

To appear in J. Greco, ed., *Philosophers and their Critics: Ernest Sosa*, Oxford: Blackwell.

Sosa on Abilities, Concepts and Externalism

Timothy Williamson

A kind of intellectual project characteristic of Ernest Sosa is to resolve an apparently flat-out dispute by showing that it is not after all a zero-sum game. His irenic goal is to do justice to both sides and give each of them most of what it wants. In his subtle paper ‘Abilities, Concepts, and Externalism’ he applies this strategy to the dispute between internalism and externalism in the philosophy of mind. It is a pleasure to engage in discussion with a philosopher of Sosa’s fair-mindedness and analytical skills.

1. The dispute is familiar. For the internalist, one’s mental features are fully determined by one’s intrinsic state: if a thinker x in a possible situation s is in the same total intrinsic state as a thinker x^* in a possible situation s^* , then for any mental feature M , x has M in s if and only if x^* has M in s^* . Externalists such as Hilary Putnam, Tyler Burge and Donald Davidson have challenged that supervenience claim with celebrated thought experiments. In Putnam’s story, the role that H_2O plays on Earth is played by the chemically quite different but observationally quite indistinguishable liquid XYZ on Twin-Earth. In 1750, Otto on Earth and his counterpart Twin-Otto on Twin-Earth are innocent of chemical theory. They are supposed to be in the same total intrinsic state. Consider the mental feature M of believing that one is near water. Otto has M ; he believes that he is near water. His belief is true. If

Twin-Otto has M too, then he believes that he is near water, and his belief is true if and only if he *is* near water, for in general one's belief that P is true if and only if P. Twin-Otto is not near water, for XYZ is not water. Thus Twin-Otto has M only if he thereby has a false belief. But there is no more reason to attribute false belief to Twin-Otto than to Otto; they are equally neutral about the underlying nature of the liquid in their environment with the relevant appearance. Therefore, Twin-Otto does not have M; he does not believe that he is near water. He has a different belief that he might express with the sounds 'I am near water'. Believing that one is near water is an externalist mental feature. Presumably, Twin-Otto fails to believe that he is near water because, unlike Otto, he lacks the concept *water*. Then grasping the concept *water* is another externalist mental feature.

This externalist account does not satisfy Sosa. He resists the claim that Twin-Otto lacks the concept *water*, on the grounds that it might be an indexical concept (pp. 321-2, 326-8; all references are to Sosa 1993 unless otherwise indicated). The very concept that on Earth refers to H₂O might on Twin-Earth refer to XYZ. It is far from obvious that such an indexicalist account of natural kind concepts coheres with an adequate general theory of concept individuation. However, it is not the aim of this paper to refute the indexicalist account, although its motivation will in effect be questioned. We should in any case note that it does not make all propositional attitude ascriptions internalist. Even if Twin-Otto has the concept *water*, he still fails to believe that he is near water. The supposed indexicality of the concept *water* would not undermine the argument for that conclusion. The point here is that if, in attributing a propositional attitude to S*, the attributor S uses an indexical C in the complement clause, then the reference of C depends on the context of S, not on that of S*. For example, if you believe that I am hungry, your belief is true if and only if I am hungry; you do not express that belief by saying 'I am hungry'. Similarly, if Twin-Oscar believes that

he is near water, his belief is true if and only if he is near water; on Twin-Earth he does not express that belief by saying 'I am near water' or 'He is near water'. Thus believing that one is near water is still an externalist mental state.

Sosa has a second line of argument, one that does not rely on the indexicalist account of natural kind concepts. In effect, he challenges externalism by attempting to ground externalist mental states in an underlying level for which internalism holds. Such a resolution of the dispute is likely to prove more acceptable to those of internalist sympathies than to those whose sympathies go the other way. In what follows, I will argue that Sosa does not succeed in revealing an underlying internalist level.

2. Sosa's strategy depends on the widespread assumption, shared by many externalists, that concept possession is a matter of abilities or dispositions. He does not say much about which abilities or dispositions are the relevant ones, or how they are related to concept possession; indeed, it is no easy task to do so. Presumably, one has a concept if and only if one has the ability to employ it in thought, but that equivalence is not very informative. We might learn more from a link between possession of a concept *C* and an ability or disposition to ϕ , where having *C* is not already an immediate consequence of ϕ ing. Sosa indicates the kind of thing he has in mind:

Having a disposition to discriminate white objects, for example, is partially constitutive of possession of the concept of white (*as* the concept of white), which in turn is required for having the thought that snow is white. (p. 314.)

