In this paper I will do three things. One, to explain why conceptual role semantics seems an attractive theory of meaning (I). Two, to sketch a version of it which has a good chance of withstanding some of the standard objections (II-III). Three, to see what follows from this version with respect to the naturalization of meaning (IV).

I

Meaning encompasses a large and rather heterogeneous group of phenomena, so the theory of meaning should be correspondingly complex. But there is a general consensus that its central component must be some version of truth-conditional semantics. The reason is that truth-conditional semantics, as it was developed by Frege, Russell and their followers, is an elegant and powerful way of capturing two all-important facts. First, that sentences are true or false partly in virtue of their meaning. Second, that meaning is compositional, i.e. the meanings of larger units are determined by the meanings of smaller units. Truth-functional semantics identifies the meaning of words with their reference or intension, the meaning of sentences with their truth conditions, and shows how the truth conditions of a sentence can be calculated from its syntactic structure and the reference of the words contained in it. So it is custom-made for the expression of these all-important facts. Its central role in the theory of meaning is fully justified.

On the other hand, there are lots of interesting issues about which truth-conditional semantics has nothing to say. I would like to mention just two of them. First, how do words acquire meaning? Truth-conditional semantics simply takes the meaning of words for granted, and does not ask where meaning comes from. Second, what does it mean to
understand a word or a sentence? If someone understands an expression, what does he know, what sort of mental representations does he have?\textsuperscript{1} The advocates of truth-conditional semantics may reject these questions by saying that they call for theories of a completely different character. Truth-conditional semantics is supposed to account for the truth-relevance of meaning and compositionality. These facts must be accounted for. If the account simultaneously illuminates other issues as well, fine. But if it does not, we should not worry about that. Even though this answer is acceptable, I think we can give a more constructive one, which is this. Truth-conditional semantics can be interpreted as a part of conceptual role semantics, and conceptual role semantics goes some way towards answering these questions—some way, not all the way.

To see why truth-conditional semantics can be reinterpreted in this fashion, consider the truth-conditional analysis of a given sentence of a natural language. The analysis itself is the translation of the natural language sentence into a transparent formal language. The translation is informative: it lays bare the syntactic and semantic structure of the sentence, it shows how the particular expressions contribute to the meaning of the sentence. The analysis is acceptable only if the translation is correct, i.e., if the truth conditions of the natural language sentence (the analysandum) agree with those of the formal language sentence (the analysans). Suppose this is indeed the case. And now here is a question. What do we know when we know this? In order to know that the truth conditions of the two sentences agree we have to know their truth conditions. But what does it mean to know the truth conditions of a sentence?

This question is puzzling, because the answers which are not false are circular. For example, take this answer: to know the truth conditions of the sentence is to understand it. But to understand a sentence is to know its meaning, and meaning is identical with the truth conditions. Nevertheless, among the circular answers there is one, which is instructive: to know the truth conditions of the sentence is to know its inferential role.\textsuperscript{2} This answer is also circular. Obviously, we should only take into account the valid inferences, since without this restriction every sentence has the same inferential role, and thus inferential role could not reveal anything about truth conditions. But an inference is valid if and only if the conclusion cannot be false provided that the premises are all true, i.e., if the satisfaction of the truth conditions of the premises guarantees that the truth conditions of the conclusion are also satisfied.

\textsuperscript{1} This might be an acute problem for the sophisticated versions truth-conditional semantics like Montague’s. How could the psychological implementation of functions ranging over possible worlds look like?

Before explaining why this answer is instructive, in spite of its circularity, here is an example. Russell, famously, analyzes the sentence, “The present King of France is bald”, in the following way: “There is an \( x \) such that \( x \) is the present King of France, and for every \( y \), if \( y \) is the present King of France, then \( y \) is identical to \( x \), and \( x \) is bald.” The same analysis can also be expressed in terms of inferential role. From “The present King of France” one can validly infer the following: (1) France has presently at least one king, (2) France does not have presently more kings, (3) if someone is presently a king of France, then he is bald. The specification of the truth conditions is equivalent to the specification of the inferential role. As it were, truth conditions and inferential roles are just different formats for expressing the same facts.

