Epistemic resistance moves

Justin Bledin (bledin@jhu.edu) and Kyle Rawlins (kgr@jhu.edu)
Johns Hopkins University

This paper examines what we term 'resistance moves', focusing on the use of epistemic possibility modals. A resistance move is a response that neither agrees nor disagrees, but raises a possibility that might lead to the respondee adjusting their commitments. For example, if A says *I'm going to the party* and B replies, *Diana might be there*, in the right context, A might well reconsider. We propose that resistance moves involve attention to issues, and that acceptance in discourse is filtered through attentional states. The *might* response draws attention to the issue of whether Diana will be there, and in a context where A has not been attending to this issue, will lead to a shift in attention. The shift in attention leads potentially to new propositions becoming 'visible' (in the sense of Yalcin), and consequently, revision. We develop the proposal in a Stalnakerian framework for discourse, using Lewis's 'subject matters' as a tool for formalizing attention and visibility. Ultimately, epistemic possibility modals are a useful tool for drawing attention, but we suggest that this mechanism is a very general one, and attention can be drawn in many ways.

Epistemic resistance moves

Epistemic possibility modals can be used to *resist* an interlocutor's claim, as in (1):

(1) (Background fact: A really likes Diana.)
A: I'm going to skip the party.
B: Diana might be there. *(resistance move)*
A: Oh, maybe I'll go after all.

B uses the modal claim to postpone acceptance of A's claim, and offers an inducement for A to change their mind. Here we address two questions: (i) what are the dynamics of this sort of resistance move, and (ii) what features of epistemic possibility modals lead to this use. The basic proposal is that, in addition to their regular modal semantics, modals *draw attention* to an issue corresponding to their prejacent proposition (cf. Ciardelli et al. 2009). To support this, we propose that discourse participants track attention to issues in context, and that their discourse commitments are 'filtered' through the issues they are actively attending to: revisions in attention (attending to issues one was previously not attending to) can result in revisions in what one is willing to commit to. In (1), B draws attention to the issue of whether Diana will be at the party, an issue that A has overlooked. The end result is that A retracts his commitment to the original claim.

Disagreements, resistance, and clarification Where do resistance moves fit in the space of 'non-agreeing' responses? First of all, they do not need to express disagreement or correct the hearer. A primary diagnostic for corrections is compatibility with 'no', and we see in the above context that the modal claim is incompatible with 'no':

(2) B′: #No, Diana might be there...

Nor are they necessarily clarification requests (Ginzburg & Cooper 2004 a.o.): while (1B) may intuitively invite some clarification of A's position, it does not address A's utterance itself, but rather A's assumptions. The closest category we are aware of is what Rawlins (2010) called 'conversational backoff triggers', which themselves would seem to be a species of resistance moves:

(3) B″: What if Diana goes?

Rawlins analyzes 'what if' responses as indicating partial acceptance of A's move in the case where the 'if'-clause is false, and re-asking some QUD for the case where it is true. In this example, B accepts that A will skip the party if Diana doesn't go, but checks the case where she does go. While this analysis is plausible for 'what if' questions, it is too strong for epistemic resistance moves in three ways: (i) moves like B's do not ask a question at all, (ii) following a resistance move A's claim seems to be entirely 'on hold' rather than partially accepted, and (iii) the analysis involves taking A to tacitly assume that Diana won't be there (similar to Lewis 1979, Stalnaker 1984 ch. 5), whereas intuitively in cases like (1), A hasn't made any such assumption, or even considered whether Diana might be there.

Point (iii), the *lack of tacit assumptions* is our starting point. We take resistance moves to involve the resistor signaling that the resistee may not have attended to some *issue*, such as whether Diana would be at the party. By drawing attention to an issue that hasn't been publicly attended to or raised in discourse, B signals that they want to ensure appropriate mutual attention before considering A's claim for acceptance or rejection.

