

A PARADOX FOR POSSIBLE WORLDS SEMANTICS[†]

MICHAEL J. SHAFFER AND JEREMY MORRIS

Abstract

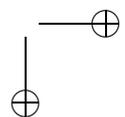
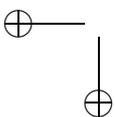
The development of possible worlds semantics for modal claims has led to a more general application of that theory as a complete semantics for various formal and natural languages, and this view is widely held to be an adequate (philosophical) interpretation of the model theory for such languages. We argue here that this view generates a self-referential inconsistency that indicates either the falsity or the incompleteness of PWS.

1. *Introduction*

The development of possible worlds semantics for modal claims has led to a more general application of that theory as a complete semantics for various formal and natural languages, and this view is widely held to be an adequate philosophical interpretation of the model theory for such languages. In its most basic form, possible world semantics (PWS) posits that the meaning of any (declarative) sentence is a set of possible worlds at which it is true.¹ More accurately, possible worlds semantics holds that the meanings of all well-formed declarative sentences in a given language L_i are defined by the set of all possible worlds at which a given sentence P is true relative to a model. Formally, where Ψ_{wwf} are the well-formed declarative sentences of language L_i , $P \in \Psi_{wwf}$, $\{w_i, w_j, \dots\}$ is the set of possible worlds at which P is true and \mathcal{W} is the set of all possible worlds such that for all w , $w \in \mathcal{W}$:

[†]The authors would like to thank Hal Brown, Keith Lehrer, Ken Williford and (especially) an anonymous referee for valuable comments.

¹See, for example, Cresswell 1988, 1994, Hintikka 1969, Stalnaker 1976, Lewis 1970, Lycan 1994 and Montague 1974 for explication and defense of PWS.



(PWS) The meaning, $|P|$, of any P of any $L_i = \{w_i, w_j, \dots\}$.²

We argue here that this generates a self-referential inconsistency that indicates the falsity or the incompleteness of PWS.

2. The Paradox

The paradox of PWS is generated as follows. PWS is itself a declarative sentence expressible in, say, English. If PWS is true, then the meaning of ‘PWS’ is the set of possible worlds at which it is true. Also, as a matter of logic, PWS must either be contingently true or necessarily true. If PWS is necessarily true, then according to the standard interpretation of necessity in terms of possible worlds (i.e., according to PWS), PWS is true at all possible worlds. Formally, where $w \in \mathcal{W}$, P is necessarily true iff $(\forall w)(w \models P)$. If PWS is contingently true, then, in terms of the standard interpretation of contingency, PWS is false at at least one possible world and true at at least one possible world. Formally, where w_i and $w_j \in \mathcal{W}$, P is contingently true iff $(\exists w_i)(w_i \models P) \ \& \ (\exists w_j)\neg(w_j \models P)$. If PWS is meaningful in terms of PWS and necessarily true, then \neg PWS must, as a matter of logic, be logically impossible. However, \neg PWS is not logically impossible. Therefore, if PWS is meaningful in terms of PWS, then it must not be necessarily true. So, if PWS is true, it must be contingently true. However, if PWS is meaningful in terms of PWS and contingently true, then, by the standard possible worlds analysis of contingency, PWS is false at one or more worlds. If, however, PWS is false at one or more worlds, then PWS is *false* as the meanings of sentences of a given language at those worlds are not the sets of possible worlds at which they are true. Therefore, possible world semantics must either be incomplete in the sense that it does not apply self-referentially or it is simply false.³

Having sketched out this problem above and in order to be more precise, we offer the following formal, but simple, rendering of the paradox:

²We assume that the relevantly important quantifiers in this expression are intended to range over all possible worlds.

³Our criticism in this respect is related to worries raised about Rosen 1990 in Rosen 1993.

