Bain, D.T. (2004) Private languages and private theorists. *Philosophical Quarterly* 54(216):pp. 427-434.

Private Languages and Private Theorists

Abstract: Simon Blackburn objects that Wittgenstein's private language argument overlooks the possibility of a private linguist equipping himself with a criterion of correctness by confirming generalisations about the patterns in which his private sensations occur. Crispin Wright responds that appropriate generalisations would be too few to be interesting. But I show that Wright's calculations are upset by his failure to appreciate both the richness of the data and the range of theories that would be available to the linguist.

Wittgenstein famously poses a problem for the idea of a private language, i.e. a language no two people could have reason to believe they share, as a language for describing sensations would be, *if* sensations were in principle inaccessible to anyone but their subjects. The problem the aspirant speaker of such a language faces, according to Wittgenstein, is that he could never reasonably convict himself of incorrect uses of its terms. He would, Wittgenstein says, "have no criterion of correctness", and hence he would not really be speaking a language at all (1953, §258).

Simon Blackburn and Crispin Wright agree this is Wittgenstein's point.[1] But Blackburn thinks Wittgenstein overlooks that a speaker might regulate his use of a private sensation language by exploiting well confirmed generalisations about the patterns in which his sensations occur.[2] Wright offers Wittgenstein an intriguing response: even if an aspirant speaker might do this, not just any generalisation will do; indeed, it turns out that the ratio of useful to useless generalisations is so small that there is only a negligible probability of one's being able to equip oneself to understand a language in the proposed way.

In what follows, I argue that Wright's assessment of the aspirant linguist's chances is flawed. Though I suspect Wittgenstein can successfully be defended against Blackburn, my business in this paper is simply to show why, in doing so, one must not concede as much to Blackburn as Wright does.

I. BLACKBURN'S PROPOSAL

Let P_1 be a phenomenological category of sensations. Suppose a subject, *A*, undergoes sensations at times t_1 and t_2. *A* judges at t_1

> s_1: I am undergoing a P_1 sensation,

and is inclined to judge at t_2 both not-s_1 and

---

[1] See Blackburn (1984) and Wright (1986), to which all page references refer.
[2] See also Harrison (1974, p. 161), Walker (1978, p. 115), and Carruthers (1986, ch. 6).

H: the sensation I am undergoing is of the same phenomenological type as the sensation
I was undergoing at t_1.

A's inclinations at t_2 are insufficient to *justify* a verdict that his earlier judgement, s_1, was false. He might just as well deny either H or not-s_1. So the example does not show that *A* has a criterion of correctness. But, Blackburn argues, *A* would have more to go on than mere classificatory inclinations if he became a theorist about his sensations, engaged in a "project … of ordering the expectation of the occurrence of sensation, with an aim at prediction, explanation, systematisation" (pp. 299-300). Instead of H being a mere impression of the phenomenological identity of two sensations, for example, *A* might have established a *correlation* between two or more sensation types. Theories are ultimately answerable to observation, of course, but such a correlation might be sufficiently well confirmed to warrant, in a given case, protecting it against a putative counterexample by rejecting a particular sensation judgement instead. Of course, *A* would still have to choose *which* particular sensation judgement to revoke. (After all, in the example above, even if there were reason to protect H, there would still be a choice as to which of $s_1(t_1)$ and not-$s_1(t_2)$ to revoke.) But if *A* confirms more correlations and has more classificatory inclinations, the idea is that he could make a principled decision on this further matter too. So equipped, Blackburn thinks, *A* could exploit such theoretical ideals as simplicity to underpin his verdicts about the correctness and incorrectness of his sensation judgements.

Wright illustrates Blackburn's proposal as follows (pp. 239-41).[3] Suppose *A* undergoes three types of sensation: P_1, P_2, and P_3. Let "S_1" abbreviate "I underwent a P_1 sensation at some point in the preceding six minutes", "not-S_1" abbreviate "I did not undergo a P_1 sensation during the preceding six minutes"; and read "S_2", "not-S_2", "S_3", and "not-S_3" similarly, *mutatis mutandis*. (Note that I intend the capital "S" to distinguish these past-tense judgements from the present tense s_1, above.) Suppose that, during an extended period, *A* confirms that the following pattern is exhibited over any six minutes: "If I did not undergo a P_1 sensation in the preceding six minutes, then I underwent a P_2 sensation; if I underwent a P_3 sensation, then I did not undergo a P_2 sensation". This can be represented using the material conditionals

H_1: $\sim S_1 \rightarrow S_2$

H_2: $S_3 \rightarrow \sim S_2$.

