Review of *Philosophical Troubles* by Saul Kripke

Graeme Forbes

Philosophical Troubles by Saul Kripke; Oxford University Press 2011; xiii + 388

This collection of thirteen papers by Saul Kripke includes six previously unpublished works: (i) 'On Two Paradoxes of Knowledge', about the Surprise Examination and Dogmatism paradoxes; (ii) 'Vacuous Names and Fictional Entities'; (iii) 'Nozick on Knowledge', (iv) 'The First Person', (v) 'Unrestricted Exportation and Some Morals for the Philosophy of Language', and (vi) 'A Puzzle about Time and Thought', concerning a paradox about thinking of sets of times which, though analogous to Russell's Paradox, is not resolved by ZF. The seven reprinted papers include four masterpieces, 'Identity and Necessity', 'Speaker's Reference and Semantic Reference', 'A Puzzle about Belief' and 'Outline of a Theory of Truth', which the reader probably already has available in multiple sources and can recite passages from by heart. So the main interest of the volume will be in the newly published works. I will comment on (ii), (iii) and (v).

 'Vacuous Names and Fictional Entities' dates from 1973 and was a precursor to Kripke's John Locke Lectures that year (now published separately by OUP). It is concerned mainly with how statements about fictional entities are to be understood,

given that description theories of names are incorrect and that a fictional term can't be assigned a unique merely possible denotation (59). Kripke suggests (63) an ontology of fictional entities, *abstracta* brought into existence by the activities of people (e.g., authors). But he also holds (68) that statements like 'Sherlock Holmes {does/doesn't} exist', 'there aren't any unicorns in Dubuque', etc., cannot express propositions, and so he spends some time wrestling with why we nevertheless ascribe them a truth-value.

I would have thought that the ontology of fictional entities handles these problems ('unicorn' refers to a particular fictional species), especially if we suppose there are operators 'in the stories' and 'in reality', roughly analogous to 'possibly' and 'actually', that we employ (usually implicitly) without much difficulty. Kripke says that he finds himself 'uneasy about invoking this as a complete solution' (71), and remarks in explanation that when atheists deny the existence of God, they may 'mean to use "God" as an empty name', which he construes as 'denying that there could be propositions of a certain kind at all' (*loc. cit.*). But atheists will say something false about a certain individual if they assert that God does not exist when in fact He does, if their use of the name is picked up from religious texts some of whose authors *met* God. So if God *doesn't* exist in reality, why wouldn't they be saying something true about a certain fictional being, that it is not a real thing? After all, they have the use-as-empty intention in both cases. In a given cultural context, 'God' may automatically refer to a certain entity, so that there is no room for a special use of 'God' as an empty name in the mouths of some speakers.

'Nozick on Knowledge' is the official version of a paper which has circulated widely as an unpublished ms. It concerns the counterfactual analysis of knowledge developed in (Nozick 1981), it is the longest paper in the volume (63 pages), and, although

Kripke occasionally allows that there may be answers to some of his points, it is relentlessly negative; indeed, some readers may find their attention wandering as the sixty-third nail is hammered into the theory's coffin.

As is well-known, Nozick (1981:172–8) attempted to replace the justification condition in the traditional account of knowledge-that-*p* as justified true belief (JTB) that *p* with two counterfactuals: (3) if it weren't that *p*, *S* wouldn't believe that *p*, and (4) if it were that *p*, *S* would believe that *p* (the idea in (4) is that belief has to persist across worlds that are very similar to the actual world and where *p* is true). (3) and (4) may be relativized to a method M of acquiring the belief (1981:179–85). Kripke objects to this, among other things, that it mishandles some central counterexamples to JTB that it was expressly designed to explain, that it produces a bizarre account of the logic of knowledge, and that it is not much help with scepticism.