P1.	PWS	Assumption
P2.	$\Box PWS \vee \neg \Box PWS$	Modal Tautology
P3.	$\Box PWS \supset \neg \Diamond \neg PWS$	Modal Equivalence
P4.	$\neg \Box PWS \supset \Diamond \neg PWS$	Modal Equivalence
P5.	$\Diamond \neg PWS \vee \neg \Diamond \neg PWS$	Dilemma
P6.	$\Diamond \neg PWS \supset \neg PWS$	Premise
P7.	$\Diamond \neg PWS$	Premise
\therefore	$\neg PWS$	Propositional Logic [P6–P7] ⁴

Of course, the defense of the claim that PWS is paradoxically self-defeating then depends on our establishing both P6 and P7 and establishing that they are true is what we aim to do in what follows. If it turns out that these two premises are defensible, then it looks as if PWS must be false or incomplete. If it is false or incomplete, we suggest that it should be rejected. In spite of this result, however, we will not advocate a specific replacement for PWS here.

3. Why PWS is not Necessarily True

The first obvious avenue of attack open to those who might wish to reject our criticism of PWS is to challenge P7 and to, thereby, simply assert that PWS really is necessarily true. This would allow the defenders of PWS to claim that PWS is not self-defeating and that it is true in all possible worlds. In other words, that $(\forall w)(w \models PWS)$ is true. We claim that this is false and that the contingency of PWS can easily be seen by reference to those semantic theories that are competitors with respect to PWS.⁵

⁴ Of course, it should be apparent that our real argument only requires P6 and P7. However, P1–P5 are included for the purpose of clarity.

⁵ An anonymous referee noted that we are correct with respect to this point and (very helpfully) added that PWS is typically understood to be “true only if it accords with the way in which speakers understand synonymy, truth, entailment and so on.” We, of course, agree, but Keith Lehrer has also suggested to us that PWS must be contingent on the basis of the following argument. That there are consistent alternatives to PWS can be observed from the fact that perfectly consistent semantic theories differ over the truth-values assigned to certain statements. For instance the statement, ‘all true sentences are true’ may be true according to one theory and ill-formed according to another. For some semantic theories entail that such statements are ill-formed, since they predicate truth of themselves or at their own semantic level. Other no less consistent theories allow the predication of truth in ‘all true statements are true’ so that the statement turns out true. Since neither alternative is inconsistent, there is at least one consistent alternative to PWS. For a complete PWS must come down one way or another with respect to such statements and, whatever way that is, the above considerations show that there will be a consistent alternative. Lehrer makes an essentially similar point to this in his 2000 (see pg. 28) and a related point is made in McGee 1990.

There are a number of viable semantic theories with which PWS competes and they include, but are not limited to, the theories (or versions of theories) that assert that (1) meanings are analytic roles in conceptual systems⁶; or that (2) meanings are atomic mental representations⁷; or that (3) meanings are social conventions⁸; or that (4) meanings are propositions⁹; or that (5) meanings are truth conditions.¹⁰ The problem for defenders of PWS who wish to avoid the paradox presented here is that in defending the claim that PWS *is* a necessary truth they are logically committed to the view that all those semantic theories with which PWS competes must be contradictory, i.e. logically impossible. But consider alternative (1), that of conceptual role semantics (CRS). In its linguistic form this semantic theory asserts that the meaning of a sentence is to be identified with the role that a sentence plays in the language to which it belongs. Where 'R(*x*, *y*)' is understood as the role that a sentence *x* plays in a language *y* CRS can then be understood as follows:

(CRS) The meaning, $|P|$, of any P of any $L_i = R(P, L_i)$.

CRS then appears to be a perfectly coherent and easily formulated theory about the nature of meaning, whatever one's opinions about its correctness. In any case, given CRS the following argument can then be generated:

- | | | |
|--------------|---|------------------------|
| Q1. | $\Box PWS \supset (\forall w)(w \models PWS)$ | Modal Definition |
| Q2. | $(\forall w)(w \models PWS) \supset \neg(\exists w)(w \models CRS)$ | Modal Semantics Truth |
| Q3. | $\neg(\exists w)(w \models CRS) \supset \Box \neg CRS$ | Modal Definition |
| Q4. | $\neg \Box \neg CRS$ | Premise |
| Q5. | $(\exists w)(w \models CRS)$ | Modus Tollens [Q3, Q4] |
| Q6. | $\neg(\forall w)(w \models PWS)$ | Modus Tollens [Q2, Q5] |
| \therefore | $\neg \Box PWS$ | Modus Tollens [Q1, Q6] |

Of course, this argument can be easily generalized for other cases of semantic theories with which PWS competes, and so what we have to say with respect to CRS will apply to a number of semantic theories that are incompatible with PWS. What is most pertinent here is that it is obvious that many

⁶ See Harman 1982.

