

Probabilistic Model-theoretic Semantics for *want*.

Dmitry Levinson
Tel-Aviv University

1. Introduction

Desire verbs usually receive little attention in the semantic literature on propositional attitude verbs. They are generally seen as a part of a larger group of the attitude verbs (e.g., Hintikka 1969), hence receiving a partial treatment. There has been some mention of properties special to desire verbs, but the need for an “independently motivated analysis” noted by Heim (1992) still exists. This paper proposes such an analysis, addressing both the problems already mentioned in earlier literature and new evidence that is incompatible with previous analyses.

The structure of this paper is as follows. Hintikka’s (1969) analysis for propositional attitudes, and other “ideal worlds” analyses of *want* based on it, are described in Section 3, and counterexamples are given in the same section. Section 4 describes the approach of Geurts (1998) and Villalta (2000). Heim’s (1992) semantics is described in Section 5. A probabilistic counterexample to previous analyses is given in Section 6, my proposal is described in Section 7, and some additional issues are discussed in Section 8.

2. Kinds of desire: partial vs. motivational (“all-things-considered”)

Before describing the various analyses of desire, I would like to make a distinction between partial and motivational desires. Consider the following argument outlined by Davis (1986, p. 77): “Chris came by my office to ask me to play tennis. ‘I really want to play’, I said, ‘but I have to teach’. A short time later Mark asked me whether I wanted to play tennis, and I replied, ‘No, I have to teach’”. The speaker once says he wants to play tennis and afterwards says he doesn’t. Did he change his mind? Probably not, but rather he wants to play in some particular way and doesn’t want it in some other way.

Let us take a closer look at this example. After being asked whether he wanted to play tennis, the speaker once answers

(1) I really want to play, but I have to teach.

and then

(2) No, I have to teach.

The desires expressed in (1) and (2) are of different kinds. In (1), it is the basic desire of the speaker to somehow play tennis, his positive attitude to the actions of the type “playing tennis”. He doesn’t consider all the consequences of

the possible actions. Some possible cases of playing tennis may have unwanted consequences, but this has nothing with the desire expressed in (1). It judges any kind of playing tennis as desirable, as long as it is playing tennis. Such desires, considering only a specific aspect of an action, I will call *partial* desires. As we can see, such a desire does not necessarily lead to action. If it did, one would play tennis any time one has a smallest desire to. As Anscombe (1963, p. 61) puts it, it would be ‘insanity’ to think so. It is important to note that a person can have different partial desires towards the same action, assessing it from different aspects.

The desire expressed in (2) is different. Here the action of playing tennis is not considered just from the aspect of playing tennis. The possible consequences of the action are also brought into consideration. Playing tennis would prevent the speaker from teaching, and not teaching would lead to undesirable consequences, so the action of playing tennis as a whole is undesirable. The kind of desire that considers an action as a whole can be called *motivational desire*, and this is the kind of desire accompanying intentional action. According to Davidson (1978), such a desire felt before acting, an “all-things-considered” judgement, *is* intending.

The distinction between these two kinds of desire is common in philosophical literature. For example, Locke (1982) distinguished between the *genuine* (partial, in my terms) and *formal* (motivational) senses of desire.

3. “Ideal worlds” analyses

3.1. Hintikka (1969): introducing the “ideal worlds” approach

Hintikka (1969) proposed a general analysis for propositional attitude verbs, which is known as a possible worlds analysis. It does not distinguish between the desire verbs such as *want* and *wish* and the epistemic attitude verbs such as *believe* and *know*. Each propositional attitude is represented by an accessibility relation between the possible worlds. The world w' is accessible to w with respect to the epistemic accessibility relation for an individual d (denoted as E_d) iff w' is consistent with d 's beliefs in w . The sentence

(3) John believes that it is raining.

is true in a world w iff it is raining in every world w' that is epistemically accessible to John in w .

The same analysis can be applied to desire verbs, of course, with a *buletic* accessibility relation, expressing desire instead of belief. The world w' would be accessible to a in w with respect to the buletic accessibility relation (denoted as B_d) iff w' conforms to everything d desires in w . The sentence,

(4) John wants to come to the party.

is true in a world w iff John comes to the party in every world w' that is accessible for John in w , with the buletic accessibility relation.

Formally,

$$(5) \quad [[d \text{ wants } \varphi]]_w = 1 \text{ iff } B_d(w) \subseteq \varphi$$

Should there be a different analysis for believing and wanting? Hintikka explicitly states some propositional attitudes for which he thinks his analysis is correct: “knowledge, belief, memory, perception, *hope*, *wish*, *striving*, *desire*” (p. 25, italics mine), that is, according to him, there should be no difference between these verbs. However, there is an important distinction between the first four verbs and the last four. The first four are what Stalnaker (1984, p.79) calls *acceptance concepts*. For him, accepting a proposition is treating it as a true proposition in some way. As a rough criterion, a propositional attitude is an acceptance concept if the attitude is said to be *correct* whenever the proposition is true. If P is true, then knowledge that P is correct, but hope that P is not. Stalnaker argues further (p. 82) that some conditions implied by Hintikka’s model hold for acceptance concepts, while they don’t hold for wanting, hence limiting the applicability of the model to such concepts only.

Hintikka’s analysis is a basic form of what I would call an “*ideal worlds*” approach. I will use this name for any approach in which ‘ d wants p in w ’ means $I(d,w) \subseteq p$ where I is some function independent of p .

