Why Externalism is not part of Cognitive Science

Eran Asoulin (e.asoulin@student.unsw.edu.au)
Program in Cognitive Science
School of History & Philosophy of Science
University of New South Wales
Sydney, NSW, 2052, AUSTRALIA.

Abstract
I will compare the Internalist and Externalist approach to meaning, and show that, given empirical scientific inclinations, the Externalist approach is somewhat misguided and that parts of it can be reinterpreted via Internalism. I give a brief overview of the nature of the problem of meaning or intentionality; I then sketch out the nature of Internalism and how it deals with meaning, and conclude with a discussion of Externalism and the ways in which Externalism can be recast in terms of Internalism without loss of explanatory power. I will argue that the foundations of Externalism and Internalism are similar, and it is the Externalists, contrary to the Internalists, who feel that these foundations are not sufficient for explaining meaning and intentionality.

Keywords: Philosophy of Mind; Semantics; Internalism; Externalism; Linguistics.

The Problem of Intentionality
The problem at the crux of the Internalism/Externalism debate concerns intentionality, or the problem of how some entities can be ‘about’ something. That is, sentences, thoughts, or symbols, among other things, display intentionality in that they are about something else; they are a representation of something. (The thought ‘I would like to see a giant squid’ is about giant squid.) The notion of intentionality can be traced back at least as far as Aristotle (Brentano, 1874/1973, p. 88), though Franz Brentano is generally credited with introducing the notion to contemporary philosophy.

While Brentano used such a definition of intentionality to argue for a distinction between physical and mental phenomena (he did not believe that anything physical could display intentionality), both the Externalists and Internalists attempt to solve the intentionality puzzle within a purely physicalistic framework. It is a puzzle because, assuming materialism and wanting a purely scientific explanation, there has been a spirited dispute over what, if anything, distinguishes the nature of some physical entities (for example, thoughts, symbols, sentences, etc.) that display intentionality from other physical entities that do not (for example, sea shells, leaves, wooden doors, etc.). The Internalists claim that intentionality is practically a brain/mind-internal matter, whereas the Externalists claim that in order to explain intentionality some kind of relation (perhaps a causal one but not necessarily so) must be postulated to exist between objects outside the brain/mind and the brain/mind itself.

Internalism
Internalism is not so much a particular thesis or a specific claim as it is a research program that articulates a certain conception of the study of language, language use, and the structure of the representations of linguistic objects inside the minds of competent users of a particular language. Those who adhere to Internalism hold that, for the purposes of scientific inquiry into the aforementioned areas of language, the internal properties of the human brain are overwhelmingly the ones that are the relevant and fruitful areas of scientific research. Internalism has thus recast the ordinary everyday notion of language, a notion that is arguably at the core of Externalism, into a form that Internalists believe is susceptible to empirical scientific inquiry. This is not to say that Internalism ignores the external environment in which an organism is embedded. As is evident in the case of language acquisition, on the Internalist conception the environment plays a crucial and necessary role of triggering and shaping the path that the internal processes of the language faculty take.

The linguistic enterprise spawned by the writings of Noam Chomsky since the 1950s is a particular instantiation, probably the most well known and influential, of the Internalist approach to meaning, thus I will use mainly his writings to illustrate the research program that is Internalism.

In order to understand Chomsky’s take on Internalism, one must become familiar with the basic framework of Chomsky’s linguistic enterprise, which can be summarised as follows:
A person who has learned a language has acquired a system of rules that relate sound and meaning in a certain specific way. He has, in other words, acquired a certain competence that he puts to use in producing and understanding speech. (Chomsky, 1972, p. 11)

There are numerous issues, some more controversial than others, raised in this short quotation that require elucidation, for, in many cases, Chomsky utilises a term in a technical sense that is often divorced from its normal everyday usage. Perhaps the best place to start, then, is the language acquisition process, in which any normal healthy child free
from specific brain damage is said by Chomsky and his followers to acquire a language (this is in contrast to the traditional or everyday notion of the child ‘learning’ a language). Furthermore, it will become obvious by looking at the language-acquisition process that, contrary to a superficial reading, Internalism does not ignore or neglect the external environment; the brain/mind and the external environment are distinct and are both essential in the Internalist conception of language and meaning.