Now, there are eight internally consistent sets of judgement *A* might make about any six minutes. Wright represents these "diary types", as I call them, as follows:[4]

[3] I have changed some minor aspects of Wright's presentation.
[4] The right-hand side of the table is my elaboration, explained below.

Diary type	S_1	S_2	S_3	Is the diary consistent with $\{H_1, H_2\}$?
1	T	T	T	✘
2	T	T	F	✓
3	T	F	T	✓
4	T	F	F	✓
5	F	T	T	✘ **OC** for S_3
6	F	T	F	✓
7	F	F	T	✘ **OC** for not-S_1
8	F	F	F	✘

Here, an "F" under S_1 on the fifth row means that one of the three judgements in a type-5 diary is not-S_1.

We can now apply Blackburn's idea to the judgement type, S_3. Suppose A records a type-5 diary, judging not-S_1, S_2, and S_3. The conjunction of S_2 and S_3 is inconsistent with H_2, so a correction is needed. Since H_2 is, unlike H in the original example, a well confirmed correlation, A can reasonably try to preserve it, narrowing the candidates for revision to two: S_2 and S_3. Of these, S_2 is corroborated by A's judgement, not-S_1: given H_1, revoking S_2 (i.e. substituting not-S_2) would require also revoking not-S_1. So it is simpler for A to revise S_3 instead. Hence Blackburn seems vindicated. A appears to have what Wittgenstein denied he could have: a criterion of correctness for S_3. Again, given $\{H_1, H_2\}$, A's recording a type-5 diary appears to be a circumstance in which he can reasonably revise a judgement of S_3, thereby deciding that the correct account of his inner life over those six minutes was a diary not of type-5, but of type-6.

2. WRIGHT'S OBJECTION

Relative to a theory, a diary type is what I call "optimally correctable" (OC) for a type of sensation judgement, S_i, if and only if any diary of that type is such that

(i) it includes a judgement of S_i

(ii) it is inconsistent with the theory,

(iii) consistency can be restored in a way that involves revising S_i within that diary, and

(iv) all other ways of restoring consistency involve more corrections to that diary than ways that involve revising, within that diary, S_i.[5]

Hence the preceding paragraph shows that $\{H_1, H_2\}$ renders the fifth diary type OC for S_3, as it does the seventh type for not-S_1. But, to appreciate Wright's claim that not just any

---

[5] The terminology and formulation are mine, but see Wright (1986), p. 241, pp. 246-8, and p. 259.

generalisation will serve the private linguist's purposes, begin by noticing that $\{H_1, H_2\}$ fails, by contrast, to generate OC diary types for S_1, S_2, not-S_2, and not-S_3,. For these judgements, any diary type satisfying the first three conditions for being OC fails the fourth.

In the case of S_1 and not-S_3, for example, they fail because, if any diary inconsistent with $\{H_1, H_2\}$ were recorded, there would be a way of restoring consistency that involved *fewer* corrections to that diary than ways that involve revising the judgement, within that diary, either of S_1 or of not-S_3. Take S_1, for example. Type-1 diaries are the only type inconsistent with $\{H_1, H_2\}$ that involve S_1. And, admittedly, if one were recorded, consistency could be restored by revising S_1 and S_3 together. But it could also be restored by revising either S_2 or S_3 alone. So S_1 lacks an OC diary relative to $\{H_1, H_2\}$.

In the case of S_2 and not-S_2, relevant diaries fail the fourth condition because, if any diary inconsistent with $\{H_1, H_2\}$ were recorded, there would be a way of restoring consistency that involved *the same number* of corrections to that diary as ways that involve revising the judgement, within that diary, either of S_2 or of not-S_2. Take S_2, for example. The only diaries that are both recalcitrant and contain S_2 are type-1 and type-5. If a type-1 diary were recorded, admittedly, consistency could be restored by revising S_2, but it could also be restored by revising S_3. As for type-5 diaries, we have already seen that, if one were recorded, the way of restoring consistency that would involve fewest corrections to that diary would be revising its judgement of S_3, not S_2. So S_2 lacks an OC diary relative to $\{H_1, H_2\}$.