I suggest, therefore, that we substitute inferential roles for truth conditions. The substitution is not a reductive explanation of truth conditions, it is not like identifying heat with the mean kinetic energy of molecules. Inferential roles are not more fundamental than truth conditions. The substitution rests on truism and not on deep insight. All the same, it gets us closer to answering the above questions, where does meaning come from, and what it is to know the meaning of words. To see this, consider the word “and” in the sense of logical conjunction. Its meaning is simply its inferential role. Any two sentences imply their conjunction, and a conjunction implies each of its members. To know the meaning of “and” is to be disposed to draw these inferences. Someone knows the meaning of the word if he has this disposition.

Conceptual role semantics is the generalization of this idea. It claims that one of the factors—mind the qualification—which determine the meaning of words is inferential role. Words have meaning in virtue of being associated with mental representations, concepts, for short. Concepts themselves can be described by the inferential connections they have to other concepts. Conceptual role is the totality of these connections. The word “and” has the meaning it has because it is associated with a concept which plays a characteristic role in inferential processes.

Although “and” is a very special case, the way we described its conceptual role can be applied to other words as well. From “\( x \) is a cat” we may infer to “\( x \) is an animal”, “\( x \) is hairy”, and “\( x \) mews”. From “\( x \) is a dog” we may infer to “\( x \) is an animal”, “\( x \) is hairy”, and “\( x \) barks”. Inferences like these are valid in this non-formal sense: if the premise is true, so is the conclusion. In addition to these, conceptual roles also involve inferences which are not valid but only plausible. From “\( x \) is hungry” we may plausibly infer that “\( x \) wants to eat”. This inference is defeasible: if \( x \) is on a diet, he may not want to eat even though he is hungry. All the same, if someone does not see the connection between being hungry and wanting to eat, he just does not possess the concept of hunger. In fact, we should not limit the conceptual role to valid and plausible inferences. What matters is the clear disposition to infer in the given way. If
we are strongly disposed to use a concept in a kind of inference which yields false conclusions most of the time, that kind of inference is still part of the conceptual role. Moreover, a given kind of inference does not have to be regularly performed so as to belong to the conceptual role. It is the disposition which matters, not the manifestation. (Of course, we cannot easily identify the “dormant” dispositions, but that is a different issue). Furthermore, the inferential dispositions which make up the conceptual role do not rule out that we occasionally infer in ways which run counter to the established patterns. For instance, it seems bizarre to infer from “x told the truth” to “x is mistaken”. But we may sometimes infer in this way. Suppose I know that x is a notorious liar as regards the issues under discussion and I also know that this time he told the truth. In this case I may infer that he is mistaken: he says what he believes to be false, but he told the truth, so he was mistaken. An inference like this does not match the conceptual role of “true” and “mistake” (although it matches that of “lie”).

After this sketch of conceptual role semantics, let me explain why it is attractive. I just want to point out three advantages. First, it can account for the fact that meaning is more fine-grained than reality, in other words, the phenomenon of intensionality. Lois Lane, the heroine from the cartoon, believes that Superman can fly, but she also believes that Clark Kent cannot. If meaning is not more fine-grained than reality, in other words, if meaning reduces to reference, then Lois Lane believes contradictory things of the same individual. However, she surely does not commit the straightforward contradiction she would if she also believed that Clark Kent is identical with Superman. So she must be using the words “Superman” and “Clark Kent” in a sense which does not imply that Superman is Clark Kent. Indeed, she must be using them in a sense which makes the identity outright implausible. So there must be some difference between the meanings of the two names which does not correspond to any real difference. These subtle differences can be easily grasped in terms of conceptual role semantics. Lois Lane is not disposed to draw the same inferences from sentences containing “Superman” and “Clark Kent”. For instance, from the sentences “There is big trouble” and “Superman is around” she infers to “Superman appears in a minute and puts things right”. But if “Superman” is replaced with “Clark Kent”, she will not draw the analogous inference.

A second feature which makes conceptual role semantics commendable is that it can account for cases in which words have meaning only in virtue of being a member of a closed set of words. The meaning of “Monday” cannot be given as a separate item. One knows its meaning only if one knows what place it marks in the order of the days of the week. And this is easy to articulate in terms of inferential relations. From “Today is Monday” it follows that “Tomorrow is Tuesday”, etc. The same applies to technical terms in science. One cannot know what “mass” means in physics without knowing things like that the acceleration caused by
the force acting on the body is inversely proportional to the mass of the body—i.e. without knowing a good deal of the theory containing the term. The dependence of meaning on the system is manifested by the fact that physicists often make reference to the system containing the term, e.g. they speak of relativistic mass or Bohr-electron. The meaning of terms like these is holistic: it can only be described as a part of a system. This fact must be expressed in any semantic theory. Conceptual role semantics can do this rather naturally, since conceptual role itself is holistic.