3.2. *Kratzer (1981) and von Fintel (1999): continuing the “ideal worlds” approach*

Kratzer (1981) proposes an analysis belonging to this group. According to this analysis, what d wants in w defines an *ordering source* $g(w)$, and the truth conditions for ‘ d wants p ’ are that p is true in every world in $g(w)$. This is consistent with Hintikka’s truth conditions, since Kratzer’s ordering source $g(w)$ actually corresponds to Hintikka’s set of buletically accessible worlds $B_d(w)$. The difference is that Kratzer specifies how it is calculated from other data.

Kratzer’s analysis is basically adopted by von Fintel (1999), with a small change: instead of Kratzer’s “what d wants”, in von Fintel’s analysis $I(d,w)$ is “the worlds that maximally correspond to d ’s preferences in w ”. This is augmented by adding Heim’s (1992) suggestion that “ d wants φ ” presupposes that both φ and $\neg\varphi$ are possible, but the final result belongs to the “ideal worlds” kind.

The “ideal worlds” analyses have common properties: according to them, Upward Entailment and Conjunction principles hold (not allowing contradicting desires). The fact that these principles do not hold is a major problem for these analyses.

3.3. *Upward Entailment*

According to “ideal worlds” analyses, *Upward Entailment* holds for *want*:

- (6) If d wants ϕ in w and ϕ entails ψ , then d wants ψ in w .

Indeed, if d wants ϕ , then ϕ holds in all the worlds in $I(d,w)$, then ψ also holds in these worlds, and a wants ψ . However, there's a number of examples suggesting that Upward Entailment does not always hold.

First, if Upward Entailment does hold, every proposition which is physically or logically necessary, or just believed to be true, is also wanted. Suppose that John lives in Europe and wants to visit Africa by taking a southbound ship. Be ψ = "the continent of Europe lies to the north of Africa". It is known to John that ψ , so visiting Africa in the actual world implies ψ , and ψ holds in every w' in $B_{John}(w)$. However, saying in this context that

- (7) John wants Europe to lie to the north of Africa.

seems very strange. The same argument holds for every proposition believed to be true in our world.

An extreme example of this kind is a *tautology*. Since every proposition p entails every tautology, it is implied by (6) that if d wants anything, he also wants every tautology. This can be derived directly from the definition of "ideal worlds" analyses, since every tautology holds in every world contained in $I(d,w)$. However, the sentence

- (8) John wants 2+2 to be 4.

seems meaningless. It seems impossible to want something which cannot conceivably be false.

A possible solution to this problem, proposed by Heim (1992) and adopted by von Stechow (1999), is to say that a *wants* ϕ presupposes that both ϕ and $\neg\phi$ are possible for a , or, in other words, that a neither believes that ϕ nor believes that $\neg\phi$. However, nothing inherent in "ideal worlds" approach suggests it should be so, and it is an *ad hoc* stipulation. As will be shown in Section 8.1, the existence of this presupposition is predicted by the analysis that will be proposed in this paper.

Another type of arguments against UE have been presented by Stalnaker (1984), Asher (1987), and Heim (1992). Consider the following example adapted from Asher (1987, p. 171):

- (9) Nicholas wants to get a free trip on the Concorde.

- (10) Nicholas wants to get a trip on the Concorde.

Heim (1992) analyzes this example as follows. Obviously, taking a free trip on the Concorde implies taking a trip on the Concorde. Now the question is whether (9) implies (10) or not. If Nicholas believes that the flight is too expensive, it may very well be that he would like to fly on the Concorde if he

could get the trip for free, but he is not willing to pay the price. In this case (9) is true, but (10) is false.

Kai von Fintel (1999) claims that arguments of this type involve a change of context, so (10) is true if getting a free trip is possible, and that Upward Entailment does hold for *want*, for a constant context. It is important to notice that desire reports in these examples concern the motivational desire. If the partial desire is considered, obviously Nicholas does want to get the Concorde trip.

I propose another counterexample for UE, which satisfies von Fintel's requirement of constant context:

Suppose you, or some other individual d , are in a situation in which you will play either game A (outcome A) or game B (outcome B), with equal probabilities. In game B the player receives \$200 unconditionally. In game A there are two possible outcomes: A1, with probability 10%, in which case the player receives \$300 and A2, with probability 90%, in which case the player receives nothing. Figure 1 illustrates the conditions of the game.

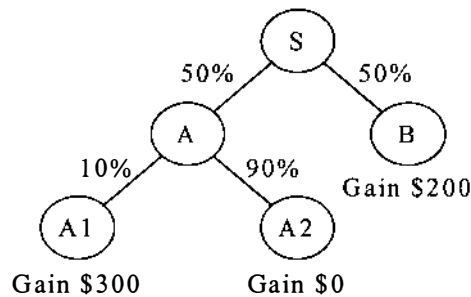


Figure 1. Game conditions

In order to reach A1, a player must reach A, so “ d reached A1” entails “ d reached A”. However, the following can be true of a reasonable person d :

- (11) d wants to reach A1
- (12) d doesn't want to reach A

The fact that (11) and (12) can be both true in the same situation, with the same context, shows that Upward Entailment does not hold for *want*.

3.4. Conjunction Introduction

Another consequence of the “ideal worlds” analyses is that according to them, *Conjunction Introduction* holds:

- (13) If a wants ϕ in w , and also a wants ψ in w , then a wants $\phi \wedge \psi$ in w

Indeed, if a wants ϕ in w , and also a wants ψ in w , then, according to “ideal worlds” approaches, both ϕ and ψ hold in every world in $I(a,w)$. This means that $\phi \wedge \psi$ holds in every world in $I(a,w)$, and a wants $\phi \wedge \psi$ in w . However,

the following example suggests that Conjunction Introduction does not necessarily hold.