Suppose you are taking a party of children around an impoverished zoo that has recently sold its famous herd of zebras and substituted painted donkeys,[1] except for one zebra, which for whatever reason remains. You are teaching the children the names of animals, and pointing into the enclosure you announce "that's a zebra", by chance picking out the one real zebra (Kripke uses the more familiar stage-prop barns case). As Kripke notes (213), there is already a problem for the relevant instance of (3), if zebras are essentially zebras, and the proposition you expressed is essentially about the zebra you actually pointed at. For then (3) has an impossible antecedent in this case, which may make it automatically true, or automatically false (take your pick). And it does seem plausible that if you can just *choose* a truth-value, the counterfactual analysis isn't *explaining* why you don't know that that's a zebra if you

---

1.   'At one zoo, workers reportedly once tried to transform two donkeys into zebras by applying stripes of hair-dye.' (*The New York Times* 22/11/13, A4, perhaps in reference to http://news.bbc.co.uk/2/hi/8297812.stm)

choose 'false'. And in this type of case, the point of placing the real *F* in a crowd of fake *F*'s is so that you fail the knowledge test because it would have been too easy to acquire a false belief; not necessarily because in the counterfactual situation there might have been a fake *F* where you are actually looking, but because you might have looked in a slightly different direction. So the falsity of the demonstrative proposition, even if it were possible, is not to the point.

In fact, in most of Kripke's cases the proposition in question is inferred, e.g., that *there is* a zebra in the enclosure.[2] But even in these cases I might still satisfy Nozick's third condition for knowledge. Perhaps the one zebra was deliberately retained because donkeys imitate the behavior of similar but more exotic animals in their company, and having a real zebra in the enclosure is crucial to the deception: without a real zebra, the zoo owner would not have dared foist his scheme on the public. Granting (4), then, (3) and (4) aren't sufficient for knowledge, since the detail we've just added obviously doesn't strengthen a claim to know.

The core of Nozick's critique of scepticism is that by his analysis, we don't know sceptical hypotheses are false but do know many ordinary truths (so knowledge isn't closed under known entailment, since that you have the body you seem to have obviously entails that you're not a disembodied brain-in-a-vat (BIV)). But non-closures licensed by Nozick's analysis are hard to swallow. Suppose that, though you are easily fooled by painted donkeys, you are an expert at discriminating Namibian from

---

2.   It's characteristic of Gettier cases that a truth is inferred from a falsehood (false lemma), but here a truth is inferred from a truth. Sometimes the problem with Gettier cases is said to be just that they violate *no false lemmas*, but Gettier and relevant-alternative counterexamples to JTB seem to be unified by the feature that in them all, it is, in the circumstances of belief acquisition, simply an accident or fortunate coincidence (in technical terminology, sheer dumb luck) that a true belief is acquired (Harman's dictator example Kripke discusses (181) may need separate accounting). It would have been fascinating to know if Kripke thinks sheer dumb luck is the crucial feature, whether *no false lemmas* is a consequence of *no accident*, if not whether the former is even correct, and whether the latter can be explained independently.

Tanzanian zebras, and the zebra before you is clearly Namibian. Then if the donkeys all look like Tanzanian zebras, it's true that if there weren't a Namibian zebra in the enclosure, you wouldn't believe there was. This means that Nozick is committed to your knowing that there's a Namibian zebra in the enclosure, but not knowing that there's a zebra in the enclosure. Since 'Namibian' is intersective, this also means knowledge isn't closed under &-Elim, the simplest of all rules of inference. Nozick embraces this form of non-closure, but Kripke shows with a range of examples how implausible this is, and how forms of closure which Nozick does embrace, e.g., under ∨-Intro, are also refuted; for example, if your belief that p satisfies Nozick's conditions for knowledge, then your belief that *p or I don't believe p* will normally fail condition (3): for if the disjunctive belief were false, you'd believe p and could easily come to believe the disjunction as the result of one step of ∨I, which leads to the conclusion that (3) isn't even necessary for knowledge (203).

