Duplicating Thoughts

KIRK LUDWIG

Suppose that a physical duplicate of me, right down to the arrangements of subatomic particles, comes into existence at the time at which I finish this sentence. Suppose that it comes into existence by chance, or at least by a causal process entirely unconnected with me. It might be so situated that it, too, is seated in front of a computer, and finishes this paragraph and paper, or a corresponding one, just as I do. (i) Would it have the same thoughts I do? (ii) Would it speak my language? (iii) Would my duplicate have any thoughts or (iv) speak any language at all?

To fix the interpretation of these questions, I will take 'thought' to cover any mental state that has a representational content, where 'representational content' is intended to be neutral with respect to psychological mode. By 'psychological mode' I mean what distinguishes kinds of thoughts, such as belief, visual perceptual experience, desire, etc. Representational content, or thought content, as I will also say, determines the conditions under which a thought is true or false, veridical or non-veridical, or, more broadly, is satisfied or fails to be satisfied, independently of relativization to circumstances, possible worlds, or the like. Beliefs, desires, hopes, intentions, and perceptual experiences will all count as thoughts on this usage. The question whether one person has the same (type of) thought as another is the question whether both have a mental state with the same representational content in the same psychological mode. I will not count epistemic verbs, however, as picking out or expressing a (pure) psychological mode. Thus, although *knowing* that the time is ripe, *seeing* that there is a goldfinch in the garden, and *remembering* that my wife's birthday is next Tuesday are all thoughts, they are not here picked out by using verbs that express a (pure) psychological mode. Therefore knowing, seeing or remembering the same things is not a requirement on having the same thoughts. Further, I will take languages to be abstract objects consisting of a recursively specifiable syntax and seman-

Address for correspondence: Department of Philosophy, University of Florida, Gainesville, FL 32611-8545, USA.
Email: kludwig@phil.ufl.edu.

tics. The English language, which is constantly changing and evolving, is not a language in this sense, but rather a series of different but closely related languages.

Questions (i)–(iv) concern the conditions under which it is possible to duplicate thoughts, and, in particular, the extent to which this can be accomplished by physical duplication. It is an important, if abstract, question, for it bears centrally on the relation of thought to other things—on the place of the mind in the natural world. For the purposes of investigating this question, it will be useful to introduce a rough taxonomy of views about the determinants of thought content. Any view that holds that the contents of a person's thoughts are at least in part determined by his relational properties I will call *externalism*. (For present purposes, a property P is a relational property iff necessarily for any x, if x has P then there is a y such that y is not identical to x or any part of x and y is a contingent existent.) *Internalism* is the negation of externalism. A view is *reductionist* if it holds that conceptually necessary and sufficient or at least sufficient conditions for someone's having a thought with a given content can be given using concepts outside the family of mental concepts, *non-reductionist* otherwise. This gives us a fourfold taxonomy of positions: reductive externalism, non-reductive externalism, reductive internalism, and non-reductive internalism.

In the following, my aim will be twofold. First it is to sketch what the consequences for our initial questions are given views that fall into the different categories of our taxonomy. Second, it will be to argue specifically that not as much as might be thought is established by the central semantical arguments for externalism about the answers to our questions, and, in particular, that there are strong reasons to think that even if the semantical arguments are correct, there is a category of thoughts to which these arguments are irrelevant.

A common view is that the answer to all of the questions (i)–(iv) is 'no', and that we can know this a priori, or at least as a priori as anything gets in philosophy. According to this view, 'causal externalism', I will call it, my duplicate (let us call him 'Duped') does not have any thoughts at all and, hence, does not have a language, because our thoughts are essentially individuated at least in part by our past causal interactions with objects in our environment. On this view, at the moment Duped comes into existence, since he has no history, he has no thoughts. This is the conclusion Davidson reaches with respect to the Swampman in his well-known thought-experiment (1987). Causal externalism, as Davidson understands it, is a non-reductive externalist view; other proponents of causal externalism are more ambitious. More broadly, any view according to which our thoughts are in part essentially determined by our past relations to our environments, a version of 'historical externalism', as we can call it, will give the same answers to questions (i)–(iv).