We may doubt whether much ability to discriminate white is really required to have the

concept *white*; certainly not much ability to discriminate liars is required to have the concept *liar*. However, Sosa's main line of argument does not rely on the premise that the abilities or dispositions at issue are discriminatory ones. Rather surprisingly, it starts just from the idea that concept possession is constituted by abilities or dispositions of some kind or other. The work is to be done by an analysis of abilities or dispositions in general.

At least for the sake of argument, we may grant the premise that concept possession is constituted by abilities or dispositions, and see what Sosa does with it. In this volume it can hardly be inappropriate to quote him at some length. He explains the idea with his usual lucidity:

if ϕ is an ability possessed by x then there will be *some* C and B such that the following is true:

Nec.: (x has ϕ iff *if* x were in C , it would emit behaviour B)

Thus consider the dispositional property that a round marble has of being a 'roller', defined as:

Nec.: (x has (the disposition of) being a roller iff if x were released at the top of an incline, it would roll)

And let's abbreviate the form of conditional involved as

$Cx \rightarrow Bx$

Such a conditional is normally true only relative to certain presupposed circumstances. Thus consider the conditional about rolling — i.e. if x were released at the top of an incline, it would roll — abbreviated as:

$$REx \rightarrow ROx$$

When this is true of a basketball on the surface of the Earth, its truth is relative to the rigid sphericity of that basketball and the downward pull of gravity. If the basketball were flat or if it were in a spaceship, the conditional would not be true of it.

Let's define now the 'grounds' of a true conditional of the form $Cx \rightarrow Bx$ as conditions $G_1, \dots, G_n$, holding of x , such that:

$$Cx \& G_1x \& \dots \& G_nx \rightarrow Bx$$

is true in *all* circumstances, but

$$G_1x \& \dots \& G_nx \rightarrow Bx$$

is *not* true in all circumstances; nor is any other such conditional that weakens the antecedent (without importing independent subject-matter): e.g.

$$Cx \& G_2x \& \dots \& G_nx \rightarrow Bx$$

The grounds of such a conditional will often include both grounds *intrinsic* to

the object involved (e.g. the rigid sphericity of the basketball) and grounds *extrinsic* to the object (e.g. the presence of gravitational pull exerted by a nearby massive body).

Let's now combine *all* extrinsic grounds into $G_E x$ and all intrinsic grounds into $G_I x$.

Then the grounds of $Cx \rightarrow Bx$ will be $G_E x$ and $G_I x$, so that:

$$Cx \ \& \ G_E x \ \& \ G_I x \rightarrow Bx$$

is true absolutely and in *all* circumstances. Consider now the true conditional

$$(*) \quad Cx \ \& \ G_E x \rightarrow Bx$$

(where, in the case of the basketball, we leave out of the antecedent all reference to the intrinsic state of the basketball: e.g. whether it is inflated or flat, etc.). This conditional (*) does in some sense *involve* matters extrinsic to the item of which it is true (e.g. the basketball) for it involves a relationship to an *incline, rolling, a massive external body*, etc. —all matters extrinsic to the item in question (the basketball). And yet conditional (*) is *concurrently determined* to hold of an item x simply in virtue of $G_I x$, which is something purely intrinsic to x (e.g. the basketball's being rigidly spherical). It is determined to hold of x by $G_I x$ in a sense entailing that

In any factually or counterfactually possible world, if an item x had G_I ,
the following would be true of x : $Cx \ \& \ G_E x \rightarrow Bx$.

(pp. 315-16).

This passage raises many questions.

3. One salient issue concerns Sosa's conditional analysis of abilities. Quite generally, conditional analyses tend to fail to provide necessary and sufficient conditions (Shope 1978). The truth-value of the conditional often depends on extraneous factors. Such problems apply to conditional analyses of dispositions (see Martin 1994, Lewis 1997 and Bird 1998), in particular of those dispositions relevant to the grasp of concepts (Heil and Martin 1998). Sosa applies his conditional analysis to both dispositions and abilities (p. 315). As we should expect, it is vulnerable to counterexamples.

The point can be illustrated by reference to Sosa's analysis of ' x has (the disposition of) being a roller' as 'if x were released at the top of an incline, it would roll'. A round marble does not lose its disposition or ability to roll merely because a malicious person follows it around with a hammer, ready to shatter it the moment it is released at the top of an incline, before it can start rolling. Equally, a plastic cube does not acquire a disposition or ability to roll merely because someone follows it around ready to squash it into a sphere the moment it is released at the top of an incline. The kinds of dispositions or abilities that seem most relevant to the grasp of concepts are also vulnerable to such problems. To take a jejune example, suppose that I am disposed to utter 'Cat!' in the presence of cats. I do not lose that disposition merely because, without my knowing it, the man with the hammer is following me around, ready to hit me on the head the moment I come into the presence of a cat, before I can utter 'Cat!'. Naturally, one can always look for more complex conditional analyses, but that has the look of a degenerating research programme.