Suppose John is thinking of his next summer vacation. He has never been to Paris and he thinks it would be nice to go to Paris this summer. In this case it can be said that

(14) John would like to visit Paris this summer.

However, he has never been to Rome, too, and he wants to visit it, too. In this case it can be true that

(15) John would like to visit Rome this summer.

If conjunction introduction holds, (14) and (15) entail (16):

(16) John would like to visit both Paris and Rome this summer.

This is not necessarily true, for example, if he does not have enough time or money to visit both cities. We may conclude that conjunction introduction doesn't always hold.

3.5. *Contradicting desires*

I will call two desires contradicting if they cannot be fulfilled together (within the model). “Ideal worlds” analyses rule out the possibility of such desires:

(17) $((a \text{ wants } \varphi) \text{ and } (\varphi \Rightarrow \neg\psi)) \Rightarrow \neg(a \text{ wants } \psi)$

However, I argue that an individual can have contradicting desires, unlike contradicting beliefs. For example, if John has a limited time for his summer vacation and this time suffices only for visiting one city, then the desires expressed in (14) and (15) are contradictory. John cannot visit Rome unless he doesn't visit Paris and vice versa. But if we hear both (14) and (15) one after another, we can understand how this sentences can be true together, even though the desires expressed by them exclude each other. On the contrary, it seems that the following sentences:

(18) John believes he'll be in Rome this summer

and

(19) John believes he'll be in Paris this summer

cannot be true together in the above circumstances, and, in general, that an individual cannot have contradicting beliefs. Having contradicting desires, unlike contradicting beliefs, is actually very common.

4. Absolute preference: Geurts (1998) and Villalta (2000)

Geurts (1998) and Villalta (2000) describe similar semantic analyses for *want*. I will call these *absolute preference* analyses. Geurts (1998) simplifies Heim's (1992) truth conditions as follows:

- (20) $[[d \text{ wants } \varphi]]_w = 1$ iff every world in $Dox_d(w) \cap \varphi$ is better for d in w than every world in $Dox_d(w) \cap \neg\varphi$.

Villalta (2000) proposes the following truth conditions:

- (21) $[[d \text{ wants } \varphi]]_w = 1$ iff every world in φ is better for d in w than every world each contextual alternative ψ .

In the case that the only contextual alternative to φ is $\neg\varphi$, and the only worlds considered are those in $Dox_d(w)$, these truth conditions are identical. These truth conditions are too strong, since other differences between the worlds can affect their desirability. The following example demonstrates such a situation.

Suppose John has two lottery tickets. The first can win \$100, and the second can win \$1000, and the results are independent. John wants the first ticket to win, and he also wants the second one to win. Let φ denote the proposition "the first ticket wins", and ψ – "the second ticket wins". The *absolute preference* truth conditions for "John wants φ " require that he prefers $\varphi \wedge \neg\psi$ to $\neg\varphi \wedge \psi$, that is, he would prefer the situation in which he wins the \$100, but not the \$1000, to the situation in which he wins the \$1000, but not the \$100. This is almost certainly not true. Moreover, the truth conditions for "John wants ψ " require the opposite preference, so the truth conditions for "John wants φ " and "John wants ψ " are, in this case, contradictory, that is, it is not possible for John both to want φ and to want ψ . This shows that *absolute preference* truth conditions are inadequate for this situation, and they are, indeed, too strong for most cases.

5. Heim: strong local preference.

Heim (1992) analyzes desire predicates as including a hidden conditional. Her analysis is based on Stalnaker's (1984, p. 89) observation that "wanting something is preferring it to certain alternatives, the relevant alternatives being those possibilities that the agent believes will be realized if he does not get what he wants". Preferring is formalized using Lewis' (1973) semantics of counterfactual conditionals. Heim's analysis for the sentence " a wants φ " can be seen as a simplified version of Lewis' analysis of "in any possible case, if the truth value of φ changed, a prefers the case in which it's true". Let me now briefly introduce this analysis.

First, Heim defines a preference relation $<_{a,w}$ between worlds as follows:

- (22) For any $w, w', w'' \in W$, $w' <_{a,w} w''$ iff w' is more desirable to a in w than w'' .

This relation is then extended to a relation between sets of worlds as follows:

- (23) For any $w \in W$, $X \subseteq W$, $Y \subseteq W$, $X <_{a,w} Y$ iff $w' <_{a,w} w''$ for all $w' \in X$, $w'' \in Y$.

The relation of comparative similarity between worlds is encoded by a family of selection functions. For each world w , there is a selection function Sim_w from propositions to propositions which maps each ϕ to the set of ϕ -worlds maximally similar to w :

- (24) $Sim_w(\phi) \equiv \{w' \in W \mid w' \in \phi \text{ and } w' \text{ resembles } w \text{ no less than any other world in } \phi\}$

If w itself is a ϕ -world, then $Sim_w(\phi)$ is a singleton of w itself.