However, there are versions of externalism besides historical externalism, which will give different answers to at least some of our questions. For example, social externalism, the view that the individuation of our thoughts

depends, at least in part, given that we intend to be using words as others in our linguistic community do, on how other members of our linguistic community use words, need not require any historical element in the determinants of thought contents (see Putnam, 1975; Burge, 1979). A social externalist will be apt to say that Duped would not have my thoughts or speak my language (the answers to (i) and (ii) are 'no'), if he finds himself in a different or in no linguistic community, but would have thoughts, and would speak some language, though perhaps not one as rich in expressive power as the one I speak, assuming I am not complete master of all the words I am in the habit of using and defer to my community on their use (the answers to questions (iii) and (iv) are 'yes'). Alternatively, one could reject social externalism, but hold that the contents of one's representational states are fixed by one's dispositions and the nature of one's current environment. We can call this 'synchronic externalism'. On this view, the answer to questions (i)–(iv) would depend not just on Duped's non-relational properties, but also on the nature of his environment. If his environment were quite different from mine, perhaps he would have quite different thoughts, and speak a quite different language, though he would have some thoughts, presumably, and speak some language. But suppose Duped were, alas, to replace me, at the instant he comes into existence. Would he not then, on this view, have my thoughts, or the thoughts I would have had, had he not replaced me, even if he would not know or remember what I would?

No, importantly, even so, he would not have (all) the thoughts I would have had, for a quite general reason which applies to all of the positions in our taxonomy. First, some of my thoughts would have been about me at a time subsequent to Duped's appearance. And since I would no longer exist if Duped replaced me, he could not have any of the thoughts about me which I would have had in his place. But this is not the central difficulty. For even if I had merely been displaced at that instant to the next room, the thoughts Duped would be thinking in the place of the thoughts I would have been thinking about myself—at least those I would have been disposed to express using the first-person pronoun—would be about Duped, not me. This result shows that the answer to question (i) is 'no' on any view, whether externalist, or internalist, reductive, or non-reductive. Duped might share some thoughts with me, but *because* he would be a physical duplicate of me, he could not share all of them.

The issue is more complicated when we turn to the question whether synchronic externalism entails that if Duped shares my environment with me he shares my language. Duped would use words the same way I do, and would speak some language or other. If one's dispositions to use words fix what language one speaks, Duped and I would speak the same language. But it is not clear that it is only one's dispositions that fix what language one speaks, even if one resists the thought that one speaks the language of one's linguistic community, even when one misuses or fails to fully understand some of its vocabulary. For consider only proper names. If proper names are, as an increasing number of philosophers of language agree,

directly referring terms, their meanings are determined by what their referents are. It is unclear, however, that any of Duped's names would refer to anything, except perhaps his own and names with which he associates as a reference determining description a purely general description. We pin down the reference of most of the names we use in part through indexically loaded descriptions, and, for many, simply through knowing that they refer to whatever they are used to refer to by the people from whom we acquired them. What works for me here would not work for Duped, since none, or few, of the corresponding indexically loaded descriptions he would use to pin down the reference of his proper names would refer. To see this, notice that many of the indexically loaded descriptions by which I secure the referents of the proper names I use, such as, for example, 'the man I met last Friday in my office', would necessarily fail of reference for Duped, since he was not around last Friday to meet anyone. In this respect, then, connected with the point of the previous paragraph, Duped would fail to speak the same language I do. To the extent to which other terms, such as natural kind terms, as many think, have their extensional or intensional properties fixed in an analogous way, the same would be true of them. The same points will hold if one accepts a causal account of reference fixing for proper names and natural kind terms. Only the pure descriptivist about proper names will want to say that Duped and I refer to the same things with the same names.