In other words, it is the story of the connection of mind to world that distinguishes Internalism from Externalism; as opposed to the uncharitable reading in which Internalism ignores the outside world and only focuses on brain/mind-internal processes.

The language-acquisition process that a child undergoes early in life arguably stands at the crux of Chomsky’s linguistic enterprise (at least in terms of evidence for the theory). This rapid and seemingly effortless process that the child undergoes en route to acquiring the ability to produce and understand speech is explained by Chomsky via the postulation of an inner (probably innate and mostly unlearned) mechanism he calls the language faculty. Such a mechanism, Chomsky argues, must exist to mediate the language acquisition process, for the child, at such a young age, does not seem to be exposed to all the necessary stimulus that would be required to acquire the ability to produce and understand speech were the child utilising general cognitive learning capacities (induction, deduction, imitation, rote memory, etc.). This claim is known as the ‘poverty of the stimulus’ argument. Furthermore, a Chomskyan would argue, even if the unlikely case could be made that the child is exposed to sufficient linguistic data that allow the general cognitive learning capacities to infer the syntax and semantics of a particular language, children acquire their first language at a remarkable rate that is entirely disproportionate to the other tasks that these general cognitive learning capacities perform.

Chomsky claims that a person inherits a Universal Grammar that provides the foundations for language to develop. What is inherited is a universal, non-language-specific, framework for language. The current formulation of the theory of Universal Grammar holds that ‘the speaker knows a set of principles that apply to all languages, and parameters that vary within clearly defined limits from one language to another’ (Cook & Newson, 1997, p. 2). In the presence of the appropriate stimuli (that is, in the presence of, usually auditory, linguistic data of a particular language, say, Zulu) the generality of Universal Grammar is focussed into the grammar of the language of the linguistic community to which the child is born. So, if a child grows up in a French linguistic community, this linguistic community would in a sense mould the child’s Universal Grammar (assumed to be represented in the language faculty in the brain) into the particularities of the French syntax and semantics. Thus, the child’s innate principles of Universal Grammar would be set to the specific parameters of the French grammar. This process of environmental triggering of the innate mechanism, however, cannot occur at any other time but during the appropriate critical time of cognitive development.

The proposed innate mechanism that mediates the acquisition of language is the language faculty, which is, crudely speaking, a (probably computational) system that relates form and meaning.

It must be understood that this linguistic hypothesis (of the internal and/or innate mechanism) is not as strong as has been claimed or as may be plausible following an uncharitable reading of Chomsky. Apart from triggering, the environment also determines the arbitrary sound-meaning relationship of words and the reference of words (in the benign sense that a particular word is used by a linguistic community to point out an object in the world), there is nothing innate or internal that specifies that the word ‘squid’ must refer to a squid and must be pronounced the way it is in English. Furthermore, the rules (or parameters) of specific grammars are arbitrary in the sense that it is an historical accident that Spanish has its particular grammatical constructions whereas Mohawk has different grammatical constructions. The internal and/or innate Universal Grammar makes specific human languages possible not by prescribing particular rules (or linguistic parameters) but by constraining the possibilities that are available for the construction of specific languages.

The language faculty, as delimited by Chomsky, is further refined by distinguishing it from both conceptual-intentional systems and performance systems (predominantly the perceptual/articulatory systems). That is, according to Chomsky, the language faculty constitutes the computational aspects of language, or the generative procedure that forms structural descriptions (SDs), each a complex of phonetic, semantic, and structural properties (Chomsky, 1992, p. 211). These internal computational processes form the building blocks of language use: they generate linguistic (syntactic) objects that are employed by the conceptual-intentional systems and performance systems to yield language production and comprehension. In this sense, then, there is only one generative engine of language: this is the syntax that constitutes the language faculty and which is submitted for sound and meaning interpretation within systems that are distinct from, though clearly closely related to, the language faculty.

Chomsky does not claim his theory fully explains how sound and meaning are interpreted from the syntax; he believes those problems are as yet unsolved, especially when compared to the relatively straightforward study of the syntactic component. Still, it is useful to briefly examine the kinds of things that fall into the category of conceptual-intentional systems and performance systems. Conceptual-intentional systems involve ‘object-reference with all of its

---

3 As we shall see later on, this framework is arguably part of a more encompassing framework of cognitive functioning that is Fodor’s ‘language of thought’.