According to Heim, the proposition ' a wants ϕ ' is true for an individual a and a proposition ϕ in w iff for every w' in $Dox_a(w)$ every ϕ -world maximally similar to w' is more desirable to a in w than any $\neg\phi$ -world maximally similar to w' . Formally,

- (25) $w \in [[a \text{ wants } \phi]]$ iff for every $w' \in Dox_a(w)$, $Sim_w(\phi) <_{a,w} Sim_w(\neg\phi)$

Heim's analysis solves the problems of "ideal worlds" analyses discussed in Sections 3.3 - 3.5. First, let's consider Upward Entailment: suppose a wants ϕ , and ϕ entails ψ . Does it entail, in Heim's model, that a wants ψ ? Assuming, without loss of generality, that neither of them holds for some w' , $Sim_w(\neg\phi) = Sim_w(\neg\psi) = \{w'\}$. From the fact that $Sim_w(\phi)$ is preferable to $\{w'\}$ it does not necessarily follow that $Sim_w(\psi)$ is also, and they may not even have common members. Consider the Concord's example. All the closest alternatives to w' in which Nicholas gets a free trip are preferable to w' . The same does not hold for simply getting a trip, since the most similar alternatives include paying the full price, which Nicholas doesn't want. We see that Heim's analysis doesn't require Upward Entailment to hold.

Tautologies require a special treatment. If ϕ is a tautology, $Sim_w(\neg\phi)$ will be empty, since there are simply no $\neg\phi$ -worlds. With the regular definitions, this implies that tautologies are always wanted, contrary to the intuition. Heim suggested that if $\neg\phi$ is empty, $Sim_w(\neg\phi)$ is undefined. Moreover, she suggests that a sentence ' a wants ϕ ' has a presupposition that both ϕ and $\neg\phi$ are possible. As with von Fintel's analysis, this presupposition has to be stipulated.

The Conjunction Introduction problem is also solved. If, for epistemically accessible worlds, the closest alternative worlds in which John visits Paris are preferable, and the same holds for visiting Rome, it is not necessarily true that the closest alternative worlds in which John visits both Paris and Rome are

preferable. Visiting both cities may have consequences that will make it unwanted.

Contradicting desires are allowed in Heim's analysis. If ϕ and ψ are contradictory, it is possible that *a* wants ϕ and *a* wants ψ are both true in w_0 . For example, if for some $w' \in \text{Dox}_a(w_0)$, $w' \notin \phi$, $w' \notin \psi$, $\text{Sim}_{w'}(\phi) = \{w''\}$, $\text{Sim}_{w'}(\psi) = \{w'''\}$, $w'' <_{a,w} w'$, $w''' <_{a,w} w'$, then the truth conditions for both desires hold for this belief world. The conditions can hold, in similar ways, for every belief world.

Heim's definition (25) contains *two* universal quantifications: first, over epistemically accessible worlds, and second, over alternatives to each accessible world. In order for *a* wants ϕ to be true, in each case the ϕ -alternative should be more desirable than $\neg\phi$ -alternative. If in one pair of an epistemically possible world and a closest alternative to it *a* prefers the $\neg\phi$ -world, "*a* wants ϕ " is judged as wrong. In other words, a person wants something only if he/she believes that it will necessarily improve the situation in any possible case. As will be shown in the next section, this excludes many actual cases of wanting¹.

6. New examples: the insurance case

Consider the following example:

John returns home in the evening. He knows that the day before there was no milk in the refrigerator in his apartment. He also knows that his roommate Bob noticed it, too, and, in fact, Bob may have already bought some milk. This evening John wants to have some coffee, and he doesn't like coffee without milk. John passes near a supermarket and decides to buy milk.

The considerations John makes in this example can be described as follows. He has a partial desire to have some milk for his coffee. Another desire, of course, is to spend as little money as possible. If Bob had already bought milk, John would lose a small amount of money and gain nothing, so the action of buying unnecessary milk is undesirable as a whole. If Bob didn't buy any milk, John's buying milk has a positive effect on the situation, allowing him to have his favorite drink, and the action as a whole is desirable, in spite of his spending the money. According to Heim's analysis, John shouldn't want to buy milk, since he isn't sure it will lead to a desirable effect. However, he thinks that the positive result in some possible cases outweighs the negative result in other possible cases, and decides that he wants to buy milk in spite of the possibility that it may make the situation worse.

A similar example is buying insurance. Usually, when someone buys house insurance, that person pays a premium that is not returned. Mostly, nothing happens to the insured property, the premium is lost, and buying insurance leads to a less desirable situation for the person who purchased it. However, if some damage actually occurs to the insured property, the cost of the damage, which is usually much more than the premium paid, is returned. The situation is the same as in the previous example: a small loss in most cases and a large gain in some cases.

People can want to buy insurance although in most cases they just lose the premium.

I'll use the following numbers in analyzing the insurance case. The premium cost is 50. The probability of damage: 0.01, which is much higher than in the real life. Damage cost: 4000, so that the insurance company has a positive expected profit. Damage cost as evaluated by the person purchasing insurance: 100000, much higher than the objective cost, since insurance is usually bought when the damage cost cannot be borne, and, unless returned by the insurance company, will cause very unwanted circumstances, such as bankruptcy.

A model for Heim's analysis of the insurance case is shown in Figure 2. Let φ denote the proposition "insurance was bought" and ψ – "the house was ruined". The worlds $w_1 \dots w_4$ correspond to the four possible combinations of truth values of the two formulas. For every world, $Sim_w(\varphi)$ and $Sim_w(\neg\varphi)$ are singletons and the one not equal to the world itself is shown with a dashed line. The epistemic accessibility relation is shown with a solid line.