What about internalism? The story here is less complicated. First, considered as the bare denial of externalism, internalism has no implications for whether Duped would have any thoughts at all. It says only that thought contents are not relationally individuated, not that they are determined by physical type. In the case of reductive internalism, we know that the answers to questions (i) and (ii) are 'no' on general grounds independent of the distinction between internalism and externalism. However, if the reduction is to properties determined by physical type, reductive internalism will entail that Duped will have, with the exception of first-person thoughts, exactly the same thoughts I do. We will reach here the same conclusions about whether Duped speaks my language as we did in the case of synchronic externalism applied to the case of Duped replacing me.

What remains is to say something about which of these views is correct and how far that takes us toward answers to our original questions. Here I concentrate on externalism, and, after an initial survey of arguments for externalism, on how far the central semantic arguments for externalism take us toward answers to our questions.

Arguments for externalism can be sorted into the following categories.

First, there are semantic arguments, supported by thought experiments such as Putnam's well-known Twin Earth thought experiment (1975). These arguments aim to show in the first instance that the meanings of some, many, or, perhaps, all of our terms depend (not just causally) upon facts about our actual environments independently of our 'narrow' psychological states—for singular terms, what actual objects exist in our environments and what relations we bear to them; for general terms, what kinds of things exist

in our environments and what relations we bear or have borne to instances of those kinds. In these arguments it is then observed that these terms are pressed into use in attributing attitudes to people, particularly in the subordinate clauses of attitude sentences that are taken to give their contents. It is assumed that such attributions often are true and correctly give the contents of the thoughts of the individual to whom they are attributed. From this it is inferred that the contents of their thoughts are individuated at least in part by their relational properties.

Second, there is the argument from interpretation which is attributed to Davidson (1986a, b). The general strategy is to argue that (a) the basic methodological stance for investigating questions about the nature of language and thought is that of the radical interpreter, and (b) that an investigation of the assumptions required in radical interpretation shows that speakers who are radically interpretable cannot be massively in error in their empirical beliefs, and that the interpreter must take their thoughts to be, by and large, in basic cases, about the events that cause them. From these two assumptions it is inferred that speakers' thoughts are partially individuated in terms of their relations to their environments. (A fuller and more careful exposition of these arguments and a critical assessment of them can be found in Ludwig, 1992b.)

Third, there is the argument from naturalism. This argument begins with the assumption that mental properties can be reduced to naturalistic properties. It assumes that there can be no successful reduction to non-relational properties of an individual, often pointing to the failures of functionalism, the main candidate, and concludes that thought contents must be individuated at least in part by reference to an individual's relations to his environment. This is an argument more often in the background than in the foreground in discussions of externalism, but I believe it plays a powerful role in sustaining the research project in the face of its failure so far to provide any detailed or workable account of how thought contents are fixed by external relations. But it is also explicit in some writers, e.g. in Chapter 1 of Putnam (1981), in which non-reductive theories are disparaged as magical, and, hence, unscientific.

Fourth, there is the argument from skepticism. This argument begins with the assumption that skepticism about the external world is false. It assumes a foundationalist epistemology in the sense that it assumes that our evidence for what we believe about the world around us is restricted to facts about our own minds and what we can know a priori. It assumes that idealism is false, and that, by and large, facts about the world are observer- and mind-independent. The conjunction of these views, together with the demand that we be able to give a reflective justification of what we know, requires that there be some a priori route from knowledge about our own minds to knowledge of the nature of the external world which is not secured by making the world depend upon the contents of our minds; the only option is to require that the contents of our minds depend on the nature of the world we are embedded in in a way that guarantees that they are, by and large,

true. This argument tends to be more of a background motivation than a primary one, but is still a powerful influence at work.

