A short note on belief under desire

Luka Crnič

`luka.crnic@mail.huji.ac.il`

Tue Trinh

`tuetrinh@uwm.edu`

Abstract

Embedded epistemic modals are infelicitous under desire predicates when they are anchored to the belief state of the attitude holder (see, esp., Anand and Hacquard 2013). We present two ways of deriving this observation from an independently motivated property of desire predicates, their anti-opinionatedness (Heim 1992; von Fintel 1999).

Keywords: Epistemics, Desideratives, Modality, Propositional Attitudes

1 A puzzle about embedded epistemics

The generalization   In (1a), the epistemic modal **must** is anchored to what John knows, and in (1b), to what the police knows.[1]

(1)   a.   John believes that Mary, given what he knows, must be the murderer
        b.   John believes that Mary, given what the police knows, must be the murderer

We say an epistemic modal embedded under an attitude verb is "subject oriented" when it is anchored to the belief state of the attitude holder. Now consider (2).

(2)   a.   #John hopes that Mary, given what he knows, must be the murderer
        b.   John hopes that Mary, given what the police knows, must be the murderer

When the embedding verb is the desire predicate **hope** instead of the doxastic predicate **believe**, the subject oriented reading of the embedded epistemic gives rise to deviance (cf. Hacquard and Wellwood, 2012; Anand and Hacquard, 2013). Let us state the generalization.

(3)   Embedded Epistemics Generalization (EEG)
        Subject oriented epistemics are infelicitous under desire predicates.

Note that without adverbials such as **given what the police knows**, embedded epistemics tend to have the subject oriented reading as default, hence the infelicity of the following sentences.[2]

(4)   a.   #John hopes that Mary must be the murderer
        b.   #John wants Mary to have to be the murderer
        c.   #John demands that Mary must be the murderer

---

[1]   For now, we remain vague about what "anchored" means. Here and in the rest of the paper, all embedded modals are intended to have the epistemic reading.

[2]   Because **want** does not take a tensed complement, the embedded modal cannot be **must**.

2 Anti-opinionatedness

This paper presents a new derivation of the EEG. We start with the semantics of desire predicates, taking the verb **want** to be their representative. Heim (1992) argues for a non-monotonic semantics of **want** which says that

(5)   Heim's analysis of **want**
        $[\![\textbf{want } \phi]\!]^{i,g}(x) = 1$ iff $\forall i' \in \mathcal{B}_x^i[sim_{i'}(\mathcal{B}_x^i \cap [\lambda i.[\![\phi]\!]^{i,g}]) <_{i,x} sim_{i'}(\mathcal{B}_x^i \cap [\lambda i.[\![\neg\phi]\!]^{i,g}])]$

where $\mathcal{B}_x^i$ is the set of indices compatible with x's beliefs at i, $sim_i(X)$ picks out those members of X that are "most similar" to i, and $X <_{i,x} Y$ is the condition that every member of X is better for x at i than every member of Y.[3] Thus, Heim's analysis of **want** involves a comparison between the ϕ and the $\neg\phi$ alternatives within the attitude holder's belief, stating, essentially, that for x to want ϕ is for x to believe that every way of making ϕ true is better than any way of making $\neg\phi$ true.

In contrast to Heim, von Fintel (1999) argues for a monotonic, Kratzerian semantics of **want** which says that

(6)   von Fintel's analysis of **want**
        $[\![\textbf{want } \phi]\!]^{i,g}(x) = 1$ iff $max_{\mathcal{D}_x^i}(\mathcal{B}_x^i) \subseteq [\lambda i.[\![\phi]\!]^{i,g}]$

where $\mathcal{D}_x^i$ is the set of x's desires at i and $max_{\mathcal{P}}(X)$ picks out those members of X that are "optimal" with respect to $\mathcal{P}$.[4] According to this analysis, for x to want ϕ is for x to find ϕ true in the most desirable scenarios among those which he considers possible.