Could Sosa explain concept possession directly in terms of the conditionals, without claiming that they analyse dispositions or abilities? That line does not seem very promising,

for where the conditionals diverge from the abilities and dispositions, the latter are usually more relevant to concept possession. For example, the mere readiness of my shadow to hit me with his hammer before I can speak does not prevent me from grasping the concept *cat* or understanding the word ‘cat’. Of course, the relation between concept possession and dispositions or abilities is itself far from clear, as we have already noted. Without resolving this problem, let us pass on to some others.

4. Sosa’s final step in the quoted passage is to infer that intrinsic features of the item in question concurrently determine its satisfaction of the conditional (*). Thus he moves from the necessity of the conditional $(Cx \ \& \ G_{Ex} \ \& \ G_{Px}) \rightarrow Bx$ to the necessity of $G_{Px} \supset ((Cx \ \& \ G_{Ex}) \rightarrow Bx)$, where $\supset$ is the truth-functional conditional. This inference is of the following general form (where $\Box$ symbolizes necessity):

(!) From $\Box((p \ \& \ q) \rightarrow r)$ conclude $\Box(q \supset (p \rightarrow r))$

Here p is $Cx \ \& \ G_{Ex}$, q is G_{Px} and r is Bx . But (!) is an invalid rule, and would still be invalid even if $\Box$ were dropped from the conclusion. We can see that easily by considering the special case in which r is q itself. For q follows by truth-tables from $p \ \& \ q$, so $(p \ \& \ q) \rightarrow q$ is necessarily true on any reasonable theory of the counterfactual conditional $\rightarrow$. Consequently, we can infer by (!) that $q \supset (p \rightarrow q)$ holds in all circumstances, which is to say that any counterfactual conditional with a true consequent is true. That is obviously wrong; John is alive now but it is false that if he had died last year he would have been alive now.

The point seems to apply to the case at hand. Even if Cx and the intrinsic and extrinsic grounds are jointly sufficient for Bx , it does not follow that the intrinsic grounds alone are

sufficient for the counterfactual conditional (*). Although Sosa's definitions are supposed to guarantee the compatibility of the intrinsic grounds with the antecedent of (*), for all that Sosa has said the intrinsic grounds might obtain in circumstances in which the following counterfactual also obtains: if the condition Cx and the extrinsic grounds had obtained, then the intrinsic grounds would not have obtained. For, in the circumstances, realizing Cx might involve an intrinsic difference in x . If so, the joint sufficiency for Bx of Cx and the intrinsic and extrinsic grounds is irrelevant to whether the intrinsic grounds are sufficient for (*). Perhaps Sosa thinks that special features of his case permit this difficulty to be resolved. One hopes that he will address the issue. As it stands, his argument does not establish that intrinsic features of the item in question concurrently determine (*) or a corresponding ability.

5. Imagine that Sosa has filled the gap in the final step of his argument. His approach still faces another and deeper problem. Why suppose that there always are intrinsic and extrinsic grounds such as he postulates? One might for a moment think that there must be, because one can pack as much detail as one likes into those grounds. But that is to overlook the logical structure of the argument. For a counterfactual conditional, the interesting cases of its truth are those in which its antecedent is false. Sosa requires the grounds of the conditional (G_Ix , ..., G_Ex , subsequently refactorized into intrinsic and extrinsic) both to obtain in the original case in which the conditional $Cx \rightarrow Bx$ is true and to conjoin with Cx to form a sufficient condition for Bx . If the antecedent is false in the original case and the sufficient condition for Bx is not vacuously sufficient merely by being impossible, then the grounds must obtain in at least two cases in which the conditional is true: one in which x has the ability in question unrealized (so Cx is false) and one in which x realizes the ability (so Cx and Bx are true). The grounds must have some kind of generality to obtain in such different cases. Packing too

much detail into the grounds will not serve Sosa's purpose.

So far, the point appears to require no more than moderation in defining the grounds. An example will help. I will take the ability to recognize cats, and pretend that one has this ability if and only if one is such that if one were confronted by a cat (Cx), one would say 'Cat!' (Bx). Obviously, satisfying this condition is not really either necessary or sufficient for being able to recognize cats, but that does not matter for present purposes, since they concern what Sosa says about the conditional itself, irrespective of its relation to abilities. The simplicity of the example does no harm to his argument.

