**What is inferentialism?**

Jaroslav Peregrin  
Institute of Philosophy, Academy of Sciences of the Czech Republic  
& Faculty of Philosophy, University of Hradec Králové

---

**Inferentialism and Representationalism**

Inferentialism is the conviction that to be meaningful in the distinctively human way, or to have a 'conceptual content', is to be governed by a certain kind of inferential rules. The term was coined by Robert Brandom as a label for his theory of language; however, it is also naturally applicable (and is growing increasingly common) within the philosophy of logic.

The rationale for articulating inferentialism as a fully-fledged standpoint is to emphasize its distinctness from the more traditional *representationalism*. The tradition of basing semantics on (such or another variant of) the concept of representation is long and rich. (Note that what is in question is representationalist *semantics*, viz. the idea that linguistic meaning is essentially a matter of representation; *not* a general thesis about the role of representations within the realm of the mental.)

The basic representationalist picture is: we are confronted with things (or other entities) and somehow make our words stand for them (individual philosophers vary, of course, about what is to be understood by *stand for*). Within this paradigm, the ‘essential’ words of our language are meaningful in so far as they represent, or stand for, something, and if there are other kinds of words, then their function is *auxiliary*: they may help compose complex representations etc. Many philosophers of the twentieth century took some form of representationalism for granted, seeing no viable alternative basis for semantics; others had more specific reasons for entertaining one or another form of it.

An alternative to representationalism was put forward by the later Wittgenstein (whose earlier *Tractatus* may be read as an exposition of a kind of representationalism): he claimed that the alternative *either an expression represents something or it is*

---

1 Thus, Tennant (2007) states: "An inferentialist theory of meaning holds that the meaning of a logical operator can be captured by suitably formulated rules of inference (in, say, a system of natural deduction)."

2 As samples, see Etchemendy’s (1990) urging of representational semantics; Fodor’s (1981; 1987; 1998) notion of semantics as an annex to his representational theory of mind, or various approaches to semantics based on the concept of reference, such as that of Devitt (1981; 1994).

3 See Wittgenstein (1922).
meaningless was wrongly conceived and that there was a third possibility, namely that “the signs can be used as in a game”\(^4\). Hence Wittgenstein’s proposal was that we should see the relation between an expression and its meaning on the model of that between a wooden piece we use to play chess and its role in chess (\textit{pawn, bishop \ldots})\(^5\). This was, of course, not a novel proposal (the comparison of language with chess had already been invoked certainly by Frege, de Saussure or Husserl). But Wittgenstein’s influence was able to bring the relationship between meaning and the rules of our language games into the limelight of discussion.

Another person to propagate the centrality of the rules of our linguistic practices for semantics was Wilfrid Sellars. He tied meaning more tightly to a specific kind of rules, namely \textit{inferential ones}\(^6\). His follower, Robert Brandom, has since made the link between meaning and inference explicit: language, he says, is principally a means of playing "the game of giving and asking for reasons", hence it is necessarily inferentially articulated, and hence meaning is the role which an expression acquires vis-à-vis inferential rules\(^7\).

\textbf{Inferentialism and logic}

In fact, roots of inferentialism can be traced back before Sellars and the later Wittgenstein. Even if we ignore its rudimentary forms which may be discernible in the writings of the early modern rationalist philosophers, such as Leibniz or Spinoza (as Brandom, 1985; 2002, argues) a very explicit formulation of an inferentialist construal of conceptual content is presented by Frege (1879, p. v). This anticipates an important thread within modern logic, maintaining that the meaning or significance of logical constants is a matter of the inferential rules, or the rules of proof, which govern them.

It would seem that inferentialism as a doctrine about the content of logical particles is plausible. For take conjunction: it seems that to pinpoint its meaning, it is enough to stipulate

\[
\begin{array}{ccc}
A \land B & A & B \\
A \land B & A & B \\
\end{array}
\]

\(\text{\footnotesize \cite{4}}\) See Waisman (1984, p. 105); cf. also Wittgenstein (1958, p. 4).

\(\text{\footnotesize \cite{5}}\) See Peregrin (to appear) for an elaboration of the model.

\(\text{\footnotesize \cite{6}}\) See esp. Sellars (1953).

