Grazer Philosophische Studien
83 (2011), xx–yy.

THE PADEREWSKI PUZZLE AND THE PRINCIPLE OF SUBSTITUTION

Max Deutsch
University of ???

Summary
In "A Puzzle about Belief," Kripke maintains that his famous *Paderewski Puzzle* cannot be solved simply by rejecting the *Substitution Principle* for coreferential proper names. Kripke thinks this shows that the usual reason given for rejecting the Substitution Principle, namely that by rejecting it we can solve Frege's puzzle about belief, is not wholly convincing. Hence, according to Kripke, we should be less quick to give up the Substitution Principle and the *Strict Millian* semantics for names of which it is a consequence. While Kripke is right that the Paderewski Puzzle does not depend on the Substitution Principle, it does, I argue, depend on a different principle, *Weak Substitution*, which I claim only Strict Millians will be willing to accept. I conclude that the Paderewski Puzzle is genuinely puzzling only for Strict Millians.

1. *Introduction*

In "A Puzzle About Belief," Saul Kripke argues that there is a flaw in a familiar Fregean argument against the *Strict Millian*[1] principle that coreferential proper names are, except in quotational contexts, intersubstitutable everywhere preserving truth. The principle can be expressed as follows:

Substitution Principle:
If *a* and *b* are coreferential proper names in *L*, then any true sentence,

1. Strict Millianism is the doctrine that a name's referent is its meaning or 'semantic content' and that, because this is so, coreferential names can be substituted for one another in any non-quotational sentential context preserving truth. Millianism is simply the first conjunct of Strict Millianism: the doctrine that a name's referent is its sole semantic value. There are Millians who are not Strict Millians, but Strict Millianism is endorsed by Nathan Salmon (1986) and Scott Soames (1987) and more recently by David Braun (1998, 2006), Theodore Sider (1995, 2006), Bryan Frances (1998), and Michael Thau (2002).

> *S*, of *L*, containing *a*, can be turned into a true sentence, *S1*, of *L*, by replacing *a* with *b*, and any true sentence *S2* of *L*, containing *b*, can be turned into a true sentence, *S3*, of *L*, by replacing *b* with *a* (unless the context is quotational).

The Fregean argument that Kripke criticizes runs roughly like this. When *a* and *b* are coreferential proper names, the Substitution Principle appears to imply that an agent who assents to both '*a* is *F*' and '*b* is not *F*' has *contradictory* beliefs, and is thus not fully rational. But this is counterintuitive. It seems obvious that Lois Lane, for example, might assent to both 'Superman is brave' and 'Clark Kent is not brave' without harboring contradictory beliefs, and while remaining fully rational. So the Substitution Principle cannot be correct.[2]

According to Kripke, the flaw in this argument is that similar sorts of puzzles (wherein a seemingly rational agent has apparently contradictory beliefs ascribed to him or her) can be generated without assuming any substitution principle whatsoever. Furthermore, Kripke claims, these 'substitution-free' puzzles about names and belief require only assumptions that *everyone*—Fregeans and Strict Millians alike—will find unobjection-

2. I am following Kripke in describing the Fregean critique of the Substitution Principle in terms that make it appear as though the notion of *rationality* is crucially involved. But in fact there is an even more familiar and straightforward Fregean argument against the Substitution Principle that does not have much to do with rationality, an instance of which is just this: Lois Lane can perfectly well believe that Superman is brave *without* believing that Clark Kent is. But the Substitution Principle implies that this is impossible. So the Substitution Principle is false. Kripke discusses this argument also, in connection with what he calls the "strengthened disquotational principle," but his remarks are brief and unfortunately most commentators have focused on Kripke's handling of the rationality-involving argument described in the main text. I say "unfortunately" because I believe the Fregean argument presented in this note to be far more threatening than the one that Kripke and his commentators (including myself in this paper) have chosen to focus on. One need not be a Strict Millian to react to the rationality-involving argument by wondering why it is so implausible, after all, to suppose that one might end up with contradictory beliefs as a matter of simple bad luck. Perhaps this is what has happened to Lois Lane. The existence of the other, more threatening Fregean argument against the Substitution Principle does not affect the conclusions I will draw in this paper, for here I am concerned mainly with demonstrating that the rationality-involving argument, even if not the strongest anti-Millian argument on offer, cannot be evaded in the way Kripke attempts to evade it. One further point about the rationality-involving argument that I will note here and return to later in the main text: By itself, the Substitution Principle does *not* imply that the ascriptions '*S* believes that *a* is *F*' and '*S* believes that *b* is not *F*' are ascriptions of *contradictory* beliefs to *S*, not even given that the latter ascription implies, by the Substitution Principle, that the ascription, '*S* believes that *a* is not *F*' is true. Perhaps *a* is an ambiguous name.

able. Kripke concludes that the Fregean argument is a red-herring: Even if the Substitution Principle were false, we would still face these other, substitution-free puzzles, which, Kripke insists, arise from "sources of just the same kind" as the puzzle about Lois Lane (Kripke 1979, 133).[3] Therefore, the Substitution Principle can't be the puzzles' source, and rejecting it, via Fregeanism, or any other way, gets us nowhere.