Now, Wright's objection to Blackburn crucially, if implicitly, involves the following conditional:

> **W**: a generalisation determines a criterion of correctness for a putative judgement type, S_i, only if it determines an OC diary type for S_i.

W can be seen to be operative, for example, in Wright's slide from the preceding account of why S_2 lacks an OC diary type to the view that there is no situation in which it would be *reasonable* for A to revoke a judgement of S_2 (pp. 241-2). Assuming that the simplest way to restore consistency is the most reasonable, and noticing that the only recalcitrant diaries involving S_2 are the first and fifth, Wright clearly has the following idea. On the one hand, if a type-1 diary were recorded, then no correction would be reasonable, since, although revoking S_2 and revoking S_3 would both be more simple corrections than any others, neither would be more simple than the other, and hence there would be no basis for choosing which to make. On the other hand, if a type-5 diary were recorded, the simplest and hence most reasonable way of restoring consistency would involve revoking S_3, not S_2. Thus using W, Wright concludes that, relative to $\{H_1, H_2\}$, A lacks a criterion of correctness for S_2. And he draws the same conclusion for S_1, not-S_2, and not-S_3. Hence these fail to be types of genuine judgement.

In three steps, Wright reaches a more ambitious conclusion. First, he suggests that a judgement can be genuine only if its truth-functional compounds are, and that *merely* putative judgements could hardly render a diary inconsistent with a theory. Hence he argues that the lack of criteria of correctness for S_1, S_2, not-S_2, and not-S_3 has a "rotten apple effect", undermining the *prima facie* claim of the remaining types to being genuine (pp. 242-3). Second, Wright thinks this rotten apple effect makes plausible a further conditional: a theory will generate criteria of correctness for judgements about *any* sensation types it concerns only if it generates criteria of correctness for judgements about *all* of those types (p. 247). Given W, this means that a theory must generate *OC diaries* for all such judgements. Third, Wright presents extensive formal work (due largely to Warren Goldfarb), aiming to show that the ratio of theories that meet this condition to theories that fail it is very small, and smaller the more types of sensation the theories concern (pp. 258-266). Thus, he concludes, A has a very low chance of confirming a theory equipping him to speak a private language. If, for instance, the theory in question is to range over four sensation types, then on Wright's calculations there is a one in 8,192 chance of an aspirant private linguist confirming a correlation that fits the bill!