A third attractive feature is that conceptual role semantics can at least make a start on understanding metaphors. Metaphors, literally speaking, are either false (“man is the wolf of man”) or trivially true (“no man is an island”). Nevertheless, we generally regard them as capable of being true and non-trivial. So they must have some content over and above their literal content. Where does this content come from? Resorting to the idea of metaphorical meaning does not help, since what needs to be explained is why a metaphor has the particular meaning it does. Moreover, we would also like to know why metaphorical meaning is indeterminate—i.e. why the content of a metaphor cannot be precisely captured by a paraphrase in which words are used literally. Conceptual role semantics gives us a clue. We understand metaphors in virtue of the fact the words as used metaphorically inherit some of the inferential connections they have when used literally. “Wolf” licenses inferences to cruel, blood-thirsty, sacrificing others to one’s own interest. “Island” licenses inferences to independence, isolation, etc. The reason why metaphors resist paraphrasing is that it is indeterminate precisely which parts of the conceptual role are inherited in the metaphorical use.

II

There is more to be said for conceptual role semantics. But I will now consider some of the things which are often said against it. The first objection is that it cannot cope with compositionality. Meaning must be compositional, otherwise we cannot make sense of two conspicuous feature of our language use and mental life, productivity and systematicity. The first means that we can understand and say sentences we have never heard before. The second means that there are systematic connections between things we can say. If we can say “The dog chases the cat”, we can also say “The cat chases the dog”. If we give up on compositionality, these features become utterly mysterious. However, conceptual roles are often not compositional. We do not think that driving cars or being drunk needs to be banned, but we believe that drunk driving does. The conceptual role of “drunk driving” cannot be derived

from the conceptual role of “drunk” and “driving”. Worse, “drunk driving” is a genuine semantically composition, as opposed to “red herring” (which doesn’t refers to herrings at all). If meaning is compositional, but conceptual role is not, meaning cannot be conceptual role.

At first blush we may be tempted to reply that meaning involves only part of the conceptual role. Although it belongs to the conceptual role of “drunk driving” that it must be banned, but it does not belong to its meaning. Knowing that drunk driving must be banned belongs to our real world knowledge about drunk driving rather than to our semantic knowledge about the expression. This reply presupposes that we can somehow divide conceptual role, that there is a criterion which distinguishes between real world knowledge and semantic knowledge. To say that certain inferences are licensed by our semantic knowledge whereas others depend on our real world knowledge is just a different way of saying that certain sentences are true solely in virtue of their meaning, whereas the truth of others depends on extralinguistic facts as well. These two claims differ only in format, just as inferential roles and truth conditions differ only in format. But claiming that some sentences are true solely in virtue of their meaning amounts to approving the notorious analytic-synthetic distinction. It seems then that conceptual role semantics faces the bleak options of either giving up compositionality or endorsing the distinction between analytic and synthetic.