Kripke spends the bulk of "A Puzzle about Belief" (henceforth, *PAB*) elaborating his substitution-free puzzles, of which there are two: the *Pierre Puzzle* and the *Paderewski Puzzle*. The Pierre Puzzle is the centerpiece of *PAB* and constitutes what Kripke takes to be his strongest defense of the Substitution Principle. In fact, Kripke introduces the Paderewski Puzzle merely to answer a potential objection to one of the assumptions required by the Pierre Puzzle. As is well known, the Pierre Puzzle depends on assumptions about *translation*. It depends on the general assumption that translation preserves truth, and it depends also on the particular assumption that the English sentence, 'Pierre believes that London is pretty,' is a translation of the French sentence, '*Pierre croit que Londres est jolie*.' Kripke introduces the Paderewski Puzzle to respond to an imagined objector who denies this latter assumption. Kripke describes this denial, rightly it seems to me, as a "desperate and implausible expedient," but goes on to point out that a puzzle, analogous in every crucial respect to the Pierre Puzzle, but which does not depend on any assumption about translation, can be formulated (128). This is the Paderewski Puzzle.

Although, as Kripke sees the matter, the function of the Paderewski Puzzle is simply to respond to an objection to an assumption required by the Pierre Puzzle, the Paderewski Puzzle is actually far more significant to the defense of the Substitution Principle than Kripke seems to realize. For one thing, the Paderewski Puzzle sidesteps a difficult issue concerning just how close the assumption that translation preserves truth comes to *entailing* the Substitution Principle.[4] If that assumption *does* entail the Substi-

3. All page references in the main text are to Kripke (1979) unless otherwise indicated.

4. Here is an argument that the claim that translation preserves truth *does* entail the Substitution Principle. The claim that translation preserves truth implies a more specific thesis concerning sentences containing proper names that can be called *Name Translation*:

Name Translation:

If a name, a, in L_1, is a translation of a name, b, in L_2, then for any true sentence of L_1 containing a, if a is replaced by b, while assuming a simultaneous translation into L_2 of the rest of the sentence, the result will be a true sentence of L_2.

But now suppose that all that is required, semantically speaking, for a name, a, in L_1, to qualify as a translation of a name, b, in L_2, is that a and b corefer. If so—if, that is, *coreference suffices for*

tution Principle, then the Pierre Puzzle won't show that the Substitution Principle is not responsible for the the "unpalatable conclusions" derivable from it, such as the conclusion about Lois Lane and her beliefs about Superman/Clark Kent (133). The assumptions required by the Paderewski Puzzle, however, quite clearly do not entail the Substitution Principle. In other words, the Pierre Puzzle perhaps does, while the Paderewski Puzzle definitely does *not*, depend on the Substitution Principle.

Kripke's Paderewski Puzzle represents an ingenious attempt to defend Strict Millianism and the Substitution Principle from the Fregean argument sketched above. Kripke is correct to claim that none of the Paderewski Puzzle's assumptions entail *the* Substitution Principle. And although he is wrong, as we shall see, to claim that no substitution principle of any kind is required by the Paderewski Puzzle, the substitution principle it does require is weaker, and seems, at least on its face, to be something that Strict Millians, Fregeans, and, indeed, anyone else, would likely agree to.

Ultimately, however, the defense fails. Certain sorts of examples, which I will call *Paderewski-ized Frege Cases*, appear to show that there is something objectionably Strict Millian about even this weaker principle. That is, the cases seem to show that even the weaker principle has consequences only a Strict Millian would be willing to accept. On the basis of cases such as these, I will argue that Kripke's best effort at defending Strict Millianism and the Substitution Principle does not succeed.

---

translatability in the case of names—then given that, say, 'Hesperus' and 'Phosphorus' corefer, and assigning English to both 'L_1' and 'L_2', we will get the result, via Name Translation, that 'Hesperus' and 'Phosphorus' can be interchanged in English sentences in which they occur, *salva veritate*. That's (an instance of) the Substitution Principle. *Does* coreference suffice for translatability in the case of names? The short answer is 'yes'. When translating a name, our translational practices demand only that we find a name in our language that corefers with the name in the language we are translating. For example, if I am trying to translate the French name, '*Londres*', into English, I've done all that I must if I find, in English, some name that corefers with '*Londres*'. Even if this argument shows that the Substitution Principle is a consequence of the assumption that translation preserves truth, is that not just what Kripke wants? As an anonymous referee (this journal) points out, if everyone, Strict Millians and Fregeans alike, are committed to the assumption that translation preserves truth, and that assumption entails the Substitution Principle, then Kripke will have succeeded, with the Pierre Puzzle, in showing that Strict Millianism is no worse off than other semantic theories of names. Firstly, however, Kripke's explicitly stated aim in presenting the Pierre Puzzle is to absolve the Substitution Principle, and this it does not do, if the argument just given is correct. Secondly, Fregeans are free simply to reject the assumption that translation preserves truth and maintain that the 'is a translation of' relation is weaker that the 'has the same Fregean sense' relation. Indeed, something along these lines is intuitively quite plausible in any case; we do not expect even good translations always to match their targets' meanings exactly.

There are similarities between the assessment of the Paderewski Puzzle I will present below and the assessment presented in David Sosa's (1996), "The Import of the Puzzle About Belief." In the penultimate section of the paper (section 5), I explain these similarities while emphasizing several important differences. Although we agree in broad outline, I will argue that Sosa gets several crucial points wrong.