4 See the interesting though somewhat controversial Marantz (2005).
subtleties and complexities, thematic structures, aitiational factors, and the like (Chomsky, 1980, p. 58). Performance systems, on the other hand, include the sensorimotor systems that are responsible for the perceptual and articulatory capacities.

**Meaning in Internalism**

In the Internalist approach to meaning, the semantic features of an expression can be thought of as mental instructions that interface with, and thus give information to, the conceptual-intentional systems. This approach mirrors the approach to phonology in which phonetic features of an expression are mental instructions that interface with, and thus give information to, the performance systems. Chomsky writes:

… it is reasonable to suppose that a generated expression includes two interface levels, one providing information and instructions for the articulatory-perceptual systems, the other for the conceptual-intentional systems. One interface is generally assumed to be phonetic representation (phonetic form, PF). The nature of the other is more controversial; call it LF (‘logical form’). (1992, p. 213, emphasis in original)

In other words, a linguistic expression has two interface levels: one that provides information to the performance systems to be used in the movement of the lips, tongue, etc. (PF), and one that provides information to the comparatively poorly understood conceptual-intentional systems (LF). Thus the expression gains its form and meaning pairing (PF, LF).

The detailed nature of the semantic and phonological interfaces remains a mystery, though presumably not a permanent one. Chomsky has always stated that it is an empirical issue, to be validated or refuted via scientific means, as to the nature of the language faculty. Though the overall framework of how this particular form of Internalism deals with meaning is clear enough: the (brain/mind internal and probably computational) syntax is the only generative engine of the grammar and it is subjected to phonological and semantic interpretations. So, in a loose sense, the semantic and phonological components of language are not distinct, they both have the syntactic (or computational) component at their foundations. It is in this sense that Chomsky believes parts of semantics should properly be labelled as a part of syntax. This does not mean that the language faculty is completely detached from the environment outside of the brain/mind (it’s not), it merely means that whatever connection the brain/mind has with the outside world, that connection does not constitute the content or meaning of mental representations or linguistic expressions. As will now be discussed, those who adhere to Externalism beg to differ.

**Externalism**

Jerry Fodor is a major figure of Externalism (along with Hilary Putnam, Fred Dretske, and Tyler Burge), but due to space constraints I will use mainly Fodor’s writings to illustrate the approach to explaining meaning that is Externalism. Even though Fodor’s approach to explaining meaning in language is somewhat different to that of Chomsky’s, I believe that both Fodor and Chomsky (and probably both the Externalists and the Internalists) assume the same computational and syntactic building blocks, but unlike Chomsky, Fodor does not believe that these building blocks are sufficient to explain meaning and intentionality.

Fodor conceives of semantics somewhat more abstractly than Chomsky; for example, Fodor seeks an answer to:

… the metaphysical question of the place of meaning in the world order. How can anything manage to be about anything; and why is it that only thoughts and symbols succeed? (Fodor, 1987, p. xi, emphasis in original)

That is, it appears that Fodor sees the explanation of meaning to be a metaphysical one. An Internalist attempts to deal with meaning via a computational/syntactic (largely mind-internal) approach, whereas Fodor’s attempt postulates a metaphysical nomological relation between the brain/mind and the external world in addition to the computational/syntactic mechanisms.

**The Language of Thought**

The main reason for my belief that the foundation blocks of Internalism and Externalism are similar (or at the very least that the building blocks of the approach to meaning of Chomsky and Fodor are similar) is Fodor’s (1975) postulation of a ‘language of thought’ (LOT). Fodor (1975) reasons that since cognitive psychology (and the generative grammar approach to linguistics) proposes that mental processes inherently consist of computations, and since the postulation of computations presupposes a medium in which to compute, the medium in which these mental processes occur is an internal (most probably an innate and mostly unlearned) ‘language of thought’, an internal representational system, akin to an internal language, in which the computational cognitive processes occur. The use of the word ‘language’ in the ‘language of thought’ is somewhat misleading, however, for hypothesising a language of thought in this way is not the same as hypothesising an additional internal natural language; rather, the hypothesis is of an underlying internal mechanism, a representational system that spans and underlies many cognitive capacities, including language.