(The impression that these three rules do institute the usual meaning of \( \wedge \) is reinforced by the fact that they may be read as describing the usual truth table: the first two saying that \( A \wedge B \) is true only if \( A \) and \( B \) are, whereas the last one that it is true if \( A \) and \( B \) are.) This led Gentzen (1934) and his followers to the description of the inferential rules that are constitutive of the functioning (and hence the meaning) of each logical constant. (For each constant, there was always an *introductory* rule or rules (in our case of \( \wedge \), above, the first one), and an *elimination* rule or rules (above, the last two.))\(^8\) Gentzen’s efforts were integrated into the stream of what is now called *proof theory*, which was initiated by David Hilbert – originally as a project to establish secure foundations for logic\(^9\) – and which has subsequently developed, in effect, into the investigation of the inferential structures of logical systems\(^10\).

The most popular objection to inferentialism in logic was presented by A. Prior (1960/61, 1964). Prior argues that if we let inferential patterns constitute (the meaning of) logical constants, then nothing prohibits the constitution of a constant *tonk* in terms of the following pattern

\[
\begin{array}{ll}
 A & A \\
 A \text{ tonk } B & B
\end{array}
\]

As the very presence of such a constant within a language obviously makes the language contradictory, Prior concluded that inferential patterns do not confer real meaning.

Defenders of inferentialism (prominently Belnap, 1962) argue that Prior only showed that not every inferential pattern is able to confer meaning *worth its name*\(^11\). This makes the inferentialist face the problem of distinguishing, in inferentialist terms, between those patterns which do, and those which do not, confer meaning (from Prior’s text it may seem that to draw the boundary we need some essentially representationalist or model-theoretic equipment, such as truth tables); but this is not fatal for inferentialism. Belnap did propose an inferentialist construal of the boundary – according to him it can be construed as the boundary between those patterns that are conservative over the base language and those that are not (i.e those that do not, and those that do, institute new links among the sentences of the base language). Prior’s tonk is obviously not; it institutes the inference of \( A \models B \) for every \( A \) and \( B \).

\(^8\) This works straightforwardly for intuitionist logic, thus making it more intimately related to inference than classical logic, for which this kind of symmetry is not really achievable.

\(^9\) See Kreisel (1968).

\(^10\) One of the early weakly inferentialist approaches to the very concept of logic was due to Hacking (1979).

Inferentialism in logic (which, at the time of Belnap’s discussion with Prior, was not a widespread view) has recently also been flourishing in connection with the acceleration of proof-theoretical studies and the widening of their scope to the newly created field of substructural logics.\(^{12}\)

**From proof theory to semantics**

The controversies over whether it is possible to base logic on (and especially to furnish logical constants with meanings by means of) proof theory, or whether it must be model theory, concern, to a great extent, the technical aspect of logic. But some logicians and philosophers have started to associate this explanatory order with certain philosophical doctrines.

In his early papers, Michael Dummett (1977) argued that basing logic on proof theory goes hand in hand with its intuitionist construal and, more generally, with founding epistemology on the concept of justification rather than on the concept of truth. This, according to him, further invites the "anti-realist" rather than "realist" attitude to ontology: the conviction that principally unknowable facts are no facts at all and hence we should not assume that every statement expressing a quantification over an infinite domain is true or false. Thus Dummett (1991) came to the conclusion that metaphysical debates are best settled by being reduced to debates about the logical backbone of our language.

The Priorian challenge has led many logicians to seek a 'clean' way of introducing logical constants proof-theoretically. Apart from Belnap’s response, this has opened the door to considerations concerning the normalizability of proofs (Prawitz, 1965) and the so called requirement of *harmony* between their introduction and elimination rules (Tennant, 1997). These notions amount to the requirement that an introductory rule and an elimination rule 'cancel out' in the sense that if you introduce a constant and then eliminate it, there is no gain.

Thus, if you use the introduction rule for conjunction and then use the elimination rule, you are no better off than in the beginning, for what you have proved is nothing more than what you already had.

\[
\begin{array}{c}
A \land B \\
A \land B \\
A
\end{array}
\]

\(^{12}\) See Došen & Schroeder-Heister (1993); Restall (2000).
The reason *tonk* comes to be disqualified by these considerations is that its elimination rule does not 'fit' its introductory rule in the required way: there is not the required harmony between them and proofs containing them would violate normalizability. If you introduce it and eliminate it, there may be a nontrivial gain:

\[ \begin{array}{c}
  & A \\
  A \text{tonk} B & B \\
\end{array} \]

Prawitz, who has elaborated on the Gentzenian theory of natural deduction, was led, by his considerations of the ways of making rules constitutive of logical constants as 'well-behaved' as possible, to consider the relationship between proof theory and semantics. He and his followers then developed their ideas, introducing the overarching heading of *proof-theoretical semantics*\(^\text{13}\).