2. *The Paderewski Puzzle*

Towards the end of *PAB*, Kripke introduces his paper's second puzzle about proper names and belief by imagining a character, Peter, who learns the name 'Paderewski' "with an identification of the person named as a famous pianist" (130). Kripke writes:

> Naturally, having learnt this, Peter will assent to "Paderewski had musical talent" and *we* can infer—using 'Paderewski', as we usually do, to name the Polish musician and statesman:
>
> (1) Peter believes that Paderewski had musical talent.
>
> Only the disquotational principle is necessary for our inference; no translation is required (130; italics in original).[5]

Inferring (1) is the first step in generating the Paderewski Puzzle. The "disquotational principle" is a principle linking sincere, reflective, and normal *assent* to a given sentence to *belief* in the proposition expressed by that sentence. In *PAB*, Kripke formulates the principle like this:

> *Disquotational Principle*:
> If a normal English speaker, on reflection, sincerely assents to 'p', then he believes that p (112f.).

The Disquotational Principle is required for our inference to (1), Kripke says, but "no translation is required" (130). Here, Kripke is alluding to his paper's first and more famous puzzle, namely the Pierre Puzzle, which does require assumptions about translation. For the reasons I described in the Introduction, the claim that "no translation is required" for generating the Paderewski Puzzle is very important. More important, however, is Kripke's claim that "*only* the disquotational principle is necessary for our inference"

5. I have renumbered the numbered belief ascription in this quotation. In *PAB* it is numbered '(8)'.

(130; italics added). The implication, of course, is that the Paderewski Puzzle does not appeal to, or depend on, the Substitution Principle.

Inferring (1) is only the first step in generating the Paderewski Puzzle. Here is how Kripke continues his story of Peter and Paderewski:

> Later, in a different circle, Peter learns of someone called 'Paderewski' who was a Polish nationalist leader and Prime Minister. Peter is skeptical of the musical abilities of politicians. He concludes that probably two people, approximate contemporaries no doubt, were both named 'Paderewski'. Using 'Paderewski' as a name for the statesman, Peter assents to 'Paderewski had no musical talent' (130).

Peter's assent to 'Paderewski had no musical talent' allows us to infer, via the Disquotational Principle only, that:

(2) Peter believes that Paderewski had no musical talent.

But now it looks as though Peter has contradictory beliefs. This seems an immediate consequence of the truth of (1) and (2). However, just as it does in Lois's case, the attribution of contradictory beliefs strikes us as unfair. Peter lacks information, not logical acumen. He believes falsely that 'Paderewski' names two different men, but we may, if we like, suppose him to be a leading logician, someone who "would *never* let contradictory beliefs pass" (122; italics in original).

So we have a case that mirrors Lois's case in important respects but does not in any obvious way depend on the Substitution Principle. The Substitution Principle is an inference principle. It validates the inference to '… *a* …' from '… *b* …', when *a* and *b* are coreferential proper names (and the context is not quotational). Nothing like this seems relevant to the Paderewski Puzzle. As Kripke says, the Disquotational Principle appears to be the only inference principle relied on in generating the Paderewski Puzzle.

3. *Weak Substitution*

Despite appearances, the Disquotational Principle is not the only principle required to generate the Paderewski Puzzle. The Paderewski Puzzle is not simply that (1) and (2) are both true; this we do get via the Disquotational Principle alone, but that (1) and (2) are both true is not, by itself, very

puzzling. Compare the claim that (1) and (2) are both true with the claim that (3) and (4) are both true:

(3) Peter believes that Aristotle was a shipping magnate.

(4) Peter believes that Aristotle was not a shipping magnate.

(3) and (4) don't, or at least need not, ascribe contradictory beliefs to Peter. 'Aristotle' is the name of Plato's most famous student, but it is also the name of a different man, one who famously married JFK's widow. Peter can consistently believe that one of these men was a shipping magnate, while believing that the other was not. In such a case, both (3) and (4) are true, but their joint truth does not imply contradictory beliefs on Peter's part.[6]

Under what circumstances must we say that (3) and (4) *do* impute contradictory beliefs to Peter? Of course, since they hold that coreferential names *mean the same thing*, Strict Millians will say that (3) and (4) ascribe contradictory beliefs to Peter just in case the 'Aristotle' in each refers to the same man. However, I mean to be asking the more general question of when *anyone*, independently of this or that semantic theory of names, must say that (3) and (4) ascribe contradictory beliefs. *If* we are Strict Millians we will give the answer just mentioned, but we will give that answer precisely because of our view of the equivalence between coreference and sameness of meaning in the case of names. In the case of a different category of singular terms—definite descriptions—Strict Millians say, along with the rest of us, that coreference does *not* suffice for sameness of meaning, and Peter can clearly have non-contradictory beliefs ascribed to him by ascriptions of the form, 'Peter believes that the *F* is *H*,' and 'Peter believes that the *G* is not-*H*,' even when the *F* is the *G*. These reflections reveal that the notion of *sameness of meaning* is the crucial notion in determining when a pair of belief ascriptions ascribe contradictory beliefs, and thus that the answer to the general question of when we must say (3) and (4) ascribe contradictory beliefs to Peter is that we must do so only if the 'Aristotle' in (3) *means the same thing as* the 'Aristotle' in (4).[7]

6. Sosa (1996) discusses a similar example (His case involves the name 'Paris' which refers to both the famous French city and a small town in Texas, USA.) and draws a similar, though not identical, moral. (See Sosa 1996, 386f..) For the difference in how Sosa and I understand the lesson of examples like these, see Section 5 of the main text.