Fifth, there are arguments for social externalism specifically, due to Tyler Burge (1979), and Hilary Putnam (1975), which appeal to thought experiments and our standard practices for attributing attitudes to others. The arguments assume that we can literally and correctly attribute to someone an attitude on an occasion using in the content clause of the attribution a sentence he would use to express it but which he does not correctly or fully understand, where the sentence used gives the content of the thought as it is interpreted relative to the community's language. Given this assumption, it is clear that the dispositions of a speaker could remain fixed while his thought contents varied.

Sixth, there are arguments which appeal directly to our intuitions about the conditions under which someone has thoughts. Davidson's thought experiment involving the Swampman is an example. Another is Putnam's (1979, Chapter 1) main line of argument, in which he presents a series of thought experiments designed to convince his readers that just as the representational content of inscriptions or physical signs is not intrinsic to them, neither is that of our phenomenal states (see Ludwig, 1992a for a criticism of Putnam's argument).

The arguments from naturalism and skepticism are obviously only as good as are the reasons we have for thinking that, necessarily, naturalism is true and skepticism about the external world false. Answering the questions whether naturalism and skepticism are true or false, however, in part depends on establishing independently whether thought contents are determined relationally or not. These arguments, then, are question-begging, and should be set aside. That externalism seems to offer a way to provide a naturalistic account of thought content, and a reply to the ancient problem of skepticism, are surely virtues of it, but could not be employed as arguments for it except insofar as we had independent reason to think that naturalism were true and skepticism false. But if the best hope of this being so depends on externalism being true, we cannot argue for externalism by appeal to these assumptions.

The argument from interpretation, and the argument for social externalism, are too complex to evaluate here. Both, I believe, fail, and I have argued for this elsewhere (Ludwig, 1992b, 1995).

The argument from intuitions about thought experiments, such as that involving the Swampman, seems to me to be extremely weak, for reasons I will come to in a moment.

What I want to take up principally is the argument from facts about the determination of the semantics of the terms we use in content clauses in attitude attributions. All of these arguments share a common assumption, which I will argue is false. The assumption is that:

    (A) If it can be demonstrated that some attitudes attributed using sentences of natural languages are relationally individuated, then

there are no thoughts which are not individuated relationally, i.e. it is of the nature of representational content that it be relational in character if the semantics of terms used to attribute attitudes are determined in such a way as to make the attitudes attributed relational in character.

If this assumption is false, then although the semantic arguments for externalism would, if sound, establish that the answers to questions (i) and (ii) were 'no', something we have good reason to believe independently of externalism, they would leave open what the answers to questions (iii) and (iv) were.

The argument for assumption (A) being false consists of two parts. First, I provide a motivation for thinking that it is not incoherent to suppose that Duped could have thoughts. Second, I argue that this result need not be in conflict with the results of the semantic arguments.

To carry out the first part of the argument, I want to begin with some remarks about Davidson's thought experiment. It is apparent the Davidson's claims about the Swampman divide into two quite distinct parts. The first set are entirely uncontroversial. The Swampman does not recognize anyone, or know anyone's name, or remember anything before its creation. The truth of these claims rests on requirements on being in specifically epistemic states. It is the second set of remarks that is supposed to do the work:

> It can't mean what I do by the word 'house', for example, since the sound 'house' it makes was not learned in a context that would give it the right meaning—or any meaning at all. Indeed, I don't see how my replica can be said to mean anything by the sounds it makes, nor to have any thoughts. (Davidson, 1987)

The unwary may be misled by the appearance of 'for example' in the first sentence of the passage into thinking that what is said here is simply a continuation of the point of the first set of remarks. But it is not. Remembering where your house is and meaning *house* by 'house' are not the same thing, or no one would ever be able to sincerely ask where he lived, something which, while unusual, is not impossible. What must do the work here is the claim that there is a connection between knowing the meaning of a word and learning it in the right context. However, while it is clear that Swampman doesn't learn the word 'house', it doesn't follow from this that he does not mean house by 'house'. For if learning a word requires recognition of its meaning by some explicit or reflective process, it is unlikely that we learn very much of the vocabulary in whose use we are competent. If learning involves merely some causal process by which we come to have the verbal dispositions that we do, then Swampman is no worse off than the rest of us. When we set aside these misleading remarks, I think we find that our untutored intuitions about the Swampman are remarkably silent. (I will return to this point below, which is an important one in reflecting on our

original questions.) Davidson, however, needs more than silence. Our intuitions have to tell us something definite about this case, not just that Swampman might not have thoughts, but that Swampman definitely does not have thoughts. It is, I think, incredible to suppose that they do.