So far as the refutation of scepticism is concerned, Kripke argues that Nozick's conditions don't yield failure to know that BIV-type sceptical hypotheses are false (217–8). Nozick thinks that a BIV hypothesis won't be known to be false because of condition (3): you don't know you're not a BIV because if you were you'd still believe you weren't, indeed, you would have exactly the beliefs you actually have. The trouble with this is that propositions you believe may have constituents (*brain*, perhaps) that you can only employ in thought if you've stood in some kind of information relation to the relevant objects, and there are versions of the BIV hypothesis that rule out such contact. So if it weren't true that I am not a BIV, indeed I wouldn't believe it, because I couldn't even entertain the proposition. Perhaps Nozick could stipulatively exclude this truthmaker for (3). Still, other versions of BIV hypotheses don't have this feature, e.g., that I was drugged last night, taken to a secret facility, and turned into a BIV. If

that had happened, I'd still believe it didn't. Whether this shows I don't know it didn't happen will turn on the other counterexamples to Nozick's analysis, though, not the point about inability to grasp certain concepts.

'Unrestricted Exportation and Some Morals for the Philosophy of Language' concerns scepticism about the *de re*/*de dicto* distinction in attitude ascriptions, as evinced in papers by Dennett and Quine, but, in Kripke's discussion, mainly in (Sosa 1970). The *de re*/*de dicto* contrast is manifest in 'there is someone I believe to be a spy' versus 'I believe there are spies', and part of the case against the distinction is the claim that the second formulation usually supports the first. Perhaps you believe that there are spies but only finitely many, and no two have the same height; then you reason to 'the shortest spy is a spy'. In that case, by exportation, the shortest spy is someone you believe to be a spy; and so there's someone you believe to be a spy. The exportation step is evidently the debatable one. Why would anyone endorse it?

Sosa proposes that a singular *de re* ascription 'α is someone S believes to be F' is equivalent to a singular *de dicto* ascription 'S believes β is F', iff 'β' is a *distinguished* term for S (Sosa 1970:891). So if 'the shortest spy' is distinguished, exportation goes through. One problem with this is that one *de re* formulation, with β exported, may simply have been switched for another, where β is in the attitude verb's complement. In modal languages, though '$\diamond(\beta$ is $F)$' is syntactically *de dicto* by one criterion, it classifies semantically with the *de re* if β is a rigid designator (Fine 1978). So if distinguishedness is the epistemic analogue of rigid designation, it is plausible that no reduction of the *de re* occurs even if Sosa's equivalence is correct.

Scepticism about the significance of the distinction also turns on Sosa's view that distinguishedness 'is a wholly pragmatic matter which can change radically from one occasion to the other'. Kripke's specific complaint is that none of Sosa's examples

establishes this. For example, suppose that a spy *s* and his accomplice *a* are watching a detective *d* through a window. *d* has just discovered evidence in a locker in fact owned by *s* that establishes that the locker-owner, whoever that is, is a spy. *a*, who knows that it is *s*'s locker that *d* is examining, says 'He knows you are a spy now. You must flee'.[3] Kripke is surely right to object that *d* does not know this (340): first, *d* must make some connection between the owner of the locker and *s*. But Sosa regards 'the owner of this locker' in 'he knows the owner of this locker is a spy' as exportable because it's in the interests of *s* to flee if *d* believes that α is a spy, where 'α' 'would enable *d* to track *s* down without much difficulty' (1970:892). It's partly because we don't know if 'the owner of this locker' is such a term that the example seems weak. But Sosa concludes that there isn't much substance to the notion of distinguishedness and we may as well drop it. If an exportation *seems* wrong, that is to be explained pragmatically, in a way that will vary from case to case (1970:895), preserving the logical correctness of the exportation.

