A matter of trust: Dynamic attitudes in epistemic logic

Rodenhauser, L.B.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Conclusion

This dissertation has explored the concept of a dynamic attitude as a formal representation of an agent’s assessment of the reliability of a source of information. Our formalization of dynamic attitudes (as functions that, given an informational input, map information states to information states) has drawn heavily on existing work in belief revision theory, dynamic epistemic logic and dynamic semantics. We have contributed a framework that has enabled us to explore new research directions. In conclusion, let me selectively highlight the main contributions of the dissertation, and point out some directions that might be taken in future work.

Contributions. The study carried out here has emphasized the importance of going beyond the discussion of specific belief revision policies, mainly in three respects. First, from the perspective adopted here, not only operations that induce belief are interesting objects of study. In this regard, our notion of a dynamic attitude generalizes the notion of a belief revision policy, and a wider range of interesting phenomena comes into view. In Chapter 2, we have seen that we can formalize notions of uniform trust, but also distrust, semi-trust, and mixed trust using our framework. Second, our notion of a fixed point, that was introduced in Chapter 1 and played a central role in the discussion of Chapter 2 and Chapter 3, embodies the idea that dynamic attitudes should be studied in tandem with the propositional attitudes they realize (an idea the roots of which go at least back to the research program outlined in van Benthem (1996)). Rather than arguing that a single belief revision policy is “the right one”, the tandem approach leads to a more pluralistic perspective: depending on the particular target of revision, different policies may be adequate. This perspective has allowed us to clarify what is special about examples of dynamic attitudes that are well-known from the literature (they have natural fixed points! cf. §2.5), and, indeed, in which sense they are unique: Chapter 3 introduced the crucial notion of optimality (§3.2), according to which, roughly, a dynamic attitude is optimal if it realizes its fixed point in a way that adheres to the principle of minimal change. As it turns
Concluding Remark. Our setting has a number of limitations that should be lifted eventually. A recent line of work in dynamic epistemic logic initiated by van Benthem and Pacuit (2011) takes non-total preorders as its starting point; in formal semantics, this seems to have been the dominant practice all along (cf., e.g., Kratzer (1981), Veltman (1996)). Considering dynamic attitudes on non-total preorders would bring the current setting closer to both. Furthermore, for reasons of generality, dynamic attitudes should be studied on infinite preorders, both well-founded and non-wellfounded ones. Finally, it would be interesting to know which dynamic attitudes can be defined by means of the action-priority operator introduced by Baltag and Smets (2008).
Going further, we mention a number of follow-up projects that naturally originate from the work presented here.

A characterization of the canonical propositional attitudes (defined as the propositional attitudes $A$ for which there exists a unique optimal dynamic attitude $\tau$ such that the fixed point $\tau$ of $\tau$ is $A$) in terms of an illuminating set of sufficient and necessary conditions remains to be found. The search may be combined with a more systematic exploration of alternative measures of similarity. §3.7, which focused on the case of simple belief, has merely scratched the surface in this direction.

The preservation results established in Chapter 4 are only the beginning of a wider-ranging investigation. We have already outlined a number of questions in this area, cf. the remarks after the proof of Theorem 76 in §4.2.

The notion of definability of dynamic attitudes could also form the starting point of a more extensive study. As we have pointed out in §5.2.4, the epistemic-doxastic language $L$ we have considered (with operators for the two “knowledges”: defeasible knowledge $\Box$ and irrecovable knowledge $K$) can define strong trust $\uparrow$ and infallible trust $\diamondsuit$, but not minimal trust $\downarrow$. The question just which dynamic attitudes $L$ can define is open. A more ambitious task is to classify different languages of epistemic logic by means of the dynamic attitudes they can define.

Another project is motivated by the notion of dynamic completeness introduced in §2.4.6. Precisely which sets of dynamic attitudes are dynamically complete? A more general (and more vague) question is: just which (types of) sources does an agent need to have at her disposal for her epistemic well-being? We make this question slightly more precise in the next paragraph, which discusses a theme that expands the scope of what we are trying to capture in our formal models.

Dynamics of Dynamic Attitudes. This thesis has explored the notion of a dynamic attitude in some detail, but has been silent on the dynamics of dynamic attitudes. We have explored the space between “informational stimulus” and “epistemic response”, a space in which an agent chooses his response, i.e., decides how to change his mind based on the content of the stimulus, but also depending on how reliable he considers the source of information to be. But we have taken the result of the choice for granted, and have not investigated the circumstances in which an agent may choose to reconsider a choice made earlier. There is a question to be asked how an agent would come to adopt a particular dynamic attitude towards a particular source, and in view of what evidence he would change it. As Annette Baier observed, a trusting agent exposes his own vulnerability: “One leaves
others an opportunity to harm one when one trusts, and also shows one’s confidence that they will not take it” (Baier 1986). So trust should not be bestowed upon one’s sources too easily. But the agent who chooses to ignore all informational inputs, regardless of origin, cannot learn. So there are occasions where we need to give others the power to harm us, by trusting them, hoping that they won’t.

Consequently, agents need to find a balance between “epistemic vigilance” (Sperber et al. 2010) and eagerness for information. Gullibility (trusting everyone) and exaggerated suspiciousness (trusting no one) are two extreme ways of making the trade-off, two examples of possible “meta-attitudes” guiding an agents’ choices of dynamic attitudes, but one would like to know more about the options that lie in-between. Prima facie, it is not even clear what the criteria are for a reasonable mix of dynamic attitudes towards one’s range of sources. A question one may ask is if an agent “too isolated” in the sense that he has “too few” sources at his disposal, or, on the contrary, “over-connected”, in the sense of being exposed to “too many”, potentially conflicting information sources. This is one direction in which the question posed at the end of the previous paragraph may be sharpened. The theme seems also related to recent application of epistemic logic to the theory of social networks (Seligman, Liu, and Girard 2013, Christoff and Hansen 2013).

Going further, an agent’s general policy for assessing the trustworthiness of others needs to take into account the risk of epistemic wounds being inflicted. To indefinitely rely on those who hurt us is a recipe for disaster (in epistemic matters as elsewhere), but to permanently exclude potentially valuable sources of information from consideration may not be in our best interest either. The agent thus faces the need for revising his dynamic attitudes as new evidence comes in. He may choose to be vigilant until a source has proven worthy of trust; or he may lose trust as soon as he has obtained evidence for past deceit. He might change his attitude to a source once another, already trusted source vouches for the former, or discredits the former.