Wright thinks this conclusion will worry friends of privacy for two reasons (p. 250): first, because he has shown that Blackburn's theorising proposal makes the possibility of a subject's speaking a private language *contingent* on the precise patterns in which his sensations occur; and second, because Wright *thinks* he has shown that the chances of a subject's sensations exhibiting an appropriate pattern is very small. Even if these two points are right, however, it is unclear why friends of privacy need be anxious. For one thing, it is surely Wittgenstein, rather than friends of privacy, who would reject the dependence of private sensation languages on the patterns the sensations exhibit. As Wright concedes, Wittgenstein seems to think a private language is *logically* impossible; he would surely not be insouciant about its being merely *improbable*. For another thing, friends of privacy might be. They might point out the following parallel: that Wittgenstein's own rule-following considerations show the possibility of a *public* language to be highly contingent.[6] That contingency is tolerable, they might say, if only because the actual world is patently one in which public language *is* possible; and they might suggest that Wright's probabilities in the private case are tolerable too, on parallel grounds. Be all that as it may, the objection I want to develop against Wright is different: namely, even if the private language issue *were* one of the aspirant linguist's odds, Wright has underestimated them.
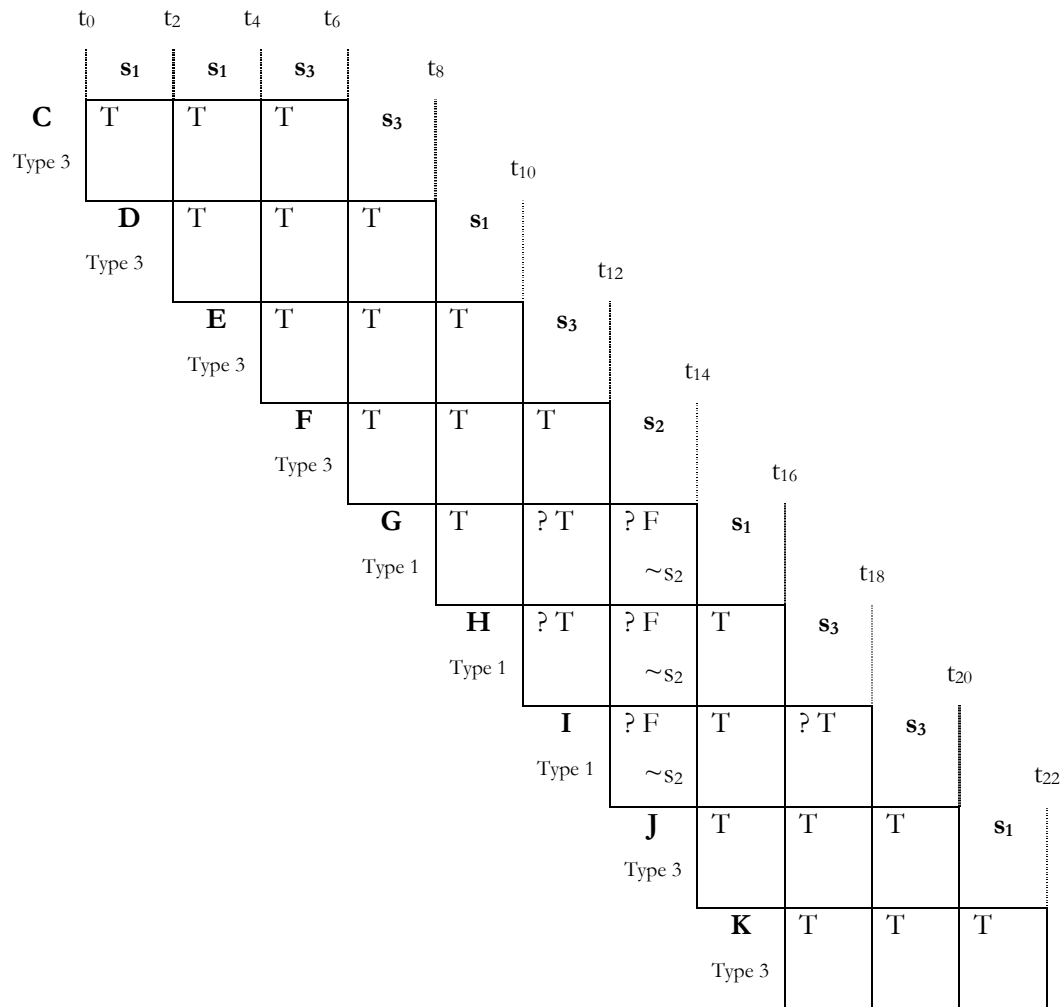
3. CONDITIONAL W

One important reason Wright underestimates the private linguist's chances is that W, the crux of Wright's calculations, is false. To take one of its counterexamples, {H_1, H_2} does

---

[6] See Moore (2003).

determine circumstances in which A would have grounds for correcting a judgement of S_2, notwithstanding the fact that $\{H_1, H_2\}$ does *not* provide an OC diary for S_2. (Or, to exercise proper caution, $\{H_1, H_2\}$ determines criteria of correctness for that judgement type *unless* such criteria are undermined by the rotten apple effect, to which I return below.)

To see this, recall that $\{H_1, H_2\}$ is a theory confirmed as holding over any six minutes. Now Wright is thinking of A's diaries being recorded in serial succession, concerning consecutive periods of six minutes. This undermines the natural reply against Wright that, if A records a recalcitrant diary that cannot be non-arbitrarily revised *now*, then A might for the time being continue to record his classificatory inclinations until he provides himself with sufficient data to enable a *later* principled revision of that earlier diary. This reply is undermined because, when we think of diaries recorded in serial succession, it is difficult to see how a collection of diaries which *individually* provide no reason to change S_2 could fare any better *collectively*. However, if $\{H_1, H_2\}$ holds over any six minute period, there is no reason why we should follow Wright in thinking of it being applied only to diaries recorded in serial succession. Surely, A can start a new diary as soon after its predecessor commences as he likes.