7. Is the 'only if' claim too strong, strictly speaking? If we assume that the occurrences of 'Aristotle' in (3) and (4) not only corefer, but also rigidly designate their shared referent, then

The same applies *mutatis mutandis* to (1) and (2). (1) and (2) ascribe contradictory beliefs to Peter only if the 'Paderewski' in (1) means the same thing as the 'Paderewski' in (2). If the 'Paderewski' in (1) means the same thing as the 'Paderewski' in (2), then those 'Paderewski"s are intersubstitutable *salva significatione* (preserving meaning). If those 'Paderewski"s are intersubstitutable *salva significatione*, then they are intersubstitutable *salva veritate* (preserving truth). Thus, it turns out that the Paderewski Puzzle does depend on *a* substitution principle. Not the fully general Substitution Principle, which licenses the substitution of coreferential names in each and every case, but a weaker substitution principle that allows the substitution of coreferential names in some cases only.[8]

---

the content clauses of (3) and (4) will not be true together in any possible world. Furthermore, assuming the rigidity of the occurrences of 'Aristotle' need not require assuming they are synonymous. Of course, it would still be assuming something more than the mere coreference of the occurrences of 'Aristotle' in (3) and (4), and the main point of the example is that not everyone party to the debate over the semantics of names would or should agree that (3) and (4) ascribe contradictory beliefs simply when it is one and the same man referred to by 'Aristotle' in each ascription. My view, however, is that the 'only if' claim is not too strong; the 'Aristotle"s must be synonymous if (3) and (4) ascribe truly *contradictory* beliefs to Peter. 'The successor of 1' and 'the smallest prime number' codenote, and presumably do so rigidly. But if 'Peter believes that the successor of 1 is divisible by an even number and that the smallest prime number is not divisible by an even number' is true, then, while Peter believes something not possibly true, he does not believe something *contradictory*. Contradictory beliefs are a subcategory of impossibly true beliefs, ones that involve believing a proposition and that very proposition's negation. The assumption that the 'Aristotle' in (3) means the same thing as the 'Aristotle' in (4) is required for the conclusion that the content clause of (4) expresses the negation of what the content clause of (3) expresses, and that, in turn, is required for the conclusion that (3) and (4) ascribe contradictory beliefs. Coreference, or even coreference-plus-rigidity, of the occurrences of 'Aristotle' in (3) and (4) is insufficient for these conclusions, unless one question-beggingly assumes some Strict Millian thesis regarding names.

8. I assume here and throughout the paper that the relevant notion of meaning is one according to which meaning determines truth-value in the case of declarative sentences and extension or reference in the case of terms. If two *names* 'have the same meaning,' in the relevant sense, then they corefer. If two *sentences* 'have the same meaning,' in this sense, then they share a truth-value. An anonymous referee (this journal) suggests that adopting this notion of meaning is optional, and that someone armed with a different notion (perhaps a 'conceptual role' notion of meaning) might simply reject the transition from the premise that the 'Paderewski"s in (1) and (2) mean the same thing to any conclusion regarding the intersubstitutability of the occurrences of 'Paderewski' in those ascriptions. In defense of the assumption that the relevant notion of meaning is a 'determines reference/extension/truth-value' notion I would say, firstly, that this is fairly clearly the notion of meaning implicit in *PAB*. Kripke's paper falls firmly within the tradition now known as 'truth conditional semantics,' a tradition that takes this notion of meaning for granted. Secondly, it is difficult to see how other notions of meaning will deliver the puzzling result, namely that (1) and (2) ascribe *contradictory* beliefs to Peter. Intuitively, (1)

Kripke is impressed by the fact that inferring that (1) and (2) are true does not involve a step at which the Substitution Principle is *applied*. Inferring that (1) and (2) are true involves the application of only the Disquotational Principle. This is a genuine difference between cases like the case of Peter and Paderewski and the kinds of cases that spawned Frege's theory of sense and reference, such as the case of Lois Lane. When Lois assents to both 'Superman is brave' and 'Clark Kent is not brave' and we infer from this that (5) and (6) are true, we must then *apply* the Substitution Principle to get (7):

(5) Lois believes that Superman is brave.

(6) Lois believes that Clark Kent is not brave.

(7) Lois believes that Superman is not brave.

(5) and (7) look like they ascribe contradictory beliefs to Lois, but in the Paderewski Puzzle we get this appearance of contradictory beliefs—in the form of the joint truth of (1) and (2)—by applying the Disquotational Principle *only*; there is no substitution step as intermediary, as there is in the step from (6) to (7).

While Kripke is right that inferring the truth of the puzzling belief ascriptions, (1) and (2), does not involve applying the Substitution Principle, it seems to me that he overlooks the possibility that the Paderewski Puzzle depends on substitution in some way, but at a later stage, *after* the truth of (1) and (2) is established. For again, that (1) and (2) are both true is not very puzzling. It would be puzzling if (1) and (2) were not only true, but also ascribed contradictory beliefs to Peter. But, as we have seen, the claim that (1) and (2) ascribe contradictory beliefs to Peter depends on the claim that the 'Paderewski' in (1) means what the 'Paderewski' in (2) means, and this implies that *those* 'Paderewski"s, at least, are intersubstitutable *salva veritate*.

Kripke is absolutely right, however, that the sort of substitution involved in the Paderewski Puzzle is not the sort that needs legitimizing from the fully general Substitution Principle, and I was careful to use 'substitu-

and (2) ascribe contradictory beliefs only if the beliefs they ascribe are "about" one and the same man, namely Paderewski. But it will follow from the fact that the 'Paderewski"s in (1) and (2) mean the same thing that the beliefs they ascribe are about one and the same man only if there is some determinative connection between a name's meaning and its reference.

tion' with a lowercase 's' when claiming above that the Paderewski Puzzle "depends on substitution." It depends on *a* substitution step, but it does not depend on *the* Substitution Principle—not, anyway, in the sense of requiring the truth of the Substitution Principle. As I pointed out earlier, the Paderewski Puzzle seems to depend on only a far weaker principle, one that licenses the substitution of coreferential names in a quite limited range of cases.