The explanation we are going to provide for the EEG does not require us to choose between (5) and (6). We can therefore remain agnostic with respect to the differences between the non-monotonic and the monotonic analysis of **want** as well as which of the two analyses may be more adequate. However, we will not remain agnostic with respect to what these analyses have in common. Both (5) and (6) make crucial reference to an information state, specifically the belief state of the attitude holder, and both turn out to predict that a **want**-sentence will be a triviality in case the subject believes the complement or its negation.[5] These predictions are intuitively false, since it does not seem valid to infer from **John believes that Mary smokes** or **John believes that Mary doesn't smoke** that **John wants Mary to smoke** is true, or is false. In fact, the last sentence is perceived as infelicitous if either of the first two sentences is true.

To resolve this problem, Heim and von Fintel both add a definedness condition to the semantics of **want**, given in (7), which requires that the subject of **want** not be opinionated about its complement. As desired, this condition not only guarantees that we cannot infer from **John believes that Mary (doesn't) smokes** to **John wants Mary to smoke**, but also entails that these sentences are incompatible (see Heim 1992; von Fintel 1999 for discussion and refinements).

---

[3]   To be explicit, $X <_{i,x} Y$ iff $\forall x \in X[\forall y \in Y[x <_{i,x} y]]$.

[4]   Specifically, $max_{\mathcal{P}}(X) = \{i \in X \mid \neg\exists i' \in X[\{p \in \mathcal{P} \mid p(i) = 1\} \subset \{p \in \mathcal{P} \mid p(i') = 1\}]\}$. Note that $\mathcal{D}_x^i$, in the lexical entry in (6), is a set of propositions, not a set of indices.

[5]   Consider Heim's analysis. If x believes ϕ, then $\mathcal{B}_x^i \cap [\lambda i.[\![\neg\phi]\!]^{i,g}] = \emptyset$, hence $sim_{i'}(\mathcal{B}_x^i \cap [\lambda i.[\![\neg\phi]\!]^{i,g}]) = \emptyset$. Similarly, if x believes $\neg\phi$, then $sim_{i'}(\mathcal{B}_x^i \cap [\lambda i.[\![\phi]\!]^{i,g}]) = \emptyset$. But given the definition of $<_{i,x}$, it is trivially true that $X <_{i,x} Y$ if either $X = \emptyset$ or $Y = \emptyset$ (see note 3). Now consider von Fintel's analysis. Given that $max_{\mathcal{D}_x^i}(\mathcal{B}_x^i) \subseteq \mathcal{B}_x^i$ by definition, if x believes ϕ, then $\mathcal{B}_x^i \subseteq [\lambda i.[\![\phi]\!]^{i,g}]$, hence $max_{\mathcal{D}_x^i}(\mathcal{B}_x^i) \subseteq [\lambda i.[\![\phi]\!]^{i,g}]$, and if x believes $\neg\phi$, then $\mathcal{B}_x^i \subseteq [\lambda i.[\![\neg\phi]\!]^{i,g}]$, hence $max_{\mathcal{D}_x^i}(\mathcal{B}_x^i) \subseteq [\lambda i.[\![\neg\phi]\!]^{i,g}]$, which means $max_{\mathcal{D}_x^i}(\mathcal{B}_x^i) \nsubseteq [\lambda i.[\![\phi]\!]^{i,g}]$.

(7)   Anti-opinionatedness
       $[\![\textbf{want } \phi]\!]^{i,g}(x)$ is defined only if $\exists i' \in \mathcal{B}^i_x([\![\phi]\!]^{i',g} = 1) \wedge \exists i' \in \mathcal{B}^i_x([\![\phi]\!]^{i',g} = 0)$

In what follows, we provide two accounts of the EEG, both of which capitalize on the anti-opinionatedness condition and which differ in their assumptions about modal semantics.

3   Resolving the puzzle

In order to appreciate the import of anti-opinionatedness for the EEG, we need to make some assumptions about epistemic modals. We discuss two common approaches to them and show that, in concert with anti-opinionatedness, they allow for a routine explanation of the EEG.

Relational semantics   We start the standard relational semantics for epistemic modals in (8), where $g(\mathcal{R})$ is the contextually determined accessibility relation.