---


6 See also Chomsky (2000, p. 90-91) or Stainton (2006, p. 934-935).

7 I refer here specifically to semantic Externalism to distinguish it from language Externalism. The former is the view to be discussed here; the latter view denies that languages are objects that supervene on the physical inner states of a person’s brain/mind (see Bezuidenhout (2006)).

8 There are of course other Externalist and Internalist approaches to explaining meaning; though for current purposes Fodor and Chomsky are sufficiently indicative of each approach to warrant a focus on their ideas.
Fodor states, and rightly so, that the range of cognitive capacities to which the LOT hypothesis applies is an empirical question.

I believe that Fodor’s LOT hypothesis is consistent with Chomsky’s Universal Grammar for I think it is plausible that the computations required to implement Universal Grammar are made possible via the underlying computational mechanisms of the LOT. If this is so, and granting that one is sympathetic to the Internalist approach to semantics, it would appear Fodor should agree with Chomsky in regard to syntax being the sole generative engine of language. In other words, since Fodor’s LOT hypothesis is a computational/syntactic approach to studying certain parts of the mind, and since Chomsky takes the same approach (which is cited favourably by Fodor) to studying language production and understanding (including meaning), it would appear that Fodor should agree with Chomsky on the nature of meaning. But this is not the case. The motivation for Fodor’s disagreement has two fronts: firstly, he sees the syntactic computational processes and their semantics as disparate entities, and he thus seeks a way to integrate the claim that mental states are computational with the claim that mental states are characteristically intentional. Secondly, he believes that a scientific/naturalistic description of (linguistic) representations must account for the truth of the content of the representations (linguistic or otherwise)9 and consequently that syntax (or purely mind-internal events) cannot adequately account for truth10.

Even though I believe that the first reason is a spurious distinction, for mental states are arguably intentional in virtue of their computational/syntactic properties, a proper expository account of Fodor’s ideas is necessary before any criticisms of them can be made.

Is Truth a Semantic Property?
The additions to the syntactic/computational theory of mind that Fodor proposes (i.e., to the LOT) in order to explain meaning and truth naturalistically, and this is what distinguishes Fodor’s (and most, if not all, of) Externalism from Internalism, are as follows:

1. Fodor’s insistence that truth is a semantic property, and thus should somehow be part of an organism’s internal representational system; and

2. that meaning is distinct from content in Dretske’s sense that a particular utterance has a particular meaning but carries additional (and perhaps different) pieces of information (content).

In regard to (1) Fodor states that:

... the notion that the agent can represent to himself salient aspects of the situations in which he finds himself presupposes that such familiar semantic properties as

truth and reference are exhibited by formulae in the representational system. (Fodor, 1975, p. 32)11

Imbuing mental representations with reference and truth is symptomatic of Fodor’s (and the Externalists’) belief that even though mental states (the symbols in our heads) are defined entirely in terms of internal, physical and computational relations in the brain, the intentional properties of mental states are defined in terms of the relations of mental states to the external world. Burge (1986), for example, states that

... mental states and events may in principle vary with variations in the environment, even as an individual’s physical... history, specified non-intentionally and individualistically, remains constant. (p. 6)

It is this conception of mental states that sets the Externalists apart from the Internalists; the Internalists state12, roughly, that the intentional properties of mental states exist in virtue of their brain/mind internal (syntactic or computational) properties. The Externalists, however, want to separate the syntactic/computational properties of the mind from its semantic or intentional properties: I will argue that this is due to, among others, their insistence (exemplified here by Fodor) that truth be assimilated into a computational, naturalistic, and scientific account of the brain/mind.

The details of (2), as well as the details of (1) and thus the Externalist conception as a whole, are very clear in Fodor’s discussion of the disjunction problem. The disjunction problem arises when Fodor attempts to amalgamate a version of Dretske’s information account of semantics13 with the computational/representational theory of mind. Considering the disjunction problem is an excellent way to comprehend the schism between Internalism and Externalism.