A second objection explores the consequences of the first option. Suppose conceptual role semantics does not divide conceptual role into meaning-constitutive and non-meaning-constitutive parts, in other words, it does not resort to the analytic-synthetic distinction. In this case our knowledge of the external world becomes semantic knowledge. But people have different beliefs of the world. If difference in beliefs becomes semantic difference, this threatens intersubjective understanding. Suppose the concept Romeo associates with the word “bachelor” is inferentially linked to the concepts male, unmarried, free and happy, and the concept Juliet associates with the same word is linked to male, unmarried, lonely and unhappy. How do they understand each other? The same problem has emerged in the philosophy of science as an aspect of Kuhn’s and Feyerabend’s incommensurability thesis. If the meaning of theoretical terms depends on inferential connections, i. e. the place within the theory, theoretical change leads to change in meaning. Newtonian mass is independent of speed. If the mass of a body at rest is such and such, no matter what premises we add regarding speed, we can never infer to a change in mass. Einsteinian mass, on the other hand, is not independent of speed. So the terms in the two theories differ in meaning. Moreover, the threat to intersubjective understanding cannot be contained. Conceptual role semantics treats meaning holistically across the board, so even the most banal expressions might be affected.
A third charge is that conceptual role semantics cannot give an account of reference. As I said, truth conditions must figure in a theory of meaning, and truth conditions demand reference. Clearly, conceptual role is not reference. Reference is a relation between concepts and things, whereas conceptual role is a structure made up of relations between concepts. The simplest way of getting from conceptual role to reference would be to say that reference is determined by conceptual role. We may conceive this along the following lines. Our mind contains an inferential network. Each concept is characterized by its place within this network. In the real world there is a system of relations between things and structures which is isomorphic to the inferential network. A concept refers to the item the place of which in the real order matches the place of the concept in the mental order. Of course, this picture is way too simplistic. First, in reality there are far more relations than among our concepts. Second, the relations between concepts do not always match the relations between things. We may be disposed to infer from the presence of property A to the presence of B, even though A is normally not accompanied by B. Simply put, we are not omniscient, and we also sometimes wrong. So what we have is not a single perfect isomorphism between the conceptual order and the real order but rather several partial isomorphisms. This does not suffice to determine reference.

We can improve the situation by supplementing the conceptual role of certain concepts with causal connections to sensory stimulation. For example, the conceptual role of “red” will then consists not only of such inferential connections that “red” implies “colored” and “blood” implies “red”, but also of the causal connection of the concept to a particular kind of sensory stimulation. But, as Putnam’s Twin-Earth argument shows, this still does not suffice to determine reference. Let us suppose with Putnam that there is a planet which is exactly like Earth except for one respect. The tasteless, odorless, colorless liquid, which flows in rivers, falls as rain, which is drunk by people and animals, etc., which is marked by the string w-a-t-e-r, does not have the chemical composition H₂O. Let us suppose that people on both planets still adhere to the Aristotelian worldview and regard their respective liquid as one of the four elements. Under these circumstances the conceptual role of “water” is identically on both planets, even when supplemented with links to sensory stimulation. But the reference of the word is different on the two planets. On Earth “water” refers to water, i.e. H₂O, on Twin-Earth it refers to a liquid different from earthly water, to twin-water, if you like. The upshot is that the identity of supplemented conceptual role does not guarantee the identity of reference, therefore, reference is not determined by conceptual role.

This argument is conclusive, and this is indeed conceded by the advocates of conceptual role semantics. As a response, they usually ac-
knowledge that conceptual role is just one factor in meaning, and there
must be another factor as well, something which determines reference.
So they opt for a two-factor view. In this way, conceptual role semantics
becomes more defensible by becoming less ambitious. But some argue
that it is still vulnerable to this objection: what ties the two factors to-
gether? As Fodor and LePore put it: “what prevents there being an ex-
pression with the inferential role appropriate to the content 4 is prime
but the truth conditions appropriate to the content water is wet?”

III

In order to answer these objections, I will first sketch a general picture.
I believe that the meaning of words is closely connected with conceptual
role, but the two are not identical. The meanings of words are located
in the intersubjective social space, whereas conceptual roles exist in indi-
vidual minds. Each of us is in possession of a particular, at many points
idiosyncratic inferential network. This network is not like an uninter-
preted formal calculus. It is causally linked to the outside world through
sensory stimulation. External objects produce sensory stimulations,
which, in turn, activate conceptual representations. The inferential net-
work changes continuously partly under the pressure from experience,
partly in the interest of coherence and simplicity. If the inferences we
draw fail to fit the sensory stimulations, we revise the network, and we
do the same if we find contradictions or possible shortcuts.

If we consider these individual conceptual networks and conceptual
roles, the three objections above are indeed justified. First, there is no
difference between meaning-constitutive and non-meaning-constitutive
inferential connections. The connections may differ in various respects.
Some are valid inferences, some are defeasible ones. Some are stable,
others can be easily revised. Some are immediately motivated by expe-
rience, others are more removed from it. But none of these differences,
nor their combinations can select the class of meaning-constitutive in-
ferences. Second, the conceptual networks of different individuals are
indeed different. There are lots of similarities, but each network is dif-
ferent from the others. Three, these networks do not determine the ref-
erence of concepts.

