

Learning to trust and trusting to learn: a theoretical framework

Asheley R. Landrum, Baxter S. Eaves Jr, and Patrick Shafto

Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY 40292, USA

Learning from other people requires integrating reasoning about an informant's psychological properties, such as knowledge and intent, with reasoning about the implications of the data the informant chooses to present. Here, we argue for an approach that considers these two reasoning paths as interrelated, reciprocal processes that develop over experience and guide learners when acquiring knowledge about the world.

Learning to trust and trusting to learn: a theoretical framework

We do not discover everything that we know on our own; much is learned from other people [1]. However, information presented by others is not necessarily reliable. In fact, it can be unrepresentative, intentionally misleading, or even completely false. This creates a difficult problem because the situations in which we need information are most often those in which we are unable to assess its quality. Thus, to learn accurate information from others requires integrating inferences about the informant (see [Glossary](#)) with inferences about the information he or she chooses to present.

In this forum, we propose a theoretical framework that goes beyond past literature to account for the interrelated, reciprocal processes involved in learning about and learning from informants ([Figure 1](#)). Under this framework, learners leverage their knowledge about the information presented by informants to draw inferences about informants (i.e., learning to trust). These include inferences about unobserved psychological properties of informants, such as knowledge and intent, which explain why informants are or are not trustworthy. Similarly, under this framework, learners leverage their knowledge about informants to draw further inferences about information presented (i.e., trusting to learn). These inferences not only include, but also go beyond merely accepting or rejecting the information presented. Thus, behavior at any point in time reflects a history of reciprocal inferences regarding whom we have learned to trust and what we trusted to learn.

Our framework is also unique in illustrating how a learner's inferences generalize beyond the specifics of an experience. Under our framework, learners treat an episode in which an informant provides information as potentially connected to stable properties of informants and stable properties of the world. Learners assume that informants'

actions are caused by their unobservable psychological properties, such as the informant's knowledge and intent. Moreover, rather than treating each individual informant as idiosyncratic, informants are treated as members of social categories to which specific inferences may be generalizable. Similarly, rather than treating each bit of information chosen as idiosyncratic or even merely generalizable, information is treated as purposefully selected by an informant to convey information about this instance as well as other, unobserved instances and properties. Thus, unlike previous work, our framework makes explicit a deep relation between specific episodes of information sharing, learners' inferential processes about people and the world, and the reciprocal relation between the episodes and inferences over experience.

Predictions of this framework

At the simplest level, the framework predicts that the information provided by informants should affect whether we ask that individual for information or endorse their assertions in the future. For example, if given the choice between two people, one who has provided correct information in the past and one who has provided incorrect information, learners should choose to ask, endorse information provided by, and attribute knowledge to the previously correct informant. Indeed, these phenomena have been the focus of the empirical and modeling literatures on epistemic trust (e.g., [2]; see [3] for a review and [4] for computational models thereof). Importantly, the framework

Glossary

Informant: someone who selects and presents information for a learner; for example, a teacher.

Intent: a latent psychological property of an informant. Here, intent deals with purpose for which information is selected. For instance, an informant can select information with the intent to help the learner. Conversely, an informant can select information that will mislead the learner.

Knowledge: a latent psychological property of an informant. Here, knowledge (or knowledgeability) deals with the extent to which an informant is knowledgeable about the world and is capable of selecting helpful, accurate information.

Representative: 'the extent to which something is a good example of a category or process' [8]. For instance, a robin is a representative single example of the category 'birds', while a cardinal and an eagle are a representative pair of examples of the category 'birds' (whereas two types of robin would not be). Representative examples reflect intentional selection (and omission) of information to facilitate learning.

Social categories: social categories are groupings of humans for which all members are treated similarly, such as gender, race, occupation, and so on.

Trust: we use the term 'trust' to refer specifically to epistemic trust: whether a learner believes an informant to be trustworthy (see below) and, thus, requests information from or endorses information presented by that informant.

Trustworthy: an informant is trust worthy when he or she is knowledgeable and intends to select information that will lead the learner to the correct answer.

Corresponding author: Landrum, A.R. (LandrumAR@gmail.com).

Keywords: epistemic trust; social learning; cognitive development.

1364-6613/

© 2014 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tics.2014.12.007>

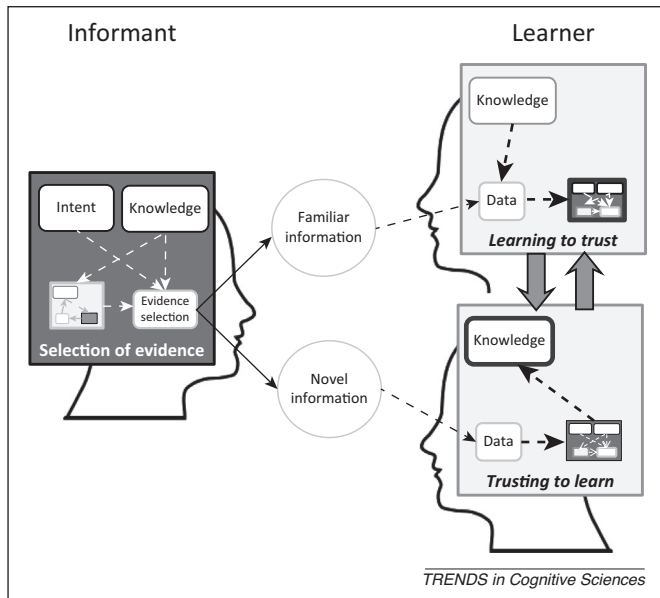


Figure 1. Illustration of theoretical framework for learning to trust and trusting to learn. The image on the left represents the informant, who selects information to present for the learner. The informant's inference about the best evidence to present is influenced by his or her knowledge, intent, and beliefs about the learner (as indicated by broken arrows). The images on the right represent the learner's inference based on the presentation of familiar and novel information. The top-right image shows inferences based on familiar information, which represent learning to trust. In this case, the learner uses the familiar information selected by the informant to update his or her beliefs about that informant (as indicated by broken arrows), including beliefs about the informant's knowledge and intent, and about informants of relevant social categories more generally. The bottom right image shows inference based on novel information, which represents trusting to learn. In this case, the learner uses his or her beliefs about the informant, including beliefs about that informant's knowledge and intent, which are informed by beliefs about the social categories to which he or she belongs, to update his or her knowledge about the world. The double arrows between learning to trust