The Disjunction Problem
Fodor is sympathetic to the fact that informational semantics separates meaning from content. This separation is necessary, he reasons, because equating the meaning of a symbol with the information its tokens carry falls prey to the disjunction problem for it cannot explain misrepresentation. The solution to the disjunction problem lies at the heart of Fodor’s theory of content, for he believes that the solution suggests sufficient conditions for intentionality. However, it should be noted that the disjunction problem is only a problem to those who take the Externalist approach to meaning (that is, those who worry not only about the nature

---

10 See also Fodor (2000), in which he argues (for reasons not unrelated but distinct from intentionality) that the computational theory of mind cannot comprise ‘more than a fragment of a full and satisfactory cognitive psychology.’ (p. 1)
11 More recently, and in a different but related context, Fodor (2000) states that ‘there is nothing in the ‘evolutionary,’ or the ‘biological,’ or the ‘scientific’ worldview that shows, or even suggests, that the proper function of cognition is other than the fixation of true beliefs.’ (p. 68, emphasis in original)
12 I believe that the Externalist notion of reference can be rephrased in terms of Internalism but that truth is merely a second-hand or meta-notion.
13 For more discussion see, for example, McGilvray (1998).
of the mental representation itself but also about whether a particular mental representation is true or false, or whether it is a valid representation or a misrepresentation).

The disjunction problem can be construed as follows:\textsuperscript{15} suppose that the tokenings of the symbol for ‘kangaroo’ are nomologically dependent on the instantiations of kangaroos so that only kangaroos would cause the symbol ‘kangaroo’ to get tokened. Hence, it is a counterfactual supporting law that only kangaroos cause ‘kangaroo’ symbol tokens and thus the ‘kangaroo’ symbol has kangaroos, and only kangaroos, in its extension. The disjunction problem arises when oneonders the status of non-kangaroo-caused ‘kangaroo’ symbol tokens, as is the case, perhaps to an observer’s untrained eye, in which the instantiation of a wallaby causes the tokening of the symbol ‘kangaroo’. This is a problem because, on the informational semantics account of content ascription, any instantiation that would cause a tokening of a symbol is by definition in that symbol’s extension. Therefore, instead of the symbol expressing the property of being a kangaroo, the symbol expresses the disjunctive property of being either a kangaroo or a wallaby.

But if the ‘kangaroo’ symbol expresses the property kangaroo or wallaby, as opposed to just kangaroo, then tokenings of the ‘kangaroo’ symbol that are caused by the instantiations of a wallaby are true (they are not a misrepresentation). So kangaroo-caused ‘kangaroo’ symbol tokens are true and wallaby-caused ‘kangaroo’ symbol tokens are true. So the symbol ‘kangaroo’ does not mean kangaroo, it means kangaroo or wallaby. Fodor reasons that if this is the case that arises by following informational semantics then error and misrepresentations have not been explained; he states that it follows from informational semantics:

... that every token of a symbol is caused by something that belongs to its extension; hence that no token of a symbol can ever be false. This is, to put the case mildly, not satisfactory. (1990, p. 60, emphasis in original)

Nevertheless, Fodor would like to keep the part of informational semantics that conjures nomic relationships between brain/mind-internal symbols and the external world, so he augments informational semantics with what he calls ‘Asymmetric Dependence’ in order to both retain the nomic relationships and explain error or misrepresentation.

Fodor holds that informational semantics is sufficient for psychology to explain what a particular symbol represents, but:

... if we’ve already used up all that to establish representation, what more could be required to establish truth? (Ibid., p. 42)

In other words, informational semantics has trouble distinguishing:

... the conditions for representation from the conditions for truth. The trouble is intrinsic; the conditions... are such that, when they’re satisfied, misrepresentation cannot, by that very fact, occur. (Ibid., p. 34, emphasis in original)\textsuperscript{16}

Thus, Fodor believes that content/meaning is constituted by a nomological relation between particular internal mental symbols expressed in the language of thought and the instantiation of their corresponding properties in the world. Furthermore, it is important to realise that Fodor’s theory of content states that:

... all that matters for meaning is ‘functional’ relations (relations of nomic covariance) between symbols and their denotations. In particular, it doesn’t matter how that covariation is mediated; it doesn’t matter what mechanisms (neurological, intentional, spiritual, psychological, or whatever) sustain the covariation. (Fodor, 1987, p. 56, emphasis in original)

The Internalist Fence has been Painted with Externalism

Both Internalism and Externalism claim that intentionality (or meaning) can be explained naturalistically within an empirical scientific framework; though Externalism (specifically Fodor) goes one step further by claiming that ‘such familiar semantic properties as truth’ can be explained within the bounds of science. I believe that it is this step that sees Externalism fall off the tightrope of naturalistic inquiry and cognitive science. This, however, does not mean that Externalism is not a valid pursuit; it merely suggests that the Externalist approach to meaning is unhelpful, perhaps in a different discipline altogether, to the pursuit to scientifically explain meaning in language production and in language understanding. Chomsky seems to concur with such a conclusion when he states that:

If ‘cognitive science’ is taken to be concerned with intentional attribution, it may turn out to be an interesting pursuit (as literature is), but it is not likely to provide explanatory theory or to be integrated into the natural sciences. (Chomsky, 1992, p. 209)

The Internalist approach to the explanation of intentionality is the most, perhaps the only, promising endeavour to couch meaning within a naturalistic (or cognitive science) framework. The nomological relations and asymmetric dependences that characterize Externalism are too vague or under-defined to be properly integrated into the natural sciences (at least for the moment, though I would argue that this predicament is an intrinsic property of these Externalist notions). In addition, it is arguable that many Externalist notions, such as reference or public language, can be rephrased in terms of Internalism\textsuperscript{17}; this gives further support to the claim that Internalism should be favoured over Externalism in terms of a scientific approach to the explanation of meaning in language. There are numerous examples of how Externalist notions can be rephrased in


\textsuperscript{16} Note, however, that Fodor assumes that misrepresentation is a valid notion, an assumption that is neither obvious nor straightforward.

\textsuperscript{17} See, for example, Jackendoff (2002).
terms of Internalist notions; I will outline just one here: reference.

In opposition to the nomological relations (of reference and meaning) and mind-world (or word-world) relations of which the Externalists speak, the Internalists do not divorce the act or cognitive mechanism of referring from the reference relation itself. Thus, Chomsky maintains that

The terms themselves do not refer, at least if the term refer is used in its natural language sense; but people can use them to refer to things… (1992, p. 221)

In other words, Chomsky is sceptical that the Externalist notion of reference (called R’ in the context of the following quotation)

… can be given a coherent and useful formulation as a relation holding between expressions and some kind of things, divorced from particular conditions and circumstances of referring. If that is so, there will also be no reasonable inquiry into a notion of ‘sense’ or ‘content’ that ‘fixes reference’ (R’); at least for natural language, though there is a promising (syntactic) inquiry into conditions for language use (including referring). (Ibid., p. 226)

In closing, Externalism appears to have different goals and methods of studying language, goals and methods that are arguably unhelpful to the scientific study of language use, production, and comprehension (some Externalist questions are altogether irrelevant to science). Internalism, on the other hand, is a much better candidate for the scientific study of language for it focuses on brain/mind internal processes.

Most probably there are Externalists who agree with this dichotomy, but Fodor is not one of them, for he believes that his theory of content is naturalistic in the same sense that Chomsky claims his theory to be a naturalistic study of the mind. I would argue, however, that Fodor’s theory of content, while likely to be useful in other disciplines such as Epistemology, is not a scientific theory (in its current form, and perhaps intrinsically) for it relies on such obscure concepts (from a scientific perspective) as metaphysical nomological relations and asymmetric dependences.

It should be clear, however, that I am not claiming that Externalist notions (specifically the commonsense notion of semantic reference) do not exist. What I am claiming is that such notions are intrinsically intractable from a scientific perspective of language because they involve concepts that are alien to science. As Stainton (2006) states, these notions are ‘as ‘real’ as governments, towns, by-laws, nearness, corporations, national dishes, and so on… but, like towns and nearness, they are invisible to the working scientist… [and] cannot be ‘seen’ from the scientific perspective.’ (pp. 925-926)

References