⁷ See Fodor 1998 and Jackendoff 1983.

⁸ See Lewis 1969 and Skyrms 1996.

⁹ See Schiffer 2003.

¹⁰ Davidson 1967.

semantic theories with which PWS competes are not contradictory. It is true that CRS might, for example, be false, but it is not *logically false* or it is at very least not obviously logically false. Such theories just do not seem to be logically impossible, and if the defenders of PWS who wish to avoid the paradox raised here were still to claim that they must be, then they would be obligated to show explicitly how they are contradictory. They have neither done this, nor is it plausible to suppose that, for example, CRS could be shown to be contradictory. It seems that one can easily and consistently understand a possible world at which CRS is true and one can also consistently imagine a world at which it is false, and one can do this for many other semantic theories in addition to those noted above. So the burden of proof lies with the defender of PWS when it comes to the issues of the necessity of PWS and of the impossibility of its competitors. Absent some compelling argument to the effect that all competing semantic theories are contradictory and that it is not possible to imagine a possible world where PWS is false, PWS *cannot* reasonably then be taken to be a necessary truth, if necessary truth is understood in terms of standard possible worlds semantics.¹¹

4. PWS and the Object-language/Meta-language Distinction

A perhaps more promising way to attack the paradox of PWS would be to deny P6 on the basis of the object-language/meta-language distinction and thus argue that the contingency of PWS does not result in self-refutation. Things are a bit complicated in this case, but this approach also ultimately results in an inevitable failure to avoid the paradox. First, let us look at bit more closely at P6. P6 asserts that if it is possible that PWS is false, then PWS is false. P6 is true because the relevant quantifiers in PWS are interpreted as unrestricted. As a result, regardless of whatever language PWS is stated in, the meaning of PWS must be the set of possible worlds at which it is true. However, if PWS is contingent, then there are worlds at which it is false, i.e., there are worlds where the meanings of sentences are *not* defined by the set of possible worlds where they are true. Thus, given its contingency, PWS cannot define all the meanings within any given language on pain of contradiction, i.e., the theory expressed by PWS is either inconsistent or incomplete. Either way, PWS is false as stated.

Avoiding the paradox by appeal to the object-language/meta-language distinction might predictably begin as follows. Either PWS is a sentence of the very language it is intended to apply to or it is a sentence in a meta-language

¹¹ The issue of the connection between conceivability and possibility is, of course, a thorny one as many of the essays in Gendler and Hawthorne 2002 indicate.

used to express the semantics of some object language. If it is a sentence in the very language it is intended to apply to, then, following Tarski, it cannot express a truth concerning the semantics of that language. Suppose then that PWS is a theory of the semantics of various languages, but is itself a sentence in some meta-language. Following Tarski, such meta-languages can be used to express the truth conditional semantics of object language sentences provided the meta-language in which PWS is expressed is sufficiently more expressive than the object languages under analysis. However, if PWS is a sentence in such a meta-language of order \mathcal{L}_3 , then if it is meaningful it must either be meaningful in virtue of PWS or in virtue of some other semantic theory. If it is meaningful in virtue of some other semantic theory, then PWS is false as there is at least one sentence in a language that is meaningful but not in terms of PWS. If PWS in the meta-language of order \mathcal{L}_2 is meaningful in virtue of PWS, then it must be so in virtue of some higher-level meta-language of order \mathcal{L}_3 that contains a translated statement of the \mathcal{L}_2 -order language expression of PWS. A vicious regress looms large, but a far deeper problem becomes apparent as well.

