To Want It is to Want to Be There
A Note on Levinson (2003)

Daniel Büring
UCLA
buring@humnet.ucla.edu

Levinson (2003) proposes an ingenious novel analysis of verbs of desire, in particular the verb *want* within a possible world semantics. He proposes that one wants φ in w with respect to g if one's subjective expectation of the g-desirability of φ is higher than that of one's subjective expectation of the g-desirability of not-g. Effectively, the 'wanter' calculates: Among all possibilities w, is the desirability of those in which π, *weighted relative to their likelihood* greater than the desirability of those in which not-φ, relative to *their* likelihood?

In this note, I want to question whether this is indeed the *semantics* of wanting. I want to explore the possibility that to want φ in w indeed just means — to a first approximation — to prefer realistic φ worlds to realistic non-φ worlds, as proposed in Heim (1992). What Levinson models is very likely what one does when forming a desire, but not what it means to want something; or so I will suggest.

In what follows, I'll exclusively discuss Levinson's argument for a probabilistic semantics of *want*, and question his arguments in its favor. The alternative I use (and favor) will be Heim's proposal, but no evaluation of Levinson's other modifications, in particular the introduction of *evaluation functions*, is intended by this. These are simply outside of the scope of this note.

1 Heim (1992)'s Semantics for *want*

The semantics for *want* given in Heim (1992) essentially say that for someone, say Al, to want φ, Al must prefer worlds in which φ is the case to worlds in which it is not. More precisely, those worlds in which φ that are maximally similar to a doxastic alternative of Al must be preferable to those in which not-φ that are maximally similar to a doxastic alternative of Al. To formalize this, Heim introduces a preference relation $<_{a,w}$ which holds between two worlds or two propositions just in case individual a finds the second more desirable than the first in w:

(1) a. for any $w, w', w'' \in W, w' <_{a,w} w''$ iff w'' is more desirable to a in w than w'.
   b. for any $w \in W, X \subseteq W, Y \subseteq W, X <_{a,w} Y$ iff for all $w' \in X, w'' \in Y, w' <_{a,w} w''$.

She furthermore introduces the notion of *comparative similarity*, which selects those worlds from a proposition that are maximally similar to a chosen world w:

(2) $Sim_w(\psi) = \{w' \in W \mid w \in \psi \text{ and } w' \text{ resembles } w \text{ no less than any other world in } \psi\}$

The use of the similarity function will be that one doesn't compare just any ψ-world to any non-ψ-world to see if ψ is desirable, but only those (non-)ψ-worlds which are similar to worlds one actually considers possible; it should be noted that if w in (2) is itself a ψ-world, $Sim_w(\psi) = \{w\}$. With these ancillary definitions at hand, Heim gives the semantics of *want* as in (3), where $Dox_{Al}(w)$ is the set of Al's *doxastic alternatives* in w, i.e. those worlds which, for all Al assumes in w, could be the real world:

(3) $w \in [\![\text{Al wants } \psi]\!]$ iff for every $w' \in Dox_{Al}(w), Sim_{w'}(\psi) <_{Al,w} Sim_{w'}(\neg\psi)$

In words, for each world w' which Al thinks could be the real world, Al must prefer (in w) those worlds maximally similar to w' in which ψ is the case to any world maximally similar to w' in which ψ is not the case; see Heim (1992) and Levinson (2003) for an appraisal of this formulation.

2   The Insurance Scenario

In response to Heim's semantics for *want* as given in the previous section, Levinson (2003) constructs a number of ingenious scenarios — which I will call 'better-safe-than-sorry scenarios' — in which he argues Heim's semantics fails. The one discussed most extensively is the insurance scenario. Al is a homeowner, and like any responsible homeowner wants to buy home insurance, i.e. an insurance that pays if any damage to the house, e.g. through fire, occurs. She is a reasonable person, so she is fully aware that it is possible, in fact likely, that nothing will happen and the insurance is superfluous, but nonetheless, like most homeowners, she wants to buy insurance.

In accordance with Levinson, I will assume that in this case it is indeed true and felicitous to say of our homeowner Al that she wants to buy insurance.

What does Heim's treatment say about these cases? Levinson asks us to consider four types of worlds.
Al's doxastic alternatives come in two basic flavors: Those in which the house will be damaged, and those in which it won't (we assume of course that Al has no knowledge of what is going to happen, and that there is no connection between buying an insurance and suffering damage — *contra* Murphy's law). Accessible φ-worlds from the former kind look like w_4, accessible non-φ-worlds like w_2. Clearly, she will prefer any w_4-type world to w_2-type worlds, the worst case scenario (no insurance, house ruined); this is good for wanting φ, since w_4 is a φ-world, and w_2 isn't.

But in those doxastic alternatives in which the house won't be ruined, φ-alternatives look like w_3, while non-φ-alternatives look like w_1. Reasonable Al should prefer w_1 to w_3, since in w_1, she doesn't spend money on an unneeded insurance, while in w_3 she does.