Let us now vary the thought experiment a bit, however, to test not just our reactions to what would be the case, but also to what could be the case. For this purpose, let us return to the thought experiment in the opening paragraph, with the variation that we will now suppose that I, the author of this paper, were Duped. (Perhaps Ludwig *began* this paper but could get no further than the end of the first sentence.) Is this inconceivable? Aren't we immediately taken up by this thought? I would be sitting here, imagining that I have been invited to write a paper on just such an object as I am, that I was working against a deadline, trying to think of something clever to say. I would be wrong on two of these counts, and about much else. No one would have invited me to write this paper. I would not be writing against a deadline. I would not be married. I would have no colleagues, no parents, no siblings, no friends (no training in philosophy). I would have had no contact with any human beings, and perhaps it would be doubtful that I were a human being myself. Fortunately for me, none of this would be believable. I would, as I do, find the suggestion incredible and would be able to advance powerful arguments against it. It would, perhaps, not be quite physically impossible, but surely everything I know (everything I would know if it were not true) shows that it so unlikely that the universe could die before it happened. But, whether or not I would be, and am, justified in believing that I was (and am) not Duped, what I would be imagining, what I am imagining, does not seem to be incoherent. We seem to know precisely what we would be imagining, what would be true of us, and false of us, in those circumstances, what thoughts we would be thinking, and perhaps we would have a clear enough idea of what belief attributions to us would not be true as well: would I have any thoughts about particulars other than the ones that had come into my view in my short existence or which I could think about relative to thoughts anchored by my immediate experience? No. But I would have thoughts, even so, and a bit of reflection is all that is required to work out what would be left and what not.

It might be charged that in this thought experiment we are not really imagining what we suppose that we are, that we are not imagining what it would be like to be Duped, but instead what it would be like for us to be in Duped's place. That is, the charge is that we project our current thoughts into Duped, but don't really imagine what it would be like for Duped himself. This does not seem to me to be very plausible. In imagining what it would be like to be Duped, we rely on what it is like for ourselves because of the way we have set up the thought experiment. But in doing so, what we are imagining is that all of our thoughts, as we are immediately aware of them, remain the same, while our past is radically different from what we have supposed it to be. If there were some obvious incoherence in this, it would strike us. It does not. What undergirds the possibility of this thought

experiment is that we recognize that our first-person perspective on our own thoughts is immune to disruption by changes in our beliefs about the world around us. That is why it seems coherent to imagine that the beliefs and conscious experiences we are currently immediately aware of are the beliefs and experiences of a being with virtually no history.

If this is right, it shows that it is not incoherent (or obviously incoherent) to suppose that Duped has thoughts. We seem to be able to imagine, in some detail, if we try, what it might like to be Duped, from his point of view. The possibility of the existence of such a point of view, articulated in terms of what thoughts and experiences Duped would be aware of having, therefore presents no conceptual difficulty, unless it can be shown that it follows from equally obvious or more fundamental principles that Duped could not have thoughts of experiences.