If the semantics for various languages at various levels in the hierarchy of meta-languages is given in terms of possible worlds, then the domain over which the expressions of those various languages are semantically interpreted is permanently fixed. There is only the total domain of possibilities, \mathcal{D} , and the various possible relations, \mathcal{R} , defined on \mathcal{D} and nothing else can possibly exist at any world. As such, the domain of possible meanings for the sentences of all languages, regardless of position in the hierarchy of meta-languages, is fixed by the combinatorial contents of those sets $\mathcal{D} \cup \mathcal{R}$. and so must be shared by all languages. So $\mathcal{D} \cup \mathcal{R}$ delimits the totality of possible meanings, independent of any particular language of any particular order in the hierarchy of languages and there are no more meanings that can be expressed in any language than those which can be expressed in terms of the combinatorial content of $\mathcal{D} \cup \mathcal{R}$. This implies that, if the truths of semantics are to be meaningful in terms of PWS and PWS is complete, then the meanings of those expressions must be such that they can be expressed in terms of $\mathcal{D} \cup \mathcal{R}$. However, it does not appear to be the case that this can be done without generating inconsistency.

Consider what the defender of PWS might say about PWS in light of the object-language/meta-language distinction. The first thing that comes to mind is that one might object to our formulation of PWS as involving unrestricted quantification over languages. What might then be suggested is that PWS must be relativized to different orders, \mathcal{L}_n , of language. So, PWS for the lowest order languages \mathcal{L}_1 , PWS_1 , would need to be a sentence of a language at linguistic order \mathcal{L}_2 and PWS_1 would apply only to sentences of languages of order \mathcal{L}_1 ; it would allow us to express sentences in an \mathcal{L}_2 -order language that concern which worlds constructed from $\mathcal{D} \cup \mathcal{R}$ the various

sentences of \mathcal{L}_1 -order languages are true at. PWS_2 would then be a sentence of a language of order \mathcal{L}_3 and would allow us to express the semantics for sentences of \mathcal{L}_2 -order languages, including sentences expressing PWS_1 itself. Of course this is where the undesirable regress of semantic theories appears, but there is a more serious problem for defenders of PWS as well. Specifically, such a reformulation of PWS is still self-refuting.

Suppose, however, that, in accord with this suggestion and to avoid the regress, we both relativize and generalize PWS as follows:

(PWS_n) The meaning, $|P|$, of any P of any L_i of order $\mathcal{L}_{n-1} = \{w_i, w_j, \dots\}$.

So, the meaning of sentences of any language of a given order is a set of possible worlds and this statement is expressible in languages of any order greater than that of the object language we might be considering. Nevertheless, PWS_n must still be contingent for the reasons given in the previous section and it must be expressed in some language of some order or other. Regardless of what language it is expressed in if it is contingent then it is not true for all sentences of all languages of any given order and so must be false. It is simply not true that the meaning of any P of any L_i of any order \mathcal{L}_{n-1} is the set of worlds constructed from $\mathcal{D} \cup \mathcal{R}$ such that $(\{w_i, w_j, \dots\} \models P)$ and this must be the case if PWS_n is only contingently true. As a result, PWS_n is self-defeating and false.

Defenders of PWS might then object to the generality of PWS_n , and suggest that PWS is really a set of theories, one for each order of language, \mathcal{L}_n . Thus the quantifiers might be further restricted so as to avoid the problem of self-refutation. But, consider PWS_1 , where PWS_1 is expressed in a language of order \mathcal{L}_2 , and let us express it formally as follows:

(PWS₁) The meaning, $|P|$, of any P of any L_i of order $\mathcal{L}_1 = \{w_i, w_j, \dots\}$.

Again, this statement is contingent as demonstrated in the previous section. As a result, it turns out that it is no less self-refuting than either PWS or PWS_n . PWS_1 contingently asserts that the meanings of sentences of \mathcal{L}_1 -order languages are sets of possible worlds constructed from $\mathcal{D} \cup \mathcal{R}$, but, as it does so contingently, it is false. It is not always true that the meanings of sentences are sets of possible worlds, even if we restrict the quantifiers in PWS_1 to \mathcal{L}_1 -order sentences because each PWS_i is contingent. The same problem will arise for whatever specific PWS_i we consider, and each possible world semantics theory, PWS_i , is itself paradoxically self-refuting. As a result, for all i , PWS_i is false.