Analysis of resistance Following Farkas & Bruce (2010), we take it that assertions do not automatically update a context, but rather introduce a proposal to update the context in a certain way. A resistance move happens after the proposal stage, but before acceptance. At this point, the resistance move introduces a new proposal to update the context, without resolving the first one:

in Farkas & Bruce's terms, the resisting proposal is pushed onto an assertion stack while the first one remains on hold. B's resisting proposal involves an epistemic possibility claim, which has two effects. The first we suggest can be modeled by a standard dynamic treatment of modals, for example (4). That is, 'might ϕ' invites the hearer to perform a test to check if their information state can be consistently updated with φ, and signals that the speaker's own state can be so updated.

(4) $i + \Diamond\varphi = \{w \in i : i + \varphi \neq \emptyset\}$ (Veltman, 1996)

The second effect is our novel proposal: 'might ϕ' *signals that the speaker is attending to the matter of whether* ϕ *holds*, e.g. is attending to the issue $\{\phi, \neg\phi\}$; this ensures that all agents are then publicly attending to the issue. We suggest that this is not part of the lexical semantics of 'might' but rather an effect that occurs whenever a proposition is mentioned. The combination of the normal modal semantics (the lack of speaker commitment to ϕ) and the attentive effect towards the prejacent ϕ is what makes a possibility modal particularly suited to resistance.

(5) **Drawing attention** If a speaker utters a sentence containing (atomic) φ, then the attentive state of any other discourse participant is immediately updated so that they attend to the matter of φ.

Attention shift may have non-monotonic results. In (1), the speaker was aware of the background fact, but simply hadn't considered whether Diana would attend. First they indicate that they have accepted this modal claim (and consequent attention shift) using the particle 'oh'. At this 'resolution' (cf. Yalcin 2011) the interaction of their background preferences about Diana and their choice to go to the party was simply not visible. However, once Alice attends to whether Diana goes, this interaction does become visible, and they change their mind, retracting their initial claim; in this example the retraction is signaled by a continuation that is incompatible with the original claim.

Sketch of formal model for attention shifts We embed the proposal in a Stalnakerian (2014) framework where the common ground consists of propositions that are *commonly accepted* by members of a conversation, paired with a dynamic epistemic logic. Here we provide only the minimal pieces needed to express the analysis of attention shift. A context representation consists of a 'Table' (Farkas & Bruce, 2010) representing proposals for common acceptance, together with the state of individual agents, representing both their informational and attentional states.

To model attention we adopt machinery from Lewis's 1988 theory of *subject matters*.

(6) A **subject matter** $M \subseteq \mathscr{W} \times \mathscr{W}$ is a total equivalence relation over $\mathscr{W}$.

(7) M_1 includes M_2 just in case M_1wv only if M_2wv.

(8) The sum $M_1 + M_w + ...$ is the least subject matter (should this exist) which includes each of $M_1, M_2, ...$ and is included in any other subject matter which includes these matters.

Following Fritz & Lederman (2015) we assume that (a) anyone attending to a subject matter M is attending to every matter included in M, (b) anyone attending to $M_1, M_2, ...$ is also attending to the least subject matter $M_1 + M_2 + ...$ (which exists), and (c) anyone conversational agent is attending to at least one subject matter. Given these assumptions, a discourse participant's attention to subject matters can be modeled by the sum Π of each of the subject matters that she is attending to. Finally, we define visibility for propositions (Yalcin, 2011):

(9) A proposition $p \subseteq \mathscr{W}$ is *visible* at M just in case: Mwv only if $(w \in p$ iff $v \in p)$

In our example scenario, the proposition that Diana will come is not visible at the resolution Π of the initial context, because A is not attending it. When B forces mutual attention to this issue, this proposition becomes visible to A, and A consequently revises his commitments.

Building towards a discourse model, we define the *view* of an agent at a world w as a pair $\langle \Pi(w), \Delta(w) \rangle$ of an attentional state $\Pi(w)$ (a subject matter over $\mathscr{W}$), and an *informational state* $\Delta(w)$, a set of possible worlds that is visible at that attentional state at w. If an agent A is not attending to whether Diana is going to the party in w, then neither the proposition that she does go, nor its negation, will be visible at that agent's attentional state, and so *neither* can be part of that agent's current view at w. We therefore have the power to capture the problematic intuition (iii) that an agent may simply not have considered some issue at all, and has no tacit beliefs about it one way or the other. The result of a resistance move is to draw mutual/public attention towards some issue, possibly leading to revision (our definition of the new concept of 'public attention' is more involved; we do not present it here):

(10) **Mutual attention** The members of *Agt* are *mutually attending* or *mutually sensitive* to M in w just in case for all $a \in Agt$, $\Pi_a(w)$ includes M.