It is clear that the inferentialist construal of the meanings of logical constants presents their semantics more as a matter of a certain know-how than of a knowledge of something represented by them. This may help not only to explain how logical constants (and hence logic) may have emerged\(^\text{14}\), but also to align logic with the Wittgensteinian trend to see language as more of a practical activity than an abstract system of signs. This was stressed especially by Dummett (1993)\(^\text{15}\).

**Brandom’s normative inferentialism**

Unlike Dummett, Brandom (1994; 2000) does not concentrate on logical constants; his inferentialism covers more uniformly the whole of language. As a pragmatist, Brandom sees language as a way of carrying out an activity, the activity of playing certain language games; but unlike many postmodern followers of Wittgenstein he is convinced that one of the games is 'principal', namely the *game of giving and asking for reasons*. It is this


\(^{14}\) See Tennant (1994).

\(^{15}\) A different approach to logic based on the 'practical' view of language is the game-theoretical semantics of Hintikka (1996). However, unlike the approach discussed here, this approach leads to the accentuation of the model-theoretic, rather than proof-theoretic foundations of logic.
game, according to him, that is the hallmark of what we are – thinking, concept-
possessing, rational beings abiding to the force of better reason.\(^\text{16}\)

It is this conviction that makes Brandom not only a pragmatist, but also an
inferentialist (and as we already stated, the initiator of inferentialism as a philosophical
document). For if our language is to let us play the game of giving and asking for reasons,
it must be *inferentially articulated*: To be able to give reasons we must be able to make
claims that can serve as reasons for other claims; hence our language must provide for
sentences that *entail* other sentences. To be able to ask for reasons, we must be able to
make claims that count as a *challenge* to other claims; hence our language must provide
for sentences that are *incompatible* with other sentences. Hence our language must be
structured by these entailment and incompatibility relations.

In fact, for Brandom the level of inference and incompatibility is merely a
deconstructible superstructure, underlain by certain normative statuses, which
communicating people acquire and maintain via using language. These statuses comprise
various kinds of *commitments* and *entitlements*. Thus, for example, when I make an
assertion, I *commit* myself to giving reasons for it when it is challenged (that is what
makes it an assertion rather than just babble); and I *entitle* everybody else to reassert my
assertion reflecting any possible challenges to me. I may commit myself to a claim
without being entitled to it, i.e. without being able to give any reasons for it, and I can be
committed to all kinds of claims, but there are certain claims commitment to which
blocks my entitlement to certain other claims.

Brandom’s idea is that living in a human society is steering within a rich network of
normative social relationships and enjoying many kinds of normative statuses, which
reach into many dimensions. Linguistic communication institutes an important stratum of
such statuses (commitments and entitlements) and to understand language means to be
able to keep track of the statuses of one’s fellow speakers – to keep score of them, as
Brandom puts it\(^\text{17}\). And the social distribution is essential because it provides for the
multiplicity of perspectives that makes the objectivity of linguistic content possible.

This interplay of commitments and entitlements is also the underlying source of the
relation of incompatibility - commitment to one claim excluding the entitlement to others.
Additionally, there is the relation of inheriting commitments and entitlements (by
committing myself to *This is a dog* I commit myself also to *This is an animal*, and being
entitled to *It is raining* I am entitled also to *The streets are wet*); and also the relation of

\(^{16}\) Therefore, Brandom rejects the view of philosophers such as Derrida (1976) or Lyotard (1979)
that all kinds of language games we play are, as it were, 'on the same level', or are even
incommensurable. According to him, it is only in terms of the game of giving and asking for
reasons that expressions can acquire real content.

\(^{17}\) The concept of *scorekeeping* was introduced, in a slightly different setting, by Lewis (1979).
co-inheritance of incompatibilities \((A\text{ is in this relation to } B \text{ iff whatever is incompatible with } B \text{ is incompatible with } A)\). This provides for the inference relation (more precisely, it provides for its several layers).

Brandom's inferentialism is a species of pragmatism and of the use-theory of meaning - he sees our expressions as tools which we employ to do various useful things (though they should not be seen as \emph{self-standing} tools like a hammer, but rather as tools, like, say, a toothwheel, that can do useful work only in cooperation with its fellow-tools). He gives pride of place to the practical over the theoretical, he sees language as a tool of social interaction rather than as an abstract system. Thus, any explication of the concepts such as \emph{language} or \emph{meaning} must be rooted in an account of what one \emph{does} when one communicates - hence semantics, as he puts it, "must be answerable to pragmatics".