5. *Two-Dimensional PWS*

Another apparently promising way in which one might attempt to avoid the paradox of PWS involves modifying PWS in the manner that a number of variously motivated philosophers have suggested.¹² The basic idea on which this maneuver might be based is that PWS could be accepted as necessarily true while restricting its application to appropriate contexts of utterance so that PWS will be necessarily true *relative to those contexts*. By adopting a "two-dimensional" view that allows for this type of context relative necessity one might hope to avoid the paradox by denying P6 or P7. We have already seen that this strategy is unsuccessful in the case of traditional PWS and we shall see that it fails in the case of two-dimensional PWS as well. In any case, a generic rendering of this more sophisticated version of PWS would take the following form:

(PWS_{2D}) The meaning, $|P|$, of any P of any L_i in a context of utterance $C = \{w_i, w_j, \dots\}$.

C is then understood to fix certain important semantic features relative to the utterance of P such as the referents of terms. How contexts are specifically to be characterized is a matter of contention among those who have advocated two-dimensional PWS.¹³ However, we argue that PWS_{2D} does not avoid the paradox on any of the characterizations of contexts of utterance.

Since we have already argued that the paradox endangers PWS on the unrelativized version of necessity and possibility, what the defender of the two-dimensional view must show is that either P6 or P7 is false given the relativized notion of that concept incorporated in PWS. Yet there is no reason to suppose that P6 is false given this more sophisticated, double-indexed, version of PWS. Even if the possibility of the falsity of PWS_{2D} is relativized to a given context, it is still the case that the meanings of some sentences are not the set of possible worlds at which they are true. Thus, if it is possible that PWS_{2D} is false, then PWS_{2D} is false. Given that there is some context such that there is some world at which PWS_{2D} is false, then the meanings of the sentences of the language at that world in that context are not the set of possible worlds at which they are true. That this must be the case can be illustrated

¹² See, for example, Humberstone and Davies 1980, Humberstone 2004, Davies 2004, Stalnaker 1978 and Kaplan 1989.

¹³ On Stalnaker's (1978) influential account contexts are themselves equated with sets of possible worlds. More specifically, they are understood to be the possible worlds compatible with what is presupposed in a given conversational situation.

easily by assuming Stalnaker's concept of context for the sake of the argument. On Stalnaker's view contexts are sets of possible worlds that represent the set of possible worlds compatible with what is assumed, or presupposed, in a conversational situation. It is clear then that there must some worlds and contexts such that PWS_{2D} is not only not assumed, but where PWS_{2D} is also false. Since this is the case PWS_{2D} is false because it will not always be the case that the meaning, $|P|$, of any P of any L_i in a context of utterance $C = \{w_i, w_j, \dots\}$ and this will be due to the contingency of PWS_{2D} . As noted earlier, this result generalizes easily to versions of PWS_{2D} other than Stalnaker's. More importantly, neither does two-dimensional PWS provide the grounds for denying P7. For PWS_{2D} is not necessarily true even relative to a context insofar as there are consistent alternatives to PWS_{2D} that concern the same context at a world. Consider a given utterance of P in a specified context C in world w_i . According to PWS_{2D} , the meaning of P at w_i in that context is the set of worlds at which P is true. Yet, as the initial argument against the necessity of PWS establishes, neither CRS nor the other contemporary alternatives to PWS are logically impossible analyses of the meaning of P in that same context at w_i . Thus it is possible relative to some context that PWS_{2D} is false at w_i .¹⁴ So the PWS_{2D} version of PWS could only be successful if one were to reject the core view of PWS, i.e. that *the* meaning of sentences are sets of worlds. In any case, given the above argument for P6, the paradox runs unhindered as before. Moreover, no particular view of the nature of contexts has been presupposed so that the above argument applies generally to all versions of two-dimensional PWS.