Consider two cases. In case α , Anna has the ability to recognize cats ($Cx \rightarrow Bx$) but is not currently realizing it; no cat confronts her ($\sim Cx$) and she does not say 'Cat!' ($\sim Bx$). The intrinsic and extrinsic grounds for Anna's ability that obtain in α are G_Ix and G_Ex respectively. In case β , G_Ix and G_Ex still obtain but Anna realizes her ability; a cat confronts her (Cx) and she says 'Cat!' (Bx). Of course, the conditional may have different grounds in many other cases, but on Sosa's approach there should be at least one case β as described corresponding to the given case α . Now consider a third case γ that is just like α in respects intrinsic to Anna but just like β in respects extrinsic to her.¹ Thus in γ a cat does confront Anna (Cx), because γ is extrinsically like β , but she does not say 'Cat!' ($\sim Bx$), because γ is intrinsically like α . Somehow, Anna fails to register the cat. The conditional $Cx \rightarrow Bx$ has a true antecedent and false consequent and is therefore false in γ (otherwise it would fail to satisfy modus ponens). According to the conditional analysis, in γ Anna lacks the ability to recognize cats. Moreover, the intrinsic grounds G_Ix obtain in γ , for they obtain in α , which is intrinsically like γ . The extrinsic grounds G_Ex likewise obtain in γ , for they obtain in β , which is extrinsically like γ . Thus the conditional $(Cx \& G_Ex \& G_Ix) \rightarrow Bx$ is false in γ , because it has a true antecedent and a false consequent. But Sosa's argument depends on the claim that this conditional 'is true

absolutely and in *all* circumstances'. At least in this straightforward example, there cannot be intrinsic and extrinsic grounds of the kind that Sosa postulates. The intrinsic and the extrinsic are too closely connected to permit such a factorization.

One might wonder whether there can really be such a case γ , intrinsically just like α and extrinsically just like β . For example, how exactly do the intrinsic part of α and the extrinsic part of β join at Anna's retinas? Now if this were a serious difficulty, it would be one for Sosa's approach. For his separation of the grounds for the conditional into intrinsic and extrinsic components requires just the kind of modal independence of the two dimensions that makes γ possible. A view on which the intrinsic state of an item necessitated external features of its environment sounds like a strong form of externalism, perhaps one that would undermine the very distinction between the intrinsic and the extrinsic. It is at least quite different from the view that Sosa suggests. In any case, we may recall that the supposed intrinsic and extrinsic grounds have a certain generality; thus γ need not match α in absolutely every intrinsic respect to reproduce the intrinsic grounds, nor need it match β in absolutely every extrinsic respect to reproduce the extrinsic grounds. Analogously, we can treat shape and colour as independent dimensions, and combine the shape of one object with the colour of another, even though in fact they are not perfectly independent, because both shape and colour depend on atomic structure (see also Williamson 2000, pp. 73-5).

How far can the Anna example be generalized? The refutation of the conditional in γ depended on the purely extrinsic nature of its antecedent (that a cat is in front of her, whether she registers it or not). Sosa can hardly object to this, since the antecedent Cx of the original conditional in his own example (that x is released at the top of an incline) is equally extrinsic. What of antecedents that are not purely extrinsic? If Cx were 'Anna sees a cat', it would be false in γ , because she would no more internally register the cat than she does in α . Thus

Sosa's conditional $(Cx \& G_{Ex} \& G_{Ix}) \rightarrow Bx$ would have a false antecedent in such a case, and the refutation above would not apply.² Obviously, it does not follow that the counterfactual conditional would be true in γ , still less that it would be a necessary truth. It may be harder to find counterexamples simply because the condition Cx has a more complex structure, mixing the extrinsic with the intrinsic. The factorization of the grounds into intrinsic and extrinsic has been shown to fail in an example in which the condition Cx is purely extrinsic. There is no special plausibility to the claim that the factorization would be more likely to succeed in examples with mixed antecedents. The natural hypothesis is that Sosa's factorization fails in a wide variety of cases.