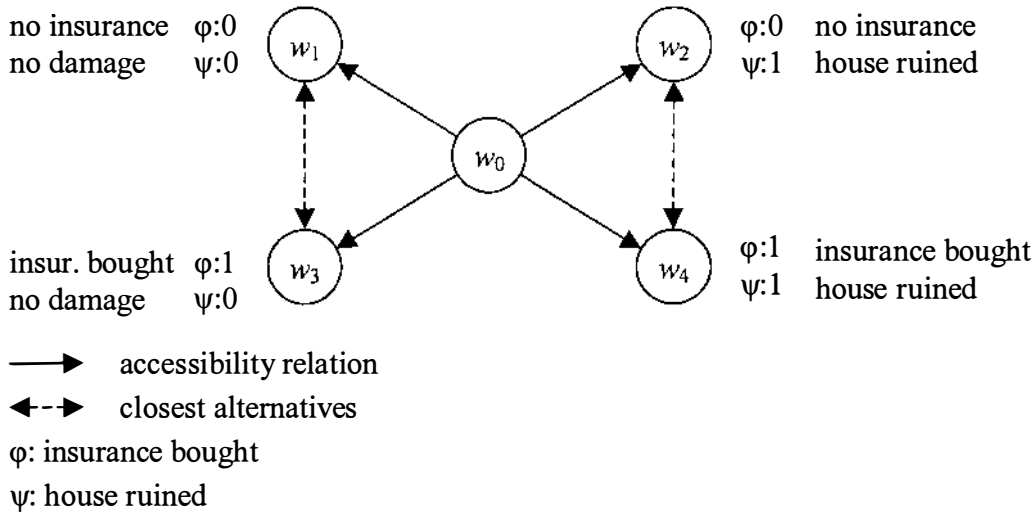


Figure 2. Heim's model for the insurance example.

It cannot be said that a wants φ , since for w_1 and w_3 the φ -alternative (w_3) is less desired than $\neg\varphi$ -alternative (w_1). On the other hand, it also cannot be said that a wants $\neg\varphi$, since for w_2 and w_4 the φ -alternative (w_4) is more desired than $\neg\varphi$ -alternative (w_2). So, the agent neither wants φ nor $\neg\varphi$. However, in the case described above the agent would want to buy insurance, i.e. a wants φ , and the prediction given by Heim's model is wrong.

"Ideal worlds" analyses cannot explain this case, either. $I(a,w)$ contains the world that maximally corresponds to a 's preferences in w , namely w_1 , in which insurance is not bought and nothing happens to the house. The prediction is that the person would never want to buy insurance, contrary to the case described above.

7. My analysis

7.1. First proposal: various evaluation functions

We have seen in Section 2 that a person can have contradicting desire attitudes towards the same proposition. Contradicting desire reports can even be given in one sentence:

- (26) Does John want to go to Rome this summer?
He does and he doesn't.

It is interesting to compare this with contradicting belief reports:

- (27) Does John believe he went to Rome last summer?
He does and he doesn't.

The answers in (26) and (27) mean different things. In (27), John's beliefs are incoherent. Maybe he lost his memory and some clues make him believe he was in Rome, while other clues, having the same importance, suggest the contrary. In (26), however, John can be a perfectly consistent person having different kinds of desire towards the same thing.

I suggest that every desire report contains an implicit *evaluation function*. In Davis' example in Section 2, when he says he wants to play he means it will cause him pleasure, which is desirable, and so expresses a partial desire. When he says he doesn't want to play, he explains that playing would prevent him from doing other important things, which is undesirable, and so expresses a lack of motivational desire.

Formally, a proposition "*a* wants ϕ " should be analysed as "*a* wants ϕ with respect to evaluation function *g*". The reply in (26) can thus be understood as follows: "He wants *with respect to some evaluation function* g_1 and doesn't want *with respect to some evaluation function* g_2 ".

This proposal can actually be combined with any other analysis of desire. For example, Heim's definition of preference relation between worlds, (22), can be changed to

- (28) For any $w, w', w'' \in W$, $w' <_{a,w,g} w''$ iff w' is more desirable to *a* in *w* than w'' with respect to *g*.

If the function *g* is defined from possible worlds to real numbers, " w' is more desirable to *a* in *w* than w'' with respect to *g*" simply means $g_{a,w}(w') > g_{a,w}(w'')$. The definitions based on (22) can be adjusted in the same way.

7.2. Second proposal: a probabilistic model

In order to deal with examples like the insurance case described in Section 6, a probability function *P* defined as follows is added to the model:

- (29) $P_{d,w}(w') = p$ iff the individual d in w assigns probability p to the possibility that w' is the actual world.

Suppose that d in w_0 assigns probability 0.01 to the outcome that the house will be ruined and probability 0.5 to the outcome that she will buy insurance. Figure 3 shows the model for this case, with the reasonable assumption that these events are independent, that is, buying insurance doesn't affect the chances of the house being ruined.

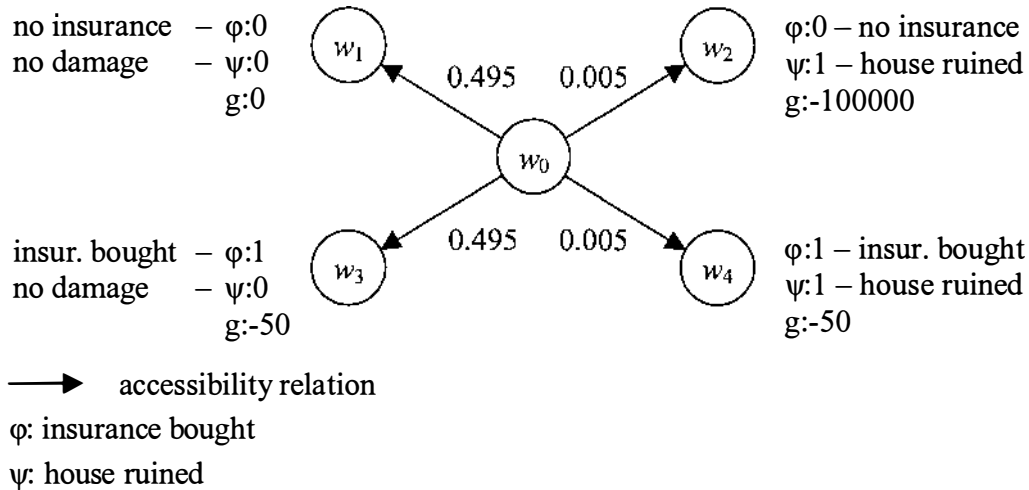


Figure 3. Model for the insurance example, absolute probabilities. Numbers on edges from w_0 to $w_{1..4}$ are absolute probabilities $P_{w_0}(w_i)$.