Which cases? Well, the Paderewski Puzzle, remember, requires the 'Paderewski' in (1) to mean the same thing as the 'Paderewski' in (2). Generalizing a bit, the operative principle might be something like the following *Meaning Principle* for coreferential tokens of a single, one-referent name:

Meaning Principle:
For a name, *n*, which has a single referent in *L*, different tokens of *n* in sentences of *L* have the same meaning in *L*.

This Meaning Principle presupposed by the Paderewski Puzzle implies the following *Weak Substitution* principle:

Weak Substitution:
For a name, *n*, which has a single referent in *L*, if n_1 and n_2 are different tokens of *n* in sentence-tokens of *L*, then n_1 may be put for n_2, and vice versa, preserving truth (even if the context is quotational).

Weak Substitution is quite weak. Unlike the far more wide-ranging Substitution Principle, Weak Substitution does not seem to be something that is merely an artifact of a semantic theory that we regard as optional anyway, such as Strict Millianism. Weak Substitution says that coreferential tokens of a *single* name, having a *single* referent, may be interchanged preserving truth. Before considering cases such as the case of Peter and Paderewski, this principle surely would have struck us as true. We wouldn't have supposed that a token of a single name having a single referent could differ in any semantic respect from other tokens of that very same name. The Paderewski Puzzle forces us to be suspicious about Weak Substitution, but that we must be suspicious of such an innocuous seeming principle is quite surprising.

Of course, this is precisely the surprise Kripke wants us to have. As was well known long before the publication of *PAB*, puzzles about names

and belief can be formulated by assuming the fully general Substitution Principle. But only the Strict Millians—a radical fringe group—accept the fully general Substitution Principle. The rest of us clear thinking moderates are happy to say that the Substitution Principle is false. It has thus seemed safe to suppose that less radical theories of naming and belief that do not entail the Substitution Principle might solve these puzzles about names and belief for us. *PAB* and the Paderewski Puzzle seem to cast severe doubt on this comforting thought. Kripke seems right when he claims there are puzzles about names and belief, e.g. the Paderewski Puzzle, which cannot be solved simply by 'going Fregean' and rejecting the Substitution Principle. Solving these deeper puzzles about names and belief appears to require giving up on something more basic and non-theoretical.

My view, however, is that although Weak Substitution appears, at first blush, to be something theorists of all stripes would or should endorse, there are substitutions licensed by the principle that are very clearly substitutions *only a Strict Millian would accept as truth-preserving*. I call the cases in which such substitutions are allowed by Weak Substitution *Paderewski-ized Frege Cases*. A Paderewski-ized Frege Case is similar to the case of Peter and Paderewski, in that it involves tokens of a single name, with a single referent, and requires assuming only Weak Substitution, not the fully general Substitution Principle. But a Paderewski-ized Frege Case is also similar to a case involving two coreferential names, such as the Lois Lane case, in that it seems intuitively plain, in a Paderewski-ized Frege Case, that the name-tokens used in characterizing the relevant agent's beliefs are *not* intersubstitutable preserving truth. The Paderewski-ized Frege Cases reveal that some of the substitutions licensed by Weak Substitution are just as objectionable as the substitution of 'Superman' for 'Clark Kent' within the scope of Lois Lane's beliefs. I present a Paderewski-ized Frege Case in the next section.

4. *A Paderewski-ized Frege Case*

Suppose that, as a joke, and because he seems so unlike the heroic and brave Superman, Clark Kent is teasingly called 'Superman' by his coworkers, including Lois Lane. They don't for a moment suppose that Clark really is the caped hero of Metropolis; they call Clark 'Superman' just to make fun of him. In the newsroom of *The Daily Planet*, Lois might sincerely and reflectively assent to 'Superman is not brave' intending to agree to

something concerning her coworker, Clark Kent. Later, arriving at the scene of yet another of Superman's exploits, Lois might assent to 'Superman is brave' without in any ordinary sense changing her mind about her earlier assent to 'Superman is not brave'. So we have the makings of a Paderewski Puzzle: A rational agent, Lois, who, given the Disquotational Principle and, it seems, *only* the Disquotational Principle, appears to have contradictory beliefs.

I said that Lois might *sincerely* and reflectively assent to 'Superman is not brave,' even while recognizing that the sentence embodies a joke concerning her coworker, Clark Kent. One might doubt this; if she is joking, then Lois is not being fully sincere, and if she is not being fully sincere, then the Disquotational Principle, which requires sincere assent, cannot be straightforwardly applied.[9] If my daughter impatiently and impolitely demands her lunch and I teasingly reply, 'The ruler of the universe wants her lunch,' it is not clearly correct to ascribe to me the belief that the ruler of the universe wants her lunch. So there is a genuine question about whether the Disquotational Principle can be properly applied to Lois's assent to 'Superman is not brave.' Her assent indicates that she thinks, de re, *of* her mild-mannered coworker that he is not brave, because she is being at least partially sincere, and using 'Superman' to make 'speaker's reference' to Clark. But the running joke behind such uses of 'Superman' perhaps makes it doubtful that her assent indicates that she thinks, de dicto, *that* Superman is not brave.