Kripke's 'morals for the philosophy of language' are first, that it's a mistake to keep theory as simple as possible by pushing recalcitrant phenomena (e.g., the intuition that the accomplice's remark isn't right) off into the domain of pragmatics, which should not be treated as a wastebasket (341); and second, that while there may be circumstances in which a certain statement is *appropriate* (as in some of Sosa's other examples) that doesn't mean we need a theory which says it is interest-relative or has a meaning that depends on context, so that it is literally true in the context.[4] To think

---

3. The knowledge-attribution seems to involve importation as well, from 'you are someone he believes to be a spy' to 'he believes you are a spy'. Since occurrence in exported, transparent position permits substitution, importation is inconsistent with substitution-failure. However, this may be because of the default presence of some mechanism, say a hidden indexical, which captures a term upon its importation, resulting in change of truth-condition. But it's quite plausible that when the imported term is 'you', 'he', 'over there', etc., the opacity-inducing mechanism is normally deleted.

4. One of Sosa's best cases is that of a Commanding Officer who tells a captain that tomorrow, (whoever is) the shortest platoon member should go first (1970:890). The shortest platoon member is

so is to commit what Kripke calls the 'toy duck' fallacy, illustrated as follows (345): a parent takes a child to a toy store where there are plastic models of various animals, the child asks "Is that a goose?" and the parent replies "No, it's a duck". We should not aim for a theory that makes the parent's response, however natural, the literal truth.

But it's not clear to me that the toy-duck fallacy is a fallacy. 'Toy' used non-intersectively appears to be a 'privative' adjective (*adj* is privative iff x is an *adj-F* implies x is not an *F*). But Partee (2007) argues, partly on the basis of data from Polish, that there *are* no privatives: context simply coerces the extension of the relevant predicate to widen (so the parent spoke the truth). I used to think 'forged' would be immune to such coercion – a forged Braque is just not a Braque, period. But if a legendary forger is arrested, there might be an exhibition of his work where you point to a painting in the analytic Cubist style and ask 'Is that a Picasso?', to which I reply 'No, it's a Braque'. One account of what goes on here, Partee's, is that the extension of 'is a Braque' is widened in the context. Another might be that the context makes available to the hearer material which is *understood*, though unpronounced, in this case 'forged' or 'meant to be' (Kripke also mentions an interesting fictionalist proposal in n.62). On Partee's view, 'it's a Braque' is the literal truth. This is so on the second view as well, if the unpronounced material enters into the proposition literally expressed. Either way, there is no toy-duck fallacy. What is wrong with Sosa's examples is simply that

---

Smith, so the captain says to Smith, 'the CO wants you to go first', or to a sergeant, 'the CO wants Smith to go first' (though the latter involves a controversial importation – see n. 2). The exportation of 'the shortest platoon member' is to be allowed because of features of the context, or the interests of the captain. Certainly, it's in the captain's interests to make sure that the CO's order is obeyed, so the captain uses a name of the shortest platoon member to specify the singular state of affairs whose obtaining will in fact satisfy the CO's quantified desire, rather than leave it up to anyone's judgement who the shortest is or invite debate on the question. But this only explains why the captain would replace the description with the name to avoid something more long-winded ('he wants the shortest p.m. to go first, and you are the shortest, so you will go first to satisfy him'); it doesn't give us any reason to think the exportation preserves the truth of the attitude ascription.

Graeme Forbes   **Aug 27, 2014**

the exportation of regularly functioning definite descriptions they rely on is invalid.[5]

BIBLIOGRAPHY

Fine, Kit. 1978. Model Theory for Modal Logic Part I: The *De Re/De Dicto* Distinction. *The Journal of Philosophical Logic* 7:125–56.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, Mass.: Harvard University Press.

Partee, Barbara. 2007. Compositionality and Coercion in Semantics: The Dynamics of Adjective Meaning. In *Cognitive Foundations of Interpretation*, edited by G. Bouma, I. Krämer and J. Swarts. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Sosa, Ernest. 1970. Propositional Attitudes *De Dicto* and *De Re*. *The Journal of Philosophy* 67:883–896.