6. Reflections and Conclusions

PWS is paradoxically self-refuting, even in its sophisticated forms, and defenders of PWS must accept that it is, at very least, incomplete and in need of serious reformulation or that it is just false. Accepting the former charge, that of incompleteness, would seem to at least require replacing the universal quantifiers in PWS with existential quantifiers and then those who wish to defend PWS face, at least, the following *prima facie* problems. First, if PWS is contingent and it is true only at some possible worlds, then how does one tell which ones it is true at? Second, is PWS true of the actual world? Third, if PWS is incomplete, then what semantic theory explains the meaning of PWS itself? Finally, as PWS is incomplete because it is contingent, can it somehow be formulated without being self-defeating? As we think

¹⁴Recall also the argument suggested by Lehrer present in fn. 5. Shifting from PWS to PWS_{2D} also does not avoid the conclusion of that argument for the contingency of possible world semantics.

that a negative answer *must* be given in response to this final question, a more promising suggestion, we believe, is simply to reject PWS because it is false and to replace it with one of the alternatives noted above, or some other self-consistent semantic theory.

Department of Philosophy
St. Cloud State University
123 Brown Hall
720 Fourth Avenue South
St. Cloud, MN 56301-4498
U.S.A.

E-mail: shaffermphil@hotmail.com

Department of Philosophy
The University of Miami
P.O. Box 248054
Coral Gables, FL 33124-4670
U.S.A.

E-mail: j.morris5@umiami.edu

REFERENCES

- Cresswell, M.J. 1988. *Semantical Essays*. Dordrecht: Kluwer.
Cresswell, M.J. 1994. *Language in the World*. Cambridge: Cambridge University Press.
Davidson, D. 1967. 'Truth and Meaning,' *Synthese* 17: 304–323.
Davies, M. 2004. 'Reference, Contingency, and the Two-Dimensional Framework,' *Philosophical Studies* 118: 83–131.
Davies, M. and L. Humberstone 1980. 'Two Notions of Necessity,' *Philosophical Studies* 38: 1–30.
Fodor, J. 1998. *Concepts*. Oxford: Oxford University Press.
Gendler, T. and J. Hawthorne 2002. *Conceivability and Possibility*. Oxford: Oxford University Press.
Harman, G. 1982. 'Conceptual Role Semantics,' *Notre Dame Journal of Formal Logic* 23: 242–256.
Hintikka, J. 1969. *Models for Modalities: Selected Essays*. Dordrecht: Reidel.
Humberstone, L. 2004. 'Two-Dimensional Adventures,' *Philosophical Studies* 118: 17–65.
Jackendoff, R. 1983. *Semantics and Cognition*. Cambridge: MIT Press.
Kaplan, D. 1989. 'Demonstratives,' in J. Almog, et al. (eds.), *Themes from Kaplan*. Oxford: Oxford University Press.

- Lehrer, K. 2000. *Theory of Knowledge*, 2nd ed. Boulder: Westview Press.
- Lewis, D. 1969. *Convention*. Cambridge: Harvard University Press.
- Lewis, D. 1970. 'General Semantics,' *Synthese* 22: 18–67.
- Lycan, W. 1994. *Modality and Meaning*. Dordrecht: Kluwer.
- McGee, V. 1990. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Cambridge: Hackett.
- Montague, R. 1974. *Formal Philosophy*. New Haven: Yale University Press.
- Rosen, G. 1990. 'Modal Fictionalism,' *Mind* 9: 327–354.
- Rosen, G. 1993. 'A Problem for Fictionalism about Possible Worlds,' *Analysis* 53: 71–81.
- Schiffer, S. 2003. *The Things we Mean*. Oxford: Oxford University Press.
- Skyrms, B. 1996. *Evolution of the Social Contract*. Cambridge: Cambridge University Press.
- Stalnaker, R. 1976. 'Propositions,' in A. MacKay and D. Merrill (eds.), *Issues in the Philosophy of Language*. New Haven: Yale University Press.
- Stalnaker, R. 1978. 'Assertion,' in P. Cole (ed.), *Syntax and Semantics* 9. New York: Academic Press.