To flesh out this possibility, suppose that the *past-tense* judgements constituting a diary, such as "I underwent a P_1 sensation at some point in the preceding six minutes" (that is, judgement S_1), are based on *present-tense* judgements made during the six minutes in question, such as "I am undergoing a P_1 sensation" (which judgement I abbreviate with the lower-case "s_1"). Suppose, then, that every two minutes, starting at t_1, A undergoes a sensation, about which he makes a present-tense judgement; and, every two minutes, starting at t_0, he begins a new six-minute diary. The crucial upshot is that, after six minutes, every token, present-tense judgement A makes will contribute not to one diary, but to three, as illustrated below:

Diary	Sensation →	s₁	s₁	s₃	s₃	s₁	s₃	s₂	s₁	s₃	s₃	s₁
	times	t₀	t₂	t₄	t₆	t₈	t₁₀	t₁₂	t₁₄	t₁₆	t₁₈	t₂₀/t₂₂
C (Type 3)		T	T	T								
D (Type 3)			T	T	T							
E (Type 3)				T	T	T						
F (Type 3)					T	T	T					
G (Type 1)						T	?T	?F ∼s₂				
H (Type 1)							?T	?F ∼s₂	T			
I (Type 1)								?F ∼s₂	T	?T		
J (Type 3)									T	T	T	
K (Type 3)										T	T	T

In this diagram, each horizontal row of three squares represents a diary, named with a letter to its left.[7] The abbreviation above each square represents the present-tense judgement the private linguist makes when he undergoes each sensation. Moreover, suppose that when A judges that a sensation of one type occurs, he simultaneously judges that sensations of the other types do not. When he judges s_1, for example, he also judges not-s_2 and not-s_3. Hence by t_6, for example, A has recorded a type-3 diary, C, since the record of his present-tense judgements between t_0 and t_6 determines the pre-theoretical, past-tense conclusion that, while sensations of types P_1 and P_3 have occurred over that period, no sensation of type P_2 has. Now, a "T" in a box indicates a post-theoretical confirmation of a pre-theoretical judgement; an "F" indicates a post-theoretical revision, the new judgement being written beneath the "F"; a question mark indicates that a judgement is one of a pair in that diary such that, though one should be revised, there is no basis at the time of completing the diary for a principled decision as to which.

---

[7] Strictly, a diary is a set of three past-tense, not present-tense, judgements. But my relaxation in usage is unproblematic, provided past-tense judgements made at the end of the six minute period reflect present-tense judgements made during it.

The significance of the diaries' overlapping is this: if A revokes the t_{13} judgement of s_2 in diary I, for example, A *thereby* revokes it in diaries G and H, since the judgement s_2 in all three diaries is one and the same token judgement. Crucially, then, overlapping diaries create the possibility that some alterations to a recalcitrant diary will solve up to three diaries (including itself) and some alterations to one diary will cause up to three diaries (including itself) to become recalcitrant. This provides more leverage for making principled revisions.

To see the counterexample to W, consider the diagram's details. After a series of four overlapping type-3 diaries (C through F), A judges s_2 at t_{13}. This is the last entry in diary G (completed at t_{14}), which is a recalcitrant type-1 diary, needing revision. At t_{14} (that is, looking only at diaries C through G), there is no principled way of deciding which of s_2 and s_3 to revoke within G. The subsequent completion of diary H is of no help either, since these two candidate corrections to G, between which A could not choose at t_{14}, are identical to the two candidates corrections to H (also type-1) between which there is still no choosing. The ratio of solved diaries to revised judgements would be 2:1 for each of s_2 and s_3. However, the completion at t_{18} of diary I (type-1 again) is helpful, since it is now the case that changing *one* token judgement of s_2 (t_{13}) would solve three diaries (G, H, and I) whereas a revision to s_3 could achieve such a reward only at the greater cost of changing *two* token judgements (t_{11} and t_{17}). Moreover, a provisional correction of s_2 would not be upset by the completion of diaries J and K, since these are not recalcitrant and thus cannot be solved (since they do not need solving) by a change to the t_{17} judgement of s_3.

On its face, this is a situation in which, guided by the ideal of simplicity, A has precisely what Wright thinks he could never have: a reason to revise his judgement S_2 (and the present-tense s_2), generated by the correlation $\{H_1, H_2\}$, despite the fact that S_2 lacks an OC diary. A similar example can be given for not-S_2. Therefore Wright's conditional, W, which makes an OC diary a necessary condition for a criterion of correctness, is false.