What distinguishes him, however, from most other pragmatists and exponents of various use-theories, is the essentially normative twist he gives to his theory. In a nutshell, we can say that what his inferentialism is about are not inferences (as actions of speakers or thinkers), but rather \emph{inferential rules}. This is extremely important to keep in mind, for it is this that distinguishes his inferentialism from other \emph{prima facie} similar approaches to meaning, which try to derive meaning from the episodes of rather than from rules (see the next section).

This brings us back to the question of the way rules of language exist. Wittgenstein realized that the rules cannot all be explicit (in pain of a vicious circle), and hence we must make sense of the idea of rule implicit to a praxis. Brandom's response to this is that rules are carried by the speakers' \emph{normative attitudes} - their treatings of the utterances of others (and indeed of their own) as correct and incorrect. But though the rules exist only as underpinned by the attitudes, which is a matter of the causal order, the rules themselves do not exist within the causal order. In other words, though we may be able to describe, in a descriptive idiom, how a community can come to employ a normative idiom, the latter is not translatable into the former.
Other Varieties of Inferentialism

An approach to meaning superficially similar to Brandom's inferentialism is constituted by what has been called *inferential role semantics* (Fodor & LePore, 1992; Boghossian, 1993), being a subspecies of *conceptual role semantics* (Harman, 1987; Peacocke, 1992), which claims, in the words of Block (2005), that "meaning of a representation is the role of that representation in the cognitive life of the agent, e.g. in perception, thought and decision-making".

It is essential not to confuse this *causal* kind of inferentialism with Brandom's *normative* kind. The drawing of inferences is something that happens within the causal world (in the mind, and hence in the brain); whereas rules, though underlain by normative attitudes, which are events within the causal world, are not themselves states or events within the causal order. Unlike normative inferentialism, causal inferentialism says that the meaning consists in, or is caused by, certain events, namely individual drawings of inferences by individual speakers. (Note that inferential rules, which, according to the normative inferentialist, are the source of meaning, though underlain by certain causal attitudes of speakers, are not themselves part of the causal order.) Hence mistaking this view for the Brandomian inferentialism is pernicious. Moreover, though there certainly are causal functionalists, whether there are any serious proponents of causal inferentialism ("inferential role semantics") is less certain; and, not infrequently, the critique of causal inferentialism ("inferential role semantics") is (mis)aimed at normative inferentialists.

Brandomian inferentialism should also not be confused with doctrines to the effect that we learn something ‘inferentially’ rather than ‘directly’, e.g. theories of linguistic communication that maintain the relevant message conveyed by a speech act is always, or often, or sometimes, *inferred* from the literal meaning or the expression employed to accomplish the speech act, rather than simply coinciding with this meaning. In this sense the term is employed, e.g., by Recanati (2002).

Now, given that it is clear that inferentialism amounts to a normative (rather than causal) enterprise concentrating on the very nature of meaning (rather than characterizing individual communicative acts), we may distinguish theories according to what they take to be the scope of inferentialism. We can speak about narrow inferentialism if the scope is restricted to (plus minus) the logical vocabulary, and about wide inferentialism if it

---

18 See Zangwill (2005) for a discussion of the difference between normative and causal functionalism.

19 This kind of misunderstanding may also obscure the discussion between Brandom, on one side, and Fodor and LePore, on the other – see Fodor and LePore (2001; 2007), Brandom (2007).
extends over the whole language. We have already discussed the former one above; the latter is, more or less, restricted to the school of Sellars and Brandom.

Brandom (2007) further distinguishes between weak inferentialism, strong inferentialism and hyperinferentialism. Weak inferentialism is the conviction that an expression cannot be meaningful without playing a role in some inferences; i.e. that each meaningful expression must be part of some sentences that are inferable from other sentences and/or from which some other sentences are inferable. Weak inferentialism is clearly not incompatible with representationalism: believing that to mean something is to represent something is not incompatible with believing that sentences are inferable from other sentences. (Therefore, Brandom himself conjectures that in fact everybody would be a weak inferentialist, but I think that some representationalists would claim that an expression may be meaningful without being part of any sentence, or at least any sentence having inferential links to other sentences.)