Attentional shifts may involve non-monotonic changes to an agent's views. We will not model exactly *how* such changes happen here, but we will represent them by abstracting away from particular informational states δ to a function $\mathscr{B}(w)(M)$ that, given world w and matter M, gives back a set of worlds that is visible at M. The mechanics of an attentional shift are quite simple – they involve refining the attention state of all agents a in the conversation:

(11) **Drawing attention** in (5) is implemented as $\oslash$:

$$\mathscr{F}_a \oslash \varphi = \begin{cases} \mathscr{B}'_a(w) = \mathscr{B}_a(w) \\ \Pi'_a(w) = \Pi_a(w) \cap M_\varphi \end{cases} .$$

(12) For any atomic sentence φ, let $M_\varphi = \{\langle w, v \rangle : \{w\} \rhd \varphi \text{ iff } \{v\} \rhd \varphi)\}$. ('The matter of whether φ')

If M_φ is already under public attention, this will have no result, but if it is not, then agents' views may change. That is, $\mathscr{B}_a(w)(\Pi_a(w))$ may not be equal to $\mathscr{B}_a(w)(\Pi_a(w) + M_\varphi)$.

The result of all this is that a declarative 'might p' move enforces that all participants are publicly attending to the matter of whether p (because p is mentioned), and consequently propositions that involve p may become visible if they aren't already. A speaker must therefore re-evaluate their commitment to the assertion currently on the table.

Conclusion In this abstract we have introduced a new class of moves in discourse: *resistance moves*. Resistance moves lie somewhere in between outright disagreement, correction, and acceptance, but we have argued are none of the above. Rather, they involve a temporary deferral of a commitment while the context reaches an equilibrium with respect to *attentional state*. In service of this analysis, we have analyzed attention as attention to subject matters, and proposed that attention governs what propositions might be factored into public commitments. When a speaker detects that there may be an attentional mismatch, they are in a form of a defective context (though not necessarily defective with respect to information, as in Stalnaker (1978)), and will attempt to move towards public equilibrium before further commitments are made.

While epistemic possibility modals provide fertile ground for resistance moves, it is part of our proposal that attention and its consequences are a much more general phenomena. This leads to two important avenues for future investigation. First, in several proposals about epistemic modals similar technical machinery has played a key role under different guises (von Fintel & Gillies, 2011; Yalcin, 2011), that we suggest might be reducible to attention. Second, we suggest that many cases of putative disagreement, and cases (Rawlins, 2010) called backoff triggers in general, might involve resistance via drawing attention.

Bibliography

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2009. Attention! *Might* in inquisitive semantics. In *Proceedings of SALT 19*, .

Farkas, Donka & Kim Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27. 81–118.

von Fintel, Kai & Anthony Gillies. 2011. 'might' made right. In Andy Egan & Brian Weatherson (eds.), *Epistemic modality*, 108–130. Oxford University Press.

Fritz, Peter & Harvey Lederman. 2015. Standard state space models of unawareness. manuscript.

Ginzburg, Jonathan & Robin Cooper. 2004. Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and Philosophy* 27. 297–365.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8. 339–359.

Lewis, David. 1988. Relevant implication. *Theoria* 54. 161–237.

Rawlins, Kyle. 2010. Conversational backoff. In *Proceedings of SALT XX*, CLC Publications.

Stalnaker, Robert. 1978. Assertion. In Peter Cole (ed.), *Pragmatics*, 315–332. New York: Academic Press.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge University Press.

Stalnaker, Robert. 2014. *Context*. Oxford University Press.

Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic* 25. 221–261.

Yalcin, Seth. 2011. Non-factualism about epistemic modality. In A. Egan & B. Weatherson (eds.), *Epistemic modality*, 295–332. Oxford University Press.