4. NEGLECTED THEORIES

My conclusion might seem premature, given Wright's claim that a theory will generate criteria of correctness for judgements about *any* of the sensation types it concerns only if it generates them for judgements about *all* those types. Overlapping diaries meant that we could generate *prima facie* criteria of correctness for more judgement types than could Wright, but since I doubt that we can use overlapping diaries to generate even *prima facie* criteria for s_1 and not-s_3, the rotten apple threat remains. Here, then, it is important to notice that Wright not only overlooks the use of overlapping diaries to enrich the putative linguist's data; he overlooks the range of theories that might be available to the linguist. Why, for example, might the linguist not consider theories concerning the temporal order of A's sensations?

Suppose, for instance, that instead of $\{H_1, H_2\}$ (a theory comprising material conditionals), A confirmed the following:

H$_3$: $P_1 \Rightarrow P_2$

H$_4$: $P_2 \Rightarrow P_3$

H$_5$: $P_3 \Rightarrow P_1$.

Read H$_3$ as "a sensation of type P_1 will be succeeded by a sensation of type P_2 before a sensation of another type", and read H$_4$ and H$_5$ similarly, *mutatis mutandis*. Now imagine that A records the following series of pre-theoretical judgements:

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}
s_1	s_2	s_3	s_1	s_2	s_3	s_1	s_1	s_3	s_1	s_2	s_3

As before, suppose that every time A judges that he is undergoing a sensation of one type, he simultaneously judges that he is not undergoing either of the other types. Now, having recorded this set of pre-theoretical judgements between t_1 and t_{12}, surely A could decide that, since $\{H_3, H_4, H_5\}$ is well confirmed, he must have been wrong in two of the three judgements he made at t_8: namely, both s_1 and not-s_2. And there are similar examples in which A makes principled revisions to tokens of the remaining four types of judgement.

Another case: Wright argues there are *no* theories about two sensation types that generate criteria of correctness for all of the judgements a subject might make (p. 248). But, once we enlarge the range of theories on offer, this seems false. Consider,

H$_6$: $P_1 \Rightarrow P_2$.

Having recorded the following series of judgements

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
s_1	s_2	s_1	s_2	s_1	s_1	s_1	s_2

A could reasonably conclude that both of his judgements at t_6 were incorrect: s_1 and not-s_2. And having recorded the following series of judgements

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
s_1	s_2	s_1	s_2	s_2	s_2	s_1	s_2

he could reasonably conclude that both of his judgements at t_5 were incorrect: s_2 and not-s_1. Wright has not explained why this theory and its more complex cousins fall short of generating criteria of correctness for all the judgements whose subject matter they concern.

———————————————

So Wright underestimates the potency of Blackburn's objection to Wittgenstein. If we concede that the aspirant private linguist might at least *attempt* to establish a criterion of correctness by theorising about his sensations, we cannot then defuse this concession's implications for the anti-privacy view by invoking Wright's meagre assessment of the linguist's odds of succeeding. For Wright's calculations are mistaken: he underestimates both the richness of the data and the range of theories that would be available to the linguist. Hence those of us who doubt the possibility of a private language must not allow the issue to come down to such a calculation of odds.[8]

David Bain

University of Nottingham

davidbain@fastmail.fm

References

Blackburn, S. (1984), "The Individual Strikes Back", *Synthese* 58, pp. 281-301.

Carruthers, P. (1986), *Introducing Persons* (London: Croome Helm)

Harrison, R. (1974), *On What There Must Be* (Oxford: Clarendon Press).

Moore, A. (2003), "On the Right Track", *Mind* 112, pp. 307-22

Walker, R. (1978), *Kant* (London: Routledge and Kegan Paul).

Wittgenstein, L (1953), *Philosophical Investigations*, translated by G.E.M. Anscombe. (Oxford: Basil Blackwell).

Wright, C. (1986), "Does *Philosophical Investigations* I. 258 Suggest a Cogent Argument Against Private Language?", in McDowell, J. and P. Pettit (eds.), *Subject, Thought, and Context* (Oxford: Oxford University Press), pp. 209-266.