It is not surprising that the factorization claim is vulnerable to a triad of cases with the structure of α , β and γ . Much of the significance of thought and language involves various kinds of co-ordination between the internal and the external. The grounds for such co-ordination typically permit its achievement in more than one way. They ground both a case α , in which the intrinsic condition I is co-ordinated with the extrinsic condition E , and a case β , in which a different intrinsic condition I^* is co-ordinated with a different extrinsic condition E^* . Therefore, in a case γ intrinsically like α and extrinsically like β , the intrinsic condition I holds together with the extrinsic condition E^* ; but I need not be co-ordinated in the appropriate sense with E^* . Yet any intrinsic grounds that hold in α and any extrinsic grounds that hold in β also hold in γ . Consequently, if the sufficient grounds for co-ordination that hold in α and β could be factorized into the conjunction of intrinsic grounds with extrinsic grounds, then they would also hold in γ , which they do not when co-ordination fails in γ . Factorization into intrinsic and extrinsic components can be shown to fail with respect to mental content, perception, belief and knowledge, and non-mental conditions too (Williamson 2000, pp. 66-72, 75-80, 88-9).

6. Sosa's position is moderate. He does not claim that all the abilities or dispositions that constitute understanding are determined by intrinsic features of the thinker, just that some of them are: but we have seen reason to doubt the analysis on which he bases even that moderate claim. He proposes, however, that the relevant abilities are the same for someone on Earth and someone on Twin-Earth, because 'it would seem that their pertinent abilities would be indistinguishable' (p. 319). But he has not supplied adequate reason to accept that claim. Consider, for example, the ability to bring water in response to the request 'Please bring me water' (for simplicity, we may individuate the request phonetically). It seems entirely natural to say that people on Earth have that ability while people on Twin-Earth do not. It is certainly legitimate to assume that Otto brought water in response to each of the hundreds of utterances of 'Please bring me water' made to him, while Twin-Otto brought water in response to none of the equally numerous utterances of 'Please bring me water' made to him. Of course, if Otto and Twin-Otto were switched without realizing it, then Twin-Otto and not Otto would bring water in response to the request. But so what? If x has the ability to ϕ and y lacks that ability, it does not follow that it is impossible to put them in other circumstances in which x unluckily fails to ϕ and y luckily ϕ s. For example, although you have the ability to find a book in a certain library and I lack that ability, a malicious librarian might then interfere so that you unluckily fail to find the book while I luckily succeed. To assume that switching between Earth and Twin-Earth makes no difference to abilities and their realization is to prejudge the very point at issue. Without that assumption, Sosa has provided no reason to think that the pertinent abilities are the same. No doubt Oscar and Twin-Oscar share some abilities; what we lack is a reason to assign those abilities any special significance for the mental.

Although Sosa allows some abilities to be individuated in an externalist way, he insists on the basis of his analysis that:

for every such environment-dependent ability possessed by a subject there is a corresponding environment-*in*dependent ability or ‘ability’, which must also characterize that subject. And this latter ability would no doubt supervene on (be concurrently determined by) the intrinsic character of the subject and nothing else. (p. 320).

Much of the interest of this claim depends on what sense of ‘corresponding’ is in play. In the sense defined in ‘Abilities, Concepts, and Externalism’, there is often no such corresponding environment-independent ability. The externalist mind does not factor into an internalist mind and a surrounding environment.³

Notes

- 1 By 'extrinsic' Sosa means more than 'not intrinsic'. For the conjunction of something intrinsic with something irrelevant to it and non-intrinsic is itself non-intrinsic; since G_{Ex} combines '*all* extrinsic grounds', it would entail G_x if 'extrinsic' meant 'not intrinsic', which is not what Sosa intends. Rather, by 'extrinsic' he means something like '*purely* non-intrinsic'. Such a sense is assumed in what follows.
- 2 It is not obvious that the new conditional ('If Anna saw a cat she would say "Cat!"') does better than the old one ('If Anna were confronted by a cat she would say "Cat!"') as an analysis of the ability to recognize cats. Anna might be unable to recognize cats *because* she is unable to see them.
- 3 Thanks to Alexander Bird for the benefit of his expertise on abilities and dispositions.

References

- Bird, Alexander. 1998. 'Dispositions and antidotes'. *Philosophical Quarterly*, 48: 227-234.
- Lewis, David. 1997. 'Finkish dispositions'. *Philosophical Quarterly*, 47: 143-158.
- Martin, C. B. 1994. 'Dispositions and conditionals'. *Philosophical Quarterly*, 44: 1-8.
- Martin, C.B., and Heil, J. 1998. 'Rules and powers', in J. Tomberlin (ed.), *Philosophical Perspectives 12: Language, Mind and Ontology*. Oxford: Blackwell.
- Shope, Robert. 1978. 'The conditional fallacy in modern philosophy'. *Journal of Philosophy*, 75: 397-413.
- Sosa, Ernest. 1993. 'Abilities, Concepts, and Externalism'. In J. Heil and A. Mele (eds.), *Mental Causation*. Oxford: Clarendon Press.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.