However, suppose we fill in a few more background details. Suppose Clark has been called 'Superman' for many, many years by his coworkers at *The Daily Planet*. They began doing so as a tease, but now the practice has become entrenched, and the original jest behind it largely forgotten. In the newsroom, 'Superman' was originally used as a teasing nickname for Clark, but is now used as simply a nickname for him. Imagine Lois assenting to 'Superman is not brave' in these circumstances, intending, as before, to use 'Superman' to denote her coworker, Clark Kent, but *not* actively intending to be making a joke by doing so. With these background details in place, the inclination to say that the Disquotational Principle cannot be applied to Lois's assent is far less strong it seems to me. The practice of using 'Superman' to refer to a certain mild-mannered reporter for *The Daily Planet*, once sufficiently entrenched, has, in effect, and so far as Clark's coworkers are concerned, christened him with a new name. Similarly, if my

9. I am grateful to an anonymous referee (this journal) for raising this objection.

daughter persists with her impatient and impolite demands, the practice of using 'the ruler of the universe' to refer to her might become entrenched, and the description might 'grow capitals,' becoming a nickname. In that event, my assent to a sentence such as 'The Ruler of the Universe wants her lunch' may become straightforwardly disquotable.

A virtue of extending the original Superman stories in this way, by adding a bit about Clark coming to be called 'Superman' as a joke, is that this is a *plausible* extension of the original stories. Unfortunately, the fact that the relevant uses of 'Superman' are jokes, or at least begin their lives as jokes, raises the worry about the applicability of the Diquotational Principle, and thus raises the question of whether the extended story gives rise to a Paderewski-type puzzle. My view is that, at least in the case in which we imagine that the uses of 'Superman' to refer to Clark have become entrenched, there is little reason to think that Lois's assent to 'Superman is not brave' cannot be disquoted. However, even if I am wrong about this, by sacrificing a little plausibility, different extensions of the original stories, which seem to eliminate completely the worry about the Disquotational Principle's applicability, are possible.

For example, suppose, perhaps as a sly joke of his own, that Clark presents himself to *The Daily Planet* as being *named* 'Superman Kent.' He signs his W-2 and I-9 forms with it, produces forged documentary 'proof' that 'Superman Kent' is indeed his name, and tells his coworkers some not completely unbelievable story about his parents' motives in so naming him. Though she finds it a bit odd, Clark's coworker, Lois Lane, accepts that 'Superman Kent' is his birth name and uses it accordingly. In particular, she assents sincerely, without meaning any sort of joke, to a variety of sentences with 'Superman' in subject position, taking 'Superman' to refer, in some of these sentences, not to the caped hero of Metropolis, but to her mild-mannered coworker who wears a Brooks Brothers suit. One day, Lois, exasperated by Clark's constant fearfulness, sincerely, reflectively, and without joking in any way, assents to 'Superman is not brave,' taking the occurrence of 'Superman' in this sentence to refer to her mild-mannered coworker. Later, positively evaluating one of Superman's heroic deeds, but without changing her mind regarding her earlier assent to 'Superman is not brave,' she assents to 'Superman is brave,' taking the occurrence of 'Superman' in this sentence to refer to the man of steel. It seems to me that the Disquotational Principle is straightforwardly applicable to each of Lois's assents. But by applying the principle to both, we seem forced to conclude that a rational agent, Lois, possesses contradictory beliefs. Only

the Disquotational Principle is required for this inference, so it appears that this extension of the Superman stories gives rise to a Paderewski-type puzzle. To have a label, I will refer to the case derivable from this most recent extension of the Superman stories as the 'modified Lois case.'

Notice that the puzzling conclusion of the modified Lois case is identical to one of the puzzling conclusions that can be derived from the actual Superman stories plus the Substitution Principle, namely that (5) and (7) from above are both true:

(5) Lois Lane believes that Superman is brave.

(7) Lois Lane believes that Superman is not brave.

For in the actual stories, (5) and (6) are certainly both true, and (7), as we saw earlier, may be obtained from (6) via the Substitution Principle:

(6) Lois Lane believes that Clark Kent is not brave.

Now, a step like the step from (6) to (7) is missing in my modified Lois case, that's true. As in the original Paderewski Puzzle, in the modified Lois case, the Substitution Principle is not *applied* in inferring the truth of the relevant belief ascriptions. Only the Disquotational Principle is used in that connection. Nonetheless, if we conclude that (5) and (7) are not just true in the circumstances described, but also that they ascribe contradictory beliefs to Lois, then we must maintain that the 'Superman' in (5) means the same thing as, and so is intersubstitutable with, the 'Superman' in (7). The trouble, of course, is that those 'Superman"s are *not* intersubstitutable *salva veritate*. Or rather (less dogmatically), we are no more inclined, with respect to the modified Lois case, to say that those 'Superman"s are intersubstitutable, than we are to say that the names 'Clark Kent' and 'Superman' are intersubstitutable.

The modified Lois case depends only on Weak Substitution and involves different tokens of one, single-referent-possessing name, 'Superman', not two different coreferential names. But the substitution allowed by Weak Substitution in the modified Lois case strikes us as no more plausible than allowing two different coreferential names, such as 'Superman' and 'Clark Kent', to be interchanged. In other words, the modified Lois case is a Paderewski-ized Frege Case; it is a case that reveals the unacceptably Strict Millian aspect of even so weak a substitution principle as Weak Substitution.

Part of the point of calling the modified Lois case a Paderewski-ized Frege Case is to emphasize the fact that it *is* a Frege Case. That is, it is not, in any important respect, different from the two-name cases discussed in connection with Frege's theory of sense and reference. Whatever inclines us to deny that (5) and (7), as uttered in connection with the modified Lois case, ascribe contradictory beliefs to Lois, it is this very same thing that inclines us to deny that (8) and (9) necessarily ascribe contradictory beliefs to the ancient astronomer:

(8) The ancient astronomer believed that Hesperus is bright.