(8)   Relational semantics of **must**
       $[\![\textbf{must}_{\mathcal{R}} \ \phi]\!]^{i,g} = 1$ iff $g(\mathcal{R})(i) \subseteq [\lambda i.[\![\phi]\!]^{i,g}]$

We assume that the subject oriented reading of embedded epistemics results from $g(\mathcal{R})$ being set to $[\lambda i.\mathcal{B}^i_x]$, where x is the subject of the embedding attitude verb. Now consider (9).

(9)   #John demands that Mary must be the murderer

Under the (infelicitous) subject oriented reading, anti-opinionatedness imposes on (9) the definedness condition in (10), with i the index of evaluation and $\phi = $ **Mary be the murderer**.

(10) $\exists i' \in \mathcal{B}^i_j([\![\textbf{must}_{\mathcal{R}} \ \phi]\!]^{i',g} = 1) \wedge \exists i' \in \mathcal{B}^i_j([\![\textbf{must}_{\mathcal{R}} \ \phi]\!]^{i',g} = 0)$, i.e.
        $\exists i' \in \mathcal{B}^i_j(\mathcal{B}^{i'}_j \subseteq [\lambda i.[\![\phi]\!]^{i,g}]) \wedge \exists i' \in \mathcal{B}^i_j(\mathcal{B}^{i'}_j \nsubseteq [\lambda i.[\![\phi]\!]^{i,g}])$

What (10) says is that John's belief does not rule out the possibility that he believes that Mary is the murderer, and does not rule out the possibility that he does not believe that Mary is the murderer. Thus, (10) says that John is not opinionated about his own belief. Obviously, nothing we have said so far prevents (10) from being true. Thus, we predict, all things being equal, that (9) should be felicitous, contrary to observation.

There may be good reasons to believe that not all things are equal, however. One common hypothesis about epistemic agents is that they are in fact opinionated about their own belief: if x believes p then x believes that x believes p, and if x does not believe p then x believes that x does not believe p (e.g., Hintikka 1962; Lewis 1969; Stalnaker 2002, among others).

(11)   Opinionatedness
        for all x, p, i, it holds that $\forall i' \in \mathcal{B}^i_x(\mathcal{B}^{i'}_x \subseteq p) \vee \forall i' \in \mathcal{B}^i_x(\mathcal{B}^{i'}_x \nsubseteq p)$

With this additional hypothesis in hand, we can now explain the deviance of (9): anti-opinionatedness and opinionatedness, together, impose contradictory demands on the attitude holder of (9), since (10) is incompatible with (11). Note, importantly, that the conflict between anti-opinionatedness and opinionatedness arises only in the case of subject oriented epistemics under desire predicates: if the embedding verb is a doxastic predicate, anti-opinionatedness does not apply – this accounts for the acceptability of (1) –, and if the reading is not subject oriented, opinionatedness is not problematic – this accounts for the acceptability of (2b).

Domain semantics Can we derive the deviance of (9) by formulating a different, non-relational, semantics for modals instead of adding opinionatedness to the theory? It turns out we can. Suppose we follow Yalcin (2008) and assume a "domain" semantics for modals, taking the index to be a pair $\langle w, \mathcal{S} \rangle$ with w a possible world and $\mathcal{S}$ an information state which is a set of possible worlds.[6]

(12)   Domain semantics for **must**
$$\llbracket \textbf{must } \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g} = 1 \text{ iff } \mathcal{S} \subseteq [\lambda w. \llbracket \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g}]$$

The subject oriented reading of an embedded epistemic modal, in this framework, will result from $\mathcal{S}$ being set to $\mathcal{B}_x^w$, with x the subject of the embedding verb and w the world of evaluation for the attitude ascription. Suppose anti-opinionatedness is formulated as in (13),

(13)   Anti-opinionatedness (domain semantics version)
$\llbracket \textbf{want } \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g}(x)$ is defined only if
$$\exists w' \in \mathcal{B}_x^w (\llbracket \phi \rrbracket^{\langle w', \mathcal{S} \rangle, g} = 1) \wedge \exists w' \in \mathcal{B}_x^w (\llbracket \phi \rrbracket^{\langle w', \mathcal{S} \rangle, g} = 0)$$

the definedness condition imposed by anti-opinionatedness on (9), under the subject oriented reading, will then be (14), with $\phi =$ **Mary be the murderer** and w the world of evaluation.