Strong inferentialism, according to Brandom, claims that this kind of ‘inferential articulation’ (i.e. being part of sentences that enter into inferential relationships) is not only a necessary, but also a sufficient condition of meaningfulness – though it construes the concept of inferential rule more broadly than we have done so far, so that it encompasses ‘inferences’, as it were, from situations to claims and from claims to actions. (Hence it accepts such ‘inferential rules’ as It is correct to claim ‘This is a dog’ when pointing at a dog.) Hyperintensionalism, then, is the claim that ‘inferential articulation’ is a necessary and sufficient condition of meaningfulness on the narrow construal of inferential rules. This version of inferentialism is clearly untenable for a language containing empirical vocabulary.

Problems of Inferentialistic Semantics

The notion of meaning that stems from the inferentialist view is that of an inferential role. Just like being a (chess) king is nothing over and above being governed by such and such rules (of chess), so the inferentialist sees meaning thus and so as nothing over and above being governed by such and such (inferential) rules. Insofar as we take the rules to be a matter of pragmatics (but then we should stress that what we have in mind is normative pragmatics), we take semantics as being, in this sense, underpinned by pragmatics. Hence inferentialism falls into the stream of recent semantic theories which constitute what has been called the pragmatic turn.

---

20 See Sellars (1949); Peregrin (2006).
The general idea that meaning might be a matter of inferences is a frequent target of criticism. We have already discussed the objection of Prior; the more general version amounts to the dual claim that

1. inference is a matter of syntax; and
2. syntax can never yield semantics.

This also underlies the often cited objection to the Turingian idea that computers might be able to think: a computer, as Searle (1984) articulates the objection, can only have syntax (inferential rules), but never a true semantics.

The inferentialist rejoinder turns on the equivocation of the word “syntax” as used in this objection: in one sense, syntax can never yield semantics (but syntax in this sense stops short before inferences\(^\text{22}\)); in another sense syntax involves inference (but in this sense it can yield semantics\(^\text{23}\)). The case of conjunction is instructive – as the inferential pattern appears to carry the same information as the truth table that is usually considered as being represented by the operator, there seems no reason to say that the inferential pattern cannot also confer meaning in so far as the table can.

A deeper objection concentrates on empirical vocabulary. This vocabulary, it would seem, cannot become meaningful without representing something (and it is a question whether we can have a language, worth its name, without this kind of vocabulary). We have seen that Brandom himself admits that to understand meanings of empirical words as their inferential roles, we have to stretch the notion of inference beyond its usual limits. Hence, is the inferentialist finally obliged to say the same as some representationalists, namely that empirical words acquire their significance through being tools of responding to objects of the extralinguistic world?

Though it is clear that the position of the inferentialist is less secure here than with logical words, the assimilation of her position to a representationalist one would be an oversimplification. First, inferentialism commits her to a sentence holism, and so the point of contact of language and the world cannot be on the word-object level, but rather on the level sentence-situation or -action. Second, she is a normativist, hence she is not interested in which responses in fact occur, but rather in which responses are correct. And third, she is convinced that no expression can become meaningful merely in force of such contacts – it must also be situated within the network of inferences proper.

Some descendants of Brandomian inferentialism, notably Lance (1998; 2000), argue that the empirical aspect of natural language must be accounted for in terms of the embodiment of language\(^\text{24}\). Language, Lance argues, is more appropriately seen as a sport like a football than as a game like chess - though the 'space of meaningfulness' is partially

\(^{22}\) It concerns merely the well-formedness of expressions.

\(^{23}\) In this sense, syntax amounts to what Carnap (1934) called logical syntax.

\(^{24}\) Cf. also Haugeland (1998).
delimited by mere rules, which can be violated, rather than inviolable natural laws, some of the rules are rules for coping with reality and hence the space is co-delimited also by laws.

Another general issue of inferentialist semantics is the relationship between inferentialism and various formal theories of semantics which have flourished since the seminal works of Montague (1970a; 1970b) and Lewis (1972). From what was said above, it might seem that the inferentialist is bound to accept a proof-theoretical rather than model-theoretical foundation of logic and automatically reject this kind of model-theoretical semantics (which, moreover, is often seen as an embodiment of the representationalist notion of language25). However, if what is in question is natural language, then the situation is less straightforward: the representationalism/inferentialism distinction cannot be too closely aligned with the model-theoretic/proof-theoretic distinction – from the inferentialist viewpoint, model theory may be a tool for explicating the inferential roles of natural language expressions no less useful than proof theory26.

References


26 See Peregrin (2001) for a more detailed discussion of some of these issues.