I argue that d wants φ does not mean that d thinks that φ will improve the situation in every possible case. Instead, it means that the *expected overall effect* for all the possible situations is positive. That is, the *subjective expectation of the desirability* (modeled by the function g) is higher in the case that φ is true than in the case than φ is false. Formally,

- (30) $w \in [[d \text{ wants } \varphi \text{ w.r.t. } g]]$ iff
 $\text{SubjExp}_{d,w}(g(w') \mid \varphi \text{ is true}) > \text{SubjExp}_{d,w}(g(w') \mid \varphi \text{ is false})$

What is SubjExp , that is, how is the expectation of desirability calculated by an individual? Van Rooy (1999) uses the answer given by Savage's (1954) theory of choice under uncertainty and Jeffrey's (1965) preference theory, according to which SubjExp is *expected utility*, that is, expectation with subjective probabilities and values:

$$(31) \quad \text{SubjExp}(g(x_i)) = E(g(x_i)) = \sum_i g(x_i) \cdot p(x_i)$$

However, there are phenomena, such as *risk aversion* (Allais 1953), that suggest that (31) does not correctly represent the way in which humans calculate

the overall desirability of a choice option. Alternative theories exist, e.g., Kahneman and Tversky (1979). I will use (31) in this paper for ease of presentation, as it is a good approximation for the actual SubjExp. While the issue of truth conditions for *want* belongs to semantics, the question what exactly SubjExp is, in my opinion, does not, and will be left open.

Thus,

$$(32) \quad w \in [[d \text{ wants } \varphi \text{ w.r.t. } g]] \text{ iff } E(g(w') \mid \varphi \text{ is true}) > E(g(w') \mid \varphi \text{ is false})$$

Expressing the expectation with absolute probabilities,

$$(33) \quad w \in [[d \text{ wants } \varphi \text{ w.r.t. } g]] \text{ iff } \frac{\sum_{[[\varphi]]_{w'}=1} P_{d,w}(w') \cdot g(w')}{\sum_{[[\varphi]]_{w'}=1} P_{d,w}(w')} > \frac{\sum_{[[\varphi]]_{w'}=0} P_{d,w}(w') \cdot g(w')}{\sum_{[[\varphi]]_{w'}=0} P_{d,w}(w')}$$

In the case that φ denotes an action an agent considers whether to do, the truth conditions in (33) include the probability of this action itself. However, if the agent is considering the action, it's unnatural to assume that he or she also estimates the probability of himself/herself performing it. It is possible to avoid this problem by reformulating (33). Instead of using the absolute probability defined in (29), a conditional probability defined as follows can be used:

$$(34) \quad P_{d,w}(w' \mid [[\varphi]] = i) = p \text{ iff the individual } d \text{ in } w \text{ assigns probability } p \text{ to the possibility that } w' \text{ is the actual world given that the value of } \varphi \text{ in the actual world is } i.$$

The final truth conditions for '*d* wants φ with respect to g ' in the proposed model, assuming the expected utility theory, are obtained from (32) by expressing the expectation with conditional instead of absolute probabilities:

$$(35) \quad w \in [[d \text{ wants } \varphi \text{ w.r.t. } g]] \text{ iff } \sum_{w' \in W} g(w') \cdot P_{d,w}(w' \mid [[\varphi]] = 1) > \sum_{w' \in W} g(w') \cdot P_{d,w}(w' \mid [[\varphi]] = 0)$$

Note that $P_{d,w}(w' \mid [[\varphi]] = i) \neq 0 \Rightarrow [[\varphi]]_{w'} = i$. For each w' , either $P_{d,w}(w' \mid [[\varphi]] = 1) = 0$ or $P_{d,w}(w' \mid [[\varphi]] = 0) = 0$ (of course, both can be zero as well), that is, at most one value of the two can be non-zero. Figure 4 gives the model for the insurance case, this time with conditional probabilities.