And here is the point on which I diverge from other varieties of con-
ceptual role semantics. What we call meaning is not conceptual role, but
something which emerges from communication on the basis of partial
similarities between individual conceptual networks. In similar situa-
tions we often say similar things, from the same premises we often draw

4 Gilbert Harman’s theory, “(Nonsolipsistic) Conceptual Role Semantics”
with a single factor, but Ned Block argues that it is in fact equivalent with the two
factor view, “Functional Role and Truth Conditions” (Proceedings of the Aristotelian

the same conclusions. As long as the significant differences in our use of words can be explained as differences in beliefs or preferences, we regard ourselves as using the words with the same meaning. If there are serious differences which cannot be explained in this way, we conclude that the meanings are different. When we are trying to come to terms with what the others say we construe an abstract entity which is supposed to include what we share. Meaning is this constructed entity—conceptual roles work merely as the basis of the construction.

To make this picture somewhat plausible, let me mention two things in its favor. First, it can make sense of our practice of defining words. Take ostensive definitions first. As Wittgenstein has famously argued, pointing at something and saying the word alone is logically insufficient for the clarification of meaning. When you point and say “This is a dog”, you do not only point at a member of a species, but also at an individual, a shape, a size, a color, a posture, etc. You can learn the meaning this way only if the place for the word is already prepared. This place can be thought of as the blueprint for a conceptual role, which involves inferential relations to concepts belonging to certain types without the specification of the concepts. Coming to verbal definitions, what is conspicuous is that we usually do not give necessary and sufficient conditions. However, the logically unsatisfactory definitions we give still work: the partner will understand what we mean. The reason is that definitions do not serve to instruct us to construct a concept from scratch; they serve to pick out and occasionally enrich some previously given concept.

Second, if we think of meaning in this way, it is easy to understand the significance of the principle of charity. The principle holds that if we seek to understand someone, we must necessarily assume that the person in question is rational: he is aware of the obvious facts, and his thinking is generally coherent. If an interpretation attributes someone a large number of inexplicable factual or logical errors, we find it dubious: our instincts suggest that the interpreter does not really understand what is going on. This fits in nicely with the picture just sketched. We find a way of thinking irrational if the conceptual roles therein do not match ours, if the other draws weird inferences, e.g. from the fact that today is Monday he draws the conclusion that today we must avoid cats. Meaning, as I said, is constructed from the similarities between conceptual roles. Therefore, if someone interprets the words of another person in such a way that conceptual roles the other person associates with those words do not correspond to the ones we associate with those words, we indeed have reason to suspect that the interpreter did not get the meaning. To put it differently, both rationality and meaning hinges on the similarity of conceptual roles. If meaning is constructed from matching conceptual roles, and rationality consists in matching conceptual roles, charity is inescapable.

And now I can return to the objections. As for the first objection, the answer should not come as a surprise. It is not the complete conceptual role which is constitutive of meaning. And the conceptual role
in itself does not determine which aspects of are meaning-constitutive. What is meaning constitutive is only that part of the conceptual role which we share with others. Of course, the situation is somewhat more complex. The conceptual roles which different people associate with the same word do not always have a rich intersection; there are few elements present in each of the idiosyncratic conceptual roles. What we have rather is a set of typical elements. Each competent speaker involves in his concept several of these elements, but there are still differences between speakers. This is why meaning is often vague. Suppose someone feels terrible fear in dangerous situations, but—shaking and trembling—he still does the right things. Is he brave, of does he merely behave bravely?

As for the second problem, the differences between individual conceptual roles do not prevent intersubjective understanding. Understanding demands similarity not identity. Similarity explains why we say that we use the word with the same meaning. Difference is explained as difference in beliefs. The vagueness of meaning surfaces here as well. Sometimes we cannot decide where we disagree, in meanings or in beliefs. Aristotle has made a sharp distinction between courage and rashness. The former is a virtue, rashness, i.e. when one does not care about danger at all, is not. Others may think that rashness is form of courage, albeit an extreme form of it. Is this a difference in meaning or a difference in opinion?