(9) The ancient astronomer believed that Phosphorus is not bright.

Kripke would perhaps agree, but point out that the conclusion that Lois, in the modified case, has contradictory beliefs does not require assuming the fully general Substitution Principle. We need only assume the Meaning Principle and the far weaker Weak Substitution principle that it implies. I think that, although this is true, it is far less significant than Kripke perhaps supposes. Reflection on the modified Lois case shows us that these weaker principles were to blame all along, and, furthermore, that any theory committed to even these weaker principles is just as unacceptable as a theory committed to the stronger Substitution Principle. The weaker principles allow for a more limited range of substitutions, but some of the substitutions they allow are just as counterintutitive as the counterintuitive substitutions allowed by the Substitution Principle.

The other reason for calling the modified Lois case a Paderewski-ized Frege Case is to highlight the fact that it is the same kind of case as Kripke's case of Peter and Paderewski. Like Kripke's case, the modified Lois case involves a one-referent name of a single language, and inferring the puzzling belief ascriptions that comprise it involves applying only the Disquotational Principle. Even if the fully general Substitution Principle were false, we would still be faced by the 'puzzle' presented by the modified Lois case. As with the original Peter/Paderewski case, to generate the modified Lois case, the only substitution principle required to be true is the far weaker Weak Substitution principle. Of course, no one in their right mind would think that one could use the modified Lois case to defend Strict Millianism or the Substitution Principle. The case is just too similar to an ordinary Frege case to be used for that purpose. What this shows, I think, is that no case of this kind, and so not even the original Peter/

Paderewski case, can be used to defend Strict Millianism and the Substitution Principle. Those who believe that the original Peter/Paderewski case can be used in that way owe us an explanation of how the original case differs in some significant respect from the modified Lois case.[10] I myself don't believe that such an explanation is forthcoming.[11]

5. *Comparison to Sosa 1996*

The assessment of the Paderewski Puzzle presented in the previous sections bears certain structural similarities to that of Sosa 1996. Sosa and I both agree (with Kripke) that the puzzles in *PAB*, including the Paderewski Puzzle, do not depend on the Substitution Principle. And we both maintain (against Kripke) that the puzzles depend on a principle which only Strict Millians would accept, and thus that the puzzles do not necessarily arise for other semantic views. Beyond this, there are important dissimilarities between the account in Sosa 1996 and the account given above.

Sosa absolves substitution entirely and regards Kripke as having shown, via his puzzles, that where and when Strict Millianism licenses the substitution of coreferential names is not the root issue. I disagree. We get the Paderewski Puzzle only if we accept Weak Substitution. Weak Substitution is weaker than the Substitution Principle by being less general, but Weak

10. Some readers may still wonder how elaborating another Paderewski-like puzzle (namely, the modified Lois case) can reveal that the original Paderewski Puzzle does not serve to indirectly defend Strict Millianism and the Substitution Principle. My view is that Kripke's presentation of the original Paderewski Puzzle obscures the fact that the original puzzle depends on Weak Substitution and thus encourages us to diagnose the puzzle in a way that places no blame on Strict Millianism. By starting with an ordinary Frege case and turning it into a Paderewski Case, as I have done with the modified Lois case, it is far easier to see that cases of this general type are not fit for defending Strict Millianism.

11. There are also 'Frege-ized' Paderewski Cases. Imagine Kripke's story of Peter and Paderewski with this small alteration: Whenever Peter wants to say something about, as he would put it, Paderewski-the-statesman, he pronounces 'Paderewski' as 'Pade-ROO-ski'. But whenever he wants to say something about Paderewski-the-musician, he pronounces 'Paderewski' as 'Pade-REV-ski'. In the story thus altered, we can derive puzzling conclusions about Peter's beliefs only if we take his 'Pade-ROO-ski"s to mean the same thing as, and so be interchangeable with, his 'Pade-REV-ski"s. The case becomes essentially the same as a two-name case of the sort that spawned Frege's theory of sense and reference. Since, with very minor alterations, one can turn a Paderewski Case into a Frege Case, and a Frege Case into a Paderewski Case, I conclude that the existence of Paderewski Cases could serve to indirectly defend Strict Millianism and the Substitution Principle only if the existence of Frege Cases serves to indirectly defend those two claims. But it is clear that the existence of Frege Cases does nothing of the kind!

Substitution licenses certain substitutions of coreferential names. So where and when substitutions of coreferential names are licensed is still the root issue, according to me. Sosa thinks Kripke has advanced the discussion by showing at least that the immediate cause of our puzzles about names and belief has been misdiagnosed. I do not think Kripke has succeeded in showing even this.

According to Sosa, the puzzle-inducing Strict Millian principle is something Sosa labels the Hermeneutic Principle:

> **(H)ermeneutic**: If a name in ordinary language has a single referent then it may correctly be represented logically by a single constant (Sosa 1996, 388).