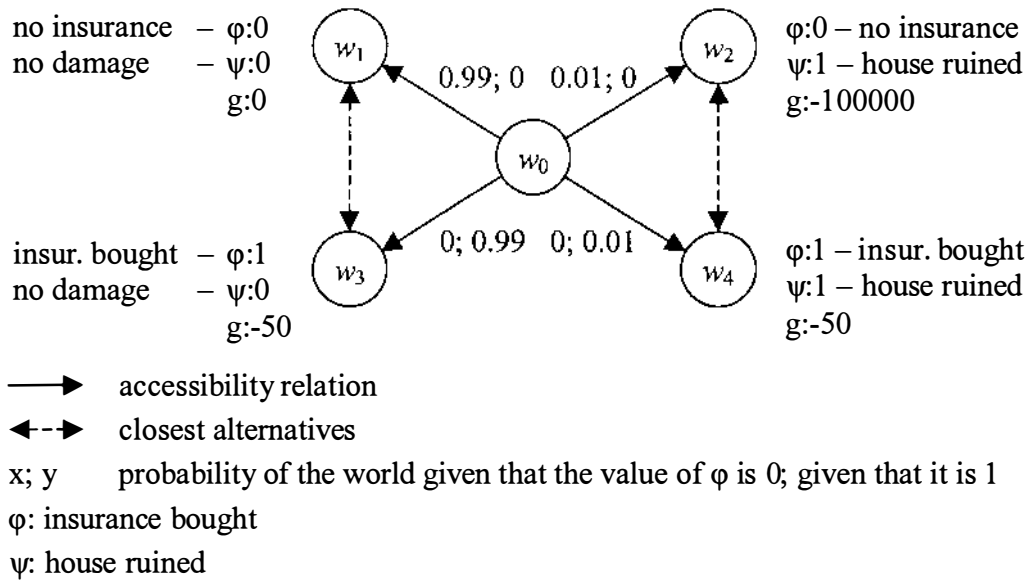


Figure 4. Model for the insurance example, conditional probabilities. The numbers on the edges from w_0 to $w_{1..4}$ are conditional probabilities $P_{d,w_0}(w_A = w_i | [[\varphi]]_{w_A} = 0)$; $P_{d,w_0}(w_A = w_i | [[\varphi]]_{w_A} = 1)$.

Let's make the computations for the insurance example using (35):

$$\begin{aligned}
 &w \in [[d \text{ wants } \varphi \text{ w.r.t. } g]] \\
 &\text{iff } (0*0 + (-50)*0.99 + (-100000)*0 + (-50)*0.01) > \\
 &\quad (0*0.99 + (-50)*0 + (-100000)*0.01 + (-50)*0) \\
 &\text{iff } -50 > -1000
 \end{aligned}$$

My analysis predicts correctly that *d wants φ* .

Example (36), which is a coherent self-report, presents the linguistic evidence for the insurance case discussed above. Heim's analysis classifies it as a contradiction, since the speaker reports both preferring of $\neg\varphi$ in most cases and desire for φ , which is taken to mean preferring of φ in all the possible cases.

(36) I want to buy insurance. I know that most probably I'll just lose the money.

According to the analysis presented here, neither upward entailment nor conjunction introduction necessarily hold for desire. We have seen in Sections 3.3 and 3.4 that these are indeed the features of the predicate *want*.

8. Additional issues

8.1. Presupposition

Heim (1992) and von Stechow (1999) propose that the want report *a wants φ* has a presupposition that both φ and $\neg\varphi$ are possible. In these analyses, this is a

stipulation. However, it follows directly from my analysis². Conditioning on the cases that φ is true/false requires that there be such cases. If either φ or $\neg\varphi$ is not possible, one of the fractions in (33) becomes 0/0 (undefined), and the truth conditions cannot be applied. In order for the truth conditions to be applicable, there should be both worlds in which φ and worlds in which $\neg\varphi$, and this presupposition is accounted for.

8.2. *Degrees of wanting*

Want is a predicate allowing degrees. One can want something ‘a little’ or ‘very much’. There is a corresponding notion for this in my analysis: the difference between $\text{SubjExp}_{d,w}(g(w') \mid \varphi \text{ is true})$ and $\text{SubjExp}_{d,w}(g(w') \mid \varphi \text{ is false})$:

$$(37) \quad \text{WantingDegree}_{d,w,g}(\varphi) = \\ \text{SubjExp}_{d,w}(g(w') \mid \varphi \text{ is true}) - \text{SubjExp}_{d,w}(g(w') \mid \varphi \text{ is false})$$

The higher the difference, the stronger the desire that φ . Wanting ‘a little’ means the difference is small, and wanting ‘very much’ means that the difference is large. No straightforward notion corresponding to the degree of desire exists in the other analyses.

8.3. *Neg-raising*

The notion of *WantingDegree* defined in (37) can be used to reformulate the truth conditions in (30):

$$(38) \quad w \in [[d \text{ wants } \varphi \text{ w.r.t. } g]] \text{ iff } \text{WantingDegree}_{d,w,g}(\varphi) > 0$$

Such a way of formulating the truth conditions is useful to explain the fact that *want* is a neg-raising verb, that is, that (39) can be used instead of (40):³

(39) I don’t want you to go there.

(40) I want you not to go there.

The meaning *want*($\neg\varphi$) is conveyed by a sentence $\neg\text{want}(\varphi)$. If the truth conditions for *want* are formulated as in (38), *want* is, in terms of Horn (1989), a *mid-scalar* predicate. This establishes *want* as a *weakly intolerant* predicate (satisfying both (41) and (42)), as opposed to *tolerant* (not satisfying (41)), and *strongly intolerant* (satisfying (41), but not (42)). Only weakly intolerant predicates allow Neg-Raising, according to Horn.

$$(41) \quad P(\neg\varphi) \Rightarrow \neg P(\varphi)$$

$$(42) \quad P(\neg\varphi) \text{ and } \neg P(\varphi) \text{ are close in their truth conditions}$$

It is easy to see the relevance of (42), since the phenomenon of Neg-Raising is, generally speaking, using an expression of the form $\neg P(\varphi)$ to convey the meaning $P(\neg\varphi)$.