The third objection was this. Even if conceptual role semantics gives up the ambition of explaining reference, it still has to answer why reference and conceptual role cannot get dissociated, i.e. why we cannot have a word with the conceptual role of water referring to the number 4. The answer to this worry can be formulated both in hermeneutic and cognitive terms. In hermeneutic terms, the situation is the following. If we want to find out about the reference of someone’s word, we have to see how he uses it: what inference he makes with sentences containing it, and whether there is a characteristic set of observable conditions under which he is willing to use it. In other words, we have to discover the conceptual role. Even if the conceptual role fails to yield a unique reference, like in the Twin-Earth case, we still start with the conceptual role. We ask what the Twin-Earth word “water” refers to if its conceptual role is such and such. It must refer to something on Twin-Earth which matches this role reasonably well. And that is surely not the number 4 but twin-water. The radical dissociation of conceptual role and reference is not possible unless the way we identify reference is completely independent of conceptual role, and that is not the case. The same answer can be put in cognitive terms as follows. Concepts, which can be characterized as conceptual roles, emerge from our interaction with the external world, and serve the purposes of this interaction. Our concepts may be inadequate in many respects. They may be so inadequate that they fail to refer. But they cannot be so grossly inadequate that they hook on to something completely different from what they
were designed to capture—to the number 4 instead of water. It is just very unlikely that the mechanism of concept-formation could yield such monsters.

IV

Finally, a few words about naturalization. I said that meaning is not identical with conceptual role, but it is partly determined by it. The naturalization of conceptual role does not present major theoretical difficulties: we may follow the path laid down in the heydays of functionalism. Conceptual roles are functional roles. Functional roles may be reduced to computational roles. Computational roles can be cashed out in terms of physical realization. Working out the outlines of this reduction in sufficient detail, i.e. transforming the philosophical sketch into a scientific account, is an enormous task. Moreover, the scientific account will be abstract and idealized. Conceptual roles are, for the most part, idiosyncratic, and describing them in complete detail is pointless. The emphasis should be on developing the theoretical concepts to be employed in the reduction, rather on actually carrying out the reduction.

But the reduction of conceptual roles is much less than the naturalization of meaning. Meanings cannot be reduced in this fashion. The mental state, functional state, computational state, physical structure path cannot be followed, since meanings are not mental states. Conceptual roles are not idiosyncratic meanings. This ties in with what I said in connection with the first objection. We know the meaning only if we can actually use the word. But our ability to use the word involves more than knowledge of meaning, it involves factual knowledge as well. And conceptual roles, as they are present in individual minds, cannot be divided into meaning-constitutive and not-meaning-constitutive parts. The division is possible only in the interpersonal sphere. So it is there that meaning has its place. Ontologically, meaning is an entity constructed for the purposes of making sense of what people say. Linguistic communication, when it has reached a certain level of sophistication, demands a metatheory. When we have things to say which are sufficiently complex, and when our language is developed enough to express these complex contents—which is the case with all natural languages we know—we cannot communicate successfully without occasionally resorting to a metatheory capable of describing what we do. Meaning is a theoretical entity of this metatheory, which we may call folk semantics. Folk semantics can be studied in a way similar to the study folk botany, or similar systems of belief. It may well be expected that folk semantics is not the same everywhere. For example, empirical studies have indicated that Asians are less prone to be convinced by Kripke’s arguments against description theory than Western philosophers are. We may also develop theories about how folk semantics works and how it takes shape. When we have done this, I think, we have done all that can be done about the naturalization of meaning.
This kind of naturalization may strike you as the naturalization of
witches. Some people believe in witches, and we can tell a story why this
is so. In this story we mention perfectly respectable naturalistic entities
like old women and sick cows. So is not an account of meaning like an
account of witches, an explanation of why people believe in things which
do not exist? Am I not saying that conceptual roles are decent naturaliz-
able entities, which really exist—like old women and sick cows—, but
meaning does not exist at all, it is a bad idea, just as witches. Briefly, is
not this an eliminativist account? I do not think so. The account about
witch-theory also reveals that witch-theory is unfounded. The account
of meaning does not reveal that the concept of meaning is unfounded.
The concept of meaning enables us to do things we cannot do without
it. It is like the many of the concepts we use in our social life—like the
concept of learning, social class, national identity, right-wing politics,
etc. We do not really understand how these concepts can be accommo-
dated in a naturalistic world view. But this does not mean either these
concepts need to be eliminated or that the naturalistic view is flawed. I
am aware this is a promissory note, but I am afraid I will have to leave
you with that.