The primary threat on this point lies in the semantic arguments for externalism. These arguments by themselves, however, establish only that attitudes attributed using sentences which contain in their that-clauses proper names, indexicals, demonstratives, and natural kind terms, whose semantic properties are fixed in part by facts about a speaker's environment beyond his immediate ken, are relationally individuated. It does not follow that our thought experiment is in some hidden way incoherent unless we add assumption (A) to these familiar arguments. Assumption (A), however, has nothing independent to be said in its favour. And in the thought experiment just recounted we have good reason to think it false. In fact, the conclusion we should draw, if we endorse the semantic arguments, is that our ordinary attitude concepts, expressed by attitude verbs such as 'believes', 'desires', 'intends', etc., do not carve psychological reality by the joints. Any concept which Duped could employ would be expressible in a natural language, and, presumably, all of them are expressible in natural languages as they are. Natural languages then should give us all the resources we need to describe correctly what thoughts Duped would have. Thus, Duped would still have many beliefs, though not all of those which he would have had without his eccentric history. Beliefs and other attitudes attributed using ordinary language psychological attitude verbs therefore include both relationally individuated and non-relationally individuated attitudes. Of these, it seems clear that those which would survive our being shorn of our histories and relations are psychologically fundamental. We could call such thoughts 'Cartesian Thoughts'. They are the attitudes we can know we have independently of investigating our environments. They are the attitudes which figure in our deliberations about what to do. It is in their light that we see ourselves as rational agents. (See Ludwig, in press, for a fuller discussion of some of these issues.)

If this is right, then the semantic arguments for externalism, even if correct, are much less important than has been assumed. They show, at most, that the answers to our initial questions (i) and (ii) are 'no'. They do not show that the answers to (iii) and (iv) are 'no', and, more importantly, they do not show anything about the kinds of thoughts in terms of which we understand

ourselves as agents. Although there is not space to pursue it here, it is clear that the same considerations can be brought to bear in the case of other arguments for externalism.

If these arguments for externalism can show at most that answers to (i) and (ii) are 'no', can we conclude that the answers to questions (iii) and (iv) are 'yes'? Not at all. While our thought experiment above establishes the possibility that Duped has thoughts, it does no more than this. As I remarked at the end of my discussion of Davidson's thought experiment involving the Swampman, our intuitions about whether Swampman has thoughts are remarkably silent. They do not tell us that he does not; nor do they tell us that he does. Whether physical duplication duplicates the class of thoughts we are concerned with is not something that philosophy could tell us. It could only if a version of conceptual reductive internalism were true; but the silence of our intuitions in the Swampman case belies this. To the extent to which physical duplication could duplicate thoughts (i.e. putting aside thoughts expressed using the first-person pronoun), it is a contingent matter whether it does. The answers to our initial questions, then, are that Duped would not have (all) the same thoughts I would, he would not speak (exactly) the same language as I do, and whether he would have any thought, or speak any language, is an empirical question.

Department of Philosophy
University of Florida

References

Burge, T. 1979: Individualism and the Mental. *Midwest Studies in Philosophy, Volume 4*. Minneapolis: University of Minnesota Press.

Davidson, D. 1984: *Inquiries into Truth and Interpretation*. Oxford University Press.

Davidson, D. 1986a: A Coherence Theory of Truth and Knowledge. In E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Cambridge, MA.: Basil Blackwell.

Davidson, D. 1986b: Empirical Content. In E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Cambridge: Blackwell.

Davidson, D. 1987: Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, 60, 441–58.

Davidson, D. 1990: The Structure and Content of Truth. *Journal of Philosophy*, 87, 279–328.

Ludwig, K. 1992a: Brains in a Vat, Subjectivity, and the Causal Theory of Reference. *Journal of Philosophical Research*, 17, 313–45.

Ludwig, K. 1992b: Skepticism and Interpretation. *Philosophy and Phenomenological Research*, 52, 317–39.

Ludwig, K. 1994: Causal Relevance and Thought Content. *Philosophical Quarterly*, 44, 334–53.

Ludwig, K. 1995: The Myth of Social Content. Manuscript.

Ludwig, K. in press: Singular Thought and the Cartesian Theory of Mind. *Noûs*.
Putnam, H. 1975: The Meaning of 'Meaning'. In *Mind, Language and Reality*. Cambridge University Press.
Putnam, H. 1981: *Reason, Truth and History*. Cambridge University Press.