World w_1, however is a non-φ-world, while w_3 (as well as w_4) is a φ-world. So she prefers a non-φ-world over a φ-world as well, contrary to (3). According to Heim's proposal, then, our homeowner neither wants to buy insurance, nor does she want to be insurance-less. But this conclusion contradict our premise that she does want to buy insurance.

What goes wrong here, Levinson claims, is that the relation between w_2 and w_4 — the one that intuitively accounts for a wanting to buy insurance — and the relation between w_1 and w_3 — the one that would make you want to stay uninsured — are treated the same. Intuitively, the dis-preference of w_2-type-worlds over w_4-type worlds should far outweigh the preference of w_1 type worlds over w_3-type worlds. Levinson calls this the *utility value* of these world types. For illustration, we can equate the utility value with the costs incurred in this case: The home insurance costs perhaps one thousand dollars a year, rebuilding the house costs a hundred times that. So the utility of not having to spend \$ 100,000 (w_4 to w_2) outweighs the utility of saving the insurance premium of \$ 1,000 (w_1 to w_3).[1]

So if instead of separating worlds into desirable and undesirable ones, you were to add up the utility values (here: the money saved) among the φ-worlds, and then among the non-φ worlds, each w_2-type world would incur a negative utility of \$ 100,000, while each w_3-type word only adds a few thousand, so that over all, the net utility value of the φ-worlds (of which w_3 is one) is much higher than that of the non-φ-worlds (to which w_2 belongs).

Accordingly, Levinson gives a semantics which, among other things, does just that: It adds up the utilities of the φ-worlds, and compares it to the combined utility values of the non-φ-worlds. If the former is higher than the latter, you want φ, which in the insurance scenario is indeed the case.

To me, the intuitive beauty of Levinson's proposal resides in the fact that this seems indeed what I do when I decide whether or not I want to do something which has negative side effects. I evaluate 'how bad' the side effects would be (as well as how likely they are to occur, another aspect Levinson elegantly builds into his analysis), and compare that to how bad it would be if I just didn't do that something at all. If the latter is worse than the former, I'll decide to do it, and hence want to do it (or want that I do it, an aspect of syntax of *want* I ignore here). And on the face of it, Heim's analysis seems incapable of capturing a desire like that.

3 What's Desirable

Let us try to grab the other end of the dilemma, though, and ask: Is w_1 really more desirable than w_3? Is it really more desirable to have no insurance and no damage, than to have insurance and no damage? The common sense answer seems to be 'yes of course', but that answer, I believe, is with hindsight. If I could choose between four courses of events, one in which I pay nothing and loose nothing, one in which I pay and loose nothing, one in which I pay something and loose a little, and one in which I pay nothing and loose a lot, there's no question which I choose, and that is, I believe, the gist of Levinson's argument.

---

[1]In actuality, utility value can of course be more sophisticated, for example that damage to the house will force the owner to sell or go bankrupt etc.

But our homeowner isn't in that comfortable position. Indeed she knows that even if she happens to live in a no-damage world, she will not know that she does. Now the question changes shape: Is it really more desirable to pay no insurance and live in constant fear of loosing ones house once and for all, than to pay a little and gain peace of mind? Now, I think, the answer depends: It is not if you are a gambler; in which case you probably won't buy insurance. But if you're a safety-first, average kind of home owner, you'll probably happily pay and know that you won't be ruined, even if the house burns down.

The crux lies with what makes a world desirable. The position I explore here is that a world w' is not desirable to Al in w if Al's house won't burn down, but only if Al were to *know in* w' that it won't. To put it perhaps a little clearer, Al should find w' more desirable than w'' only if Al would prefer *to be Al (or Al's counterpart) in* w' *over being Al (Al's counterpart) in* w''. To want φ to be the case is to want to be oneself in a φ-world, not just to abstractly determine that a φ-world is preferable.

This is the heart of my comment. It is one question how I form a desire. Levinson's algorithm gives an extremely plausible and elegant formalization of how such formation of desires comes about. It is a different question what it means to have a desire, and I suspect that to want something means indeed more or less what Heim thought it did. I content that there is no need to build the rational behavior of determining what is desirable into the semantics of verbs of desire like *want*. Potential counter-examples such as the insurance scenario only seem so pervasive only of one equates a world being desirable to A with a world being desirable given Al's objectives. But, I have argued, a world would only be desirable to Al if Al would prefer to *be* Al (or Al's counterpart) in that world, which includes not knowing the nature of that world. Once we take this into consideration, I believe the Heimian semantics for verbs like *want* does not run into problems with worst-case-scenarios like the insurance scenario.

References

Heim, Irene. 1992. Presupposition Projection and the Semantics of Attitude Verbs. JOS 9:183–221.

Levinson, Dmitry. 2003. Probabilistic Model-theoretic Semantics for *want*. paper presented at Semantics and Linguistic Theory 13, Seattle; ms. Stanford University .