According to my analysis,

(43) $w \in [[d \text{ wants } \varphi \text{ w.r.t. } g]]$ iff $\text{WantingDegree}_{d,w,g}(\varphi) > 0$

(44) $w \in [[\neg(d \text{ wants } \varphi \text{ w.r.t. } g)]]$ iff $\text{WantingDegree}_{d,w,g}(\varphi) \leq 0$

(45) $w \in [[d \text{ wants } (\neg\varphi) \text{ w.r.t. } g]]$ iff $\text{WantingDegree}_{d,w,g}(\neg\varphi) > 0$ iff $\text{WantingDegree}_{d,w,g}(\varphi) < 0$

The truth conditions in (45) entail the ones in (44), and the only difference between (44) and (45) is in the case that the value of $\text{WantingDegree}_{d,w,g}(\varphi)$ is 0. Therefore, the proposed truth conditions for *want* satisfy (41) and (42), and *want* under my analysis is indeed a weakly tolerant predicate. According to the other analyses, $want(\neg\varphi)$ and $\neg want(\varphi)$ have very different truth conditions, and *want* is *strongly intolerant*. The analysis proposed in this paper, unlike the others, is consistent with Horn's account of Neg-Raising. This is additional evidence supporting the proposed analysis.

8.4. Analyzing desire verbs as propositional attitudes

It is a convention to analyze *want* as a propositional attitude, that is, a relation between an individual and a proposition. However, unlike the case with *believe*, *want* is normally not complemented by a proposition. The following examples demonstrate the three options for the second complement of *want*: an infinitival phrase with a subject, a subjectless infinitival phrase, and an NP.

(46) I want you to go.

(47) I want to eat.

(48) I want an apple.

None of the examples contains an explicit proposition as a complement of *want*. Rephrasing makes the sentences ungrammatical. In some languages (49) is accepted, but (50) and (51) are not. Interestingly, the same sentences with *believe* are grammatical.

(49) I believe/*want that you go.

(50) I believe/*want that I eat.

(51) I believe/*want that I eat/have an apple.

These facts caused Ben-Yami (1997) to claim that mental states, including desires, are not propositional attitudes. However, he does not provide any alternative semantics.

Analyzing *want* as a propositional attitude raises the problem of explaining the transition from the actual complement of *want* to a full tensed proposition. The contracted form of (46) and (47) and the lack of an overt subject in the clause of (47) can be explained by syntactic means. In addition, in (46) and (47) the tense of the verb is missing, and while Stowell (1982) suggests that it is “a possible future”, the semantic literature ignores this issue. Cases like (48), in which the lexical entry of the verb is missing as well, were also unnoticed by the previous analyses. Future research of the semantics of *want* should try to provide explanations for these phenomena.

Endnotes

* I would like to thank Nirit Kadmon for invaluable help with the research that led to this paper. Thanks also to Robert van Rooy for helpful discussions and to anonymous reviewers for helpful suggestions.

1. For another kind of counterexample to Heim’s analysis, see Villalta (2000) and van Rooy (1999).
2. I would like to thank Chung-chieh (Ken) Shan for this observation.
3. For more examples, see Borkin (1972).

References

- Anscombe, G. E. M. (1963). *Intention*. Oxford: Basil Blackwell. Second edition.
- Asher, Nicholas (1987). A typology for attitude verbs and their anaphoric properties. *Linguistics and Philosophy* 10, 125-197.
- Ben-Yami, Hanoch (1997). Against characterizing mental states as propositional attitudes. *Philosophical Quarterly* 47, 84-89.
- Borkin, Ann (1972). Two Notes on *Want* and *Desire*. *Linguistic Inquiry* 3(3), 378-385.
- Davidson, Donald (1978). Intending. In Yirmiyahu Yovel (ed.), *Philosophy of History and Action*. Reprinted in Davidson (1980).
- Davidson, Donald (1980). *Essays on Actions and Events*. Oxford.
- Davis, Wayne A. (1986). The Two Senses of Desire. In: Marks, Joel (ed.), *The Ways of Desire*. Chicago: Precedent Publishing.
- von Fintel, Kai (1999). NPI-Licensing, Strawson-Entailment, and Context-Dependency. *Journal of Semantics* 16, 97-148.
- Geurts, Bart (1998). Presuppositions and Anaphors in Attitude Contexts. *Linguistics and Philosophy* 21, 545-601.
- Heim, Irene (1992). Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics* 9, 183-221.

- Hintikka, Jaakko (1969). Semantics for Propositional Attitudes, in J. W. Davis *et al.* (ed.), *Philosophical Logic*, Dordrecht, 21-45.
- Horn, Laurence R. (1989). *A Natural History of Negation*. University of Chicago Press, Chicago.
- Jeffrey, Richard C. (1965). *The Logic of Decision*. University of Chicago Press, Chicago.
- Kahneman, Daniel, and Tversky, Amos (1979). Prospect Theory: an Analysis of Decision under Risk. *Econometrica* 47, 263-291.
- Kratzer, Angelika (1981). The Notional Category of Modality, in H.-J. Eikmeyer and H. Reiser (eds.), *Words, Worlds, and Contexts*, Berlin – New York: Walter de Gruyter, 38-74.
- Lewis, David K. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Locke, Don (1982). Beliefs, Desires and Reasons for Actions. *American Philosophical Quarterly* 19, 241-249.
- van Rooy, Robert (1999). Some analyses of pro-attitudes. In: H. de Swart (ed.), *Proceedings from Logic, Game Theory and Social Choice 1*.
- Savage, Leonard J. (1954). *The Foundations of Statistics*. New York: John Wiley.
- Searle, John R. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Stalnaker, Robert C. (1984). *Inquiry*. Cambridge, MA: The MIT Press.
- Stowell, Tim (1982). The Tense of Infinitives. *Linguistic Inquiry* 13, 561-570.
- Villalta, Elisabeth (2000). Spanish Subjunctive Clauses Require Ordered Alternatives. *Proceedings of SALT X*.