It is **(H)**, and not any principle to do with substitution, which, by Sosa's lights, is both necessary for generating Kripke's puzzles, and such that it would command assent only from those already convinced by Strict Millianism. Sosa intends **(H)** to play the role that the Meaning Principle plays in my rendition of the Paderewski Puzzle. It is supposed to move us from the truth of (1) and (2), from above, to the conclusion that Peter has contradictory beliefs.[12]

The problem with **(H)** is that it does not do its job. From **(H)** and the truth of (1) and (2) we can conclude only that the name, 'Paderewski,' in (1) and (2) should be represented logically using a single constant. How is this supposed to guarantee that Peter has contradictory beliefs? Sosa claims that, "for any formal constant φ, 'Fφ and ~Fφ' is a contradiction" (Sosa 1996, 388). In fact, however, it is not in the nature of a formal constant that for any such constant φ, 'Fφ and ~Fφ' expresses a contradiction. Whether this is true will depend the particular formal system in question and the stipulations in force with respect to it. Indeed, Alonzo Church (1951) has claimed on a variety of grounds that logical constants ought to be assigned senses![13] So Sosa is mistaken in thinking that **(H)** bridges the gap between the truth of (1) and (2) and the truth of the claim that the beliefs ascribed to Peter by (1) and (2) are contradictory. Even if the logical forms of (1) and (2) are represented by using a single constant in

12. See Sosa's rendition of the Paderewski Puzzle (1996, 380).

13. In Church (1951), constants are said to require not just values, but "sense-values" as well, partly in order to formally render the semantic difference between 'a = a' and 'a = b.' In Church (1982), it is even suggested that variables and their natural language counterparts, pronouns, must be assigned senses in order to avoid the George IV paradox discussed therein. In any case, there is nothing intrinsically Millian about a logical constant.

place of 'Paderewski,' the question of whether the beliefs so represented are contradictory remains open.

To bridge the gap, something like the Meaning Principle described in Section 3 is needed. The Meaning Principle licenses certain substitutions that should not be licensed; it implies Weak Substitution. These two principles are the real culprits in generating the Paderewski Puzzle—**(H)** is irrelevant. As we saw above, it is consistent with the existence of the Paderewski Puzzle that the fully general Substitution Principle is false. But there is no Paderewski Puzzle if Weak Substitution is not true. Hence, as has long been thought, it is precisely illegitimate substitutions of coreferential names that is the root cause of various puzzles about names and belief, from Frege's Puzzle to Kripke's.

The existence of Paderewski-ized Frege Cases reinforces this thought. As the existence of such cases shows, there is no substantial difference between Frege's Puzzle and the Paderewski Puzzle. In his rush to accept that Kripke's Puzzles, though not independent of Strict Millianism, are genuinely new puzzles about names and belief, Sosa misses this underlying similarity between the two sorts of puzzle and ends up blaming a principle, **(H)**, which is neither a consequence of Strict Millianism nor even necessary for generating Kripke's puzzles. The assessment presented in the previous sections corrects these mistakes.

6. Conclusion

The Paderewski Puzzle appears to be an important part of the Strict Millian arsenal. It can be generated without assuming the truth of the Substitution Principle, but the belief ascriptions that comprise it strike us as just as puzzling as those inferable when we assume the Substitution Principle. This suggests, as Kripke says in *PAB*, that neither the Substitution Principle nor Strict Millianism is responsible for the 'substitution troubles' about names and belief that Frege's theory of sense and reference was meant to solve for us.

The suggestion is misleading, however. The Paderewski-ized Frege Cases show that our intuitive resistance to the Substitution Principle extends even to the weaker principles required by the Paderewski Puzzle, namely the Meaning Principle and Weak Substitution. Strict Millians are committed to these principles; for the rest of us, they are optional. This means, I think, that the Paderewski Puzzle is really only a *puzzle* for the Strict

Millian. The Paderewski Puzzle cannot be used as an indirect defense of the Substitution Principle or Strict Millianism; in fact, the puzzle arises only on assumptions that everyone but a Strict Millian is free to reject.

REFERENCES

Braun, David 1998: "Understanding Belief Reports". *Philosophical Review* 107, 555–595.

Church, Alonzo 1951: "The Need for Abstract Entities in Semantic Analysis". *American Academy of Arts and Sciences Proceedings* 80, 100–113.

— 1982: "A Remark Concerning Quine's Paradox About Modality". In: Nathan Salmon & Scott Soames (eds.) 1988: *Propositions and Attitudes*. New York: Oxford University Press, 58–65.

Frances, Bryan 1998: "Defending Millian Theories". Mind 107, 703–727.

Kripke, Saul 1979: "A Puzzle About Belief". In: Nathan Salmon & Scott Soames (eds.) 1988: *Propositions and Attitudes*. New York: Oxford University Press, 102–148.

— 1980: *Naming and Necessity*. Cambridge, MA.: Harvard University Press.

Lewis, David 1981: "What Puzzling Pierre Does not Believe". *Australasian Journal of Philosophy* 59, 283–289.

Loar, Brian 1987: "Names in Thought". *Philosophical Studies* 51, 169–185.

Salmon, Nathan 1986: Frege's Puzzle. Cambridge, MA.: MIT Press.

Salmon, Nathan & Soames, Scott (eds.) 1988: *Propositions and Attitudes*. New York: Oxford University Press.

Sider, Theodore 1995: "Three Problems for Richard's Theory of Belief Ascription". *Canadian Journal of Philosophy* 25, 487–514.

Sider, Theodore & Braun, David 2006: "Kripke's Revenge". *Philosophical Studies* 128, 669–682.

Soames, Scott 1987: "Direct Reference, Propositional Attitudes, and Semantic Content". In: Nathan Salmon & Scott Soames (eds.) 1988: *Propositions and Attitudes*. New York: Oxford University Press, 197–239.

Sosa, David 1996: "The Import of the Puzzle About Belief". *Philosophical Review* 105, 373–402.

Thau, Michael 2002: *Consciousness and Cognition*. New York: Oxford University Press.