(14) $\exists w' \in \mathcal{B}_j^w (\llbracket \textbf{must } \phi \rrbracket^{\langle w', \mathcal{B}_j^w \rangle, g} = 1) \wedge \exists w' \in \mathcal{B}_j^w (\llbracket \textbf{must } \phi \rrbracket^{\langle w', \mathcal{B}_j^w \rangle, g} = 0)$, i.e.
$\exists w' \in \mathcal{B}_j^w (\mathcal{B}_j^w \subseteq [\lambda w. \llbracket \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g}]) \wedge \exists w' \in \mathcal{B}_j^w (\mathcal{B}_j^w \not\subseteq [\lambda w. \llbracket \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g}]$, i.e.
$\mathcal{B}_j^w \subseteq [\lambda w. \llbracket \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g}] \wedge \mathcal{B}_j^w \not\subseteq [\lambda w. \llbracket \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g}]$

Of course, (14) is a contradiction. This means that we can say that the deviance of (9) is due to its having a non-satisfiable definedness condition. Note, again, that this situation arises only when the epistemic is embedded under a desire predicate and has the subject oriented reading. If the embedding verb is a doxastic predicate, anti-opinionatedness does not apply, and if the reading is not subject oriented, the existential quantification in (14) would not be superfluous. To illustrate the latter scenario, consider (15).

(15)   John demands that Mary must be the murderer according to the police

The complement of **demands** in (15), clearly, should be evaluated with respect to John's belief about what the police believes. Thus, anti-opinionatedness would impose on (15) the condition in (16), where $\mathcal{B}_p^{w'}$ is the police's belief at w'.

(16) $\exists w' \in \mathcal{B}_j^w (\mathcal{B}_p^{w'} \subseteq [\lambda w. \llbracket \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g}] \wedge \exists w' \in \mathcal{B}_j^w (\mathcal{B}_p^{w'} \not\subseteq [\lambda w. \llbracket \phi \rrbracket^{\langle w, \mathcal{S} \rangle, g}])$

The existential quantifications in (16) are not superfluous, hence (16) is not contradictory.

Distinguishing between the two approaches? The relational semantics account requires two definedness conditions, namely anti-opinionatedness and opinionatedness, while the domain semantics account requires only one, namely anti-opinionatedness. It follows, then, that an observation which requires abandoning at least one of these conditions will count as evidence in favor of domain semantics. With that said, consider (17).

(17)   John wants to believe that Mary is the murderer

---

[6]   Being true at an index will now mean being true with respect to w, as in the case of non-modalized sentences, or being true with respect to $\mathcal{S}$, as in the case of modalized sentences. See Yalcin (2008) for more details.

The data we have discussed until now involve epistemics embedded under desires and doxastic predicates. We have not looked at cases where a doxastic predicate is embedded under a desire predicate, which is what (17) is. It turns out that to the extent that (17) is felicitous, it supports the domain semantics account. Anti-opionatedness imposes on (17) the condition in (18a), and opinionatedness imposes on it the condition in (18b), with p = 'that Mary is the murderer.'

(18)   a. $\neg$(John believes (he believes p)) $\wedge$ $\neg$(John believes $\neg$(he believes p))
         b.   (John believes (he believes p)) $\vee$ (John believes $\neg$(he believes p))

As (18a) and (18b) contradict each other, the felicity of (17), to the extent that it is real, means that one of these conditions is false, hence constitutes evidence in favor of domain semantics.

4   Residual issues

Existential modals   Anti-opinionatedness and opinionatedness impose the same conditions on **must** ϕ as they do on **might** ψ, where ψ is $\neg\phi$. Thus, we expect the observations we have made about **must** to hold for its dual **might** also. This is true to a large extent.

(19)   a.   #John hopes that Mary, given what he knows, might be the murderer
         b.   John hopes that Mary, given what the police knows, might be the murderer

(20)   a.   John believes that Mary might be the murderer
         b.   #John demands that Mary might be the murderer

However, there is a difference between **must** and **might**: it seems that in sentences without adverbials such as **given what x knows**, the non-subject oriented reading under **hope** is more easily obtained with **might** than with **must** (cf. Anand and Hacquard 2013).

(21)   a.   #John hopes that Mary must be the murderer
         b.   John hopes that Mary might be the murderer

We have no solution to this puzzle and will have to leave it to future work.

Variability with respect to anti-opinionatedness   We have formulated anti-opinionatedness as a felicity condition which is to be imposed on all desire predicates. Data show that reality is more fine-grained, and that anti-opinionatedness should likely be lexically constrained, with each verb, in principle, determining the relevant domain in its own way (cf., e.g., Heim 1992; von Fintel 1999; Scheffler 2008; Anand and Hacquard 2013, among others). Consider the contrast in (22).

(22)   a.   I know Mary is playing video games but I want her to be swimming now
         b.   #I know Mary is playing video games but I hope that she is swimming now

(23)   a.   #I want to have been sick
         b.   I wish to have been sick

Again, we have no solution to this puzzle and will leave it to future work.

Anand and Hacquard (2013) In their pioneering work on the EEG, Anand and Hacquard (2013) adopt a preference-based, Heimian analysis of **want**, combining it with a domain semantics for modals, and with the assumption that the complement of **want** will be evaluated with respect to a "special" information state, namely $\emptyset$.[7]

(24)   Anand and Hacquard's analysis of **want**
$[\![\mathbf{want}\ \phi]\!]^{\langle w,\mathcal{S}\rangle,g}(x) = 1$ iff $[\lambda w'.[\![\phi]\!]^{\langle w',\emptyset\rangle,g}] <_{w,x} [\lambda w'.[\![\neg\phi]\!]^{\langle w',\emptyset\rangle,g}]$

Coupled with the assumption that $[\![\mathbf{must}\ \phi]\!]^{\langle w,\mathcal{S}\rangle,g}$, and $[\![\mathbf{might}\ \phi]\!]^{\langle w,\mathcal{S}\rangle,g}$, are undefined if $\mathcal{S} = \emptyset$, this semantics has the consequence that a **want** sentence is defined only when the complement of **want** is such that its evaluation makes no reference to the information state, i.e. that it is non-modalized.

In contrast to Anand and Hacquard's account, our proposal is agnostic with respect to the assertive component of desire predicates as well as compatible with the tight connection between belief and desire argued for in Heim (1992) and von Fintel (1999). Furthermore, the proposal does not require any special assumptions about what information states epistemic modals can be relativized to in specific constructions. While there may well be strong reasons for adopting Anand and Hacquard's treatment of **want** and their split between doxastic and desire predicates, the EEG does not necessarily furnish one.

References

Anand, Pranav, and Valentine Hacquard. 2013. Epistemics and attitudes. Semantics and Pragmatics 6:1–59.

von Fintel, Kai. 1999. NPI licensing, Strawson entailment, and context dependency. Journal of Semantics 16:97–148.

Hacquard, Valentine, and Alexis Wellwood. 2012. Embedding epistemic modals in english: A corpus-based study. Semantics and Pragmatics 5:1–29.

Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. Journal of Semantics 9:183.

Hintikka, Jaakko. 1962. Knowledge and Belief: An Introduction to the Logic of Two Notions. Cornell University Press, Ithaca, NY.

Lewis, David. 1969. Convention: A Philosophical Study. Harvard University Press, Cambridge, MA.

Scheffler, Tatjana. 2008. Semantic Operators in Different Dimensions. University of Pennsylvania, PhD thesis.

Stalnaker, Robert. 2002. Common ground. Linguistics and Philosophy 25:701–721.

Yalcin, Seth. 2008. Modality and inquiry. Doctoral Dissertation, Massachusetts Institute of Technology.

---

[7]   The formulation in (24) renders Anand & Hacquard's (36), i.e. their proposed semantics for **want**, in notation more consistent with that used in this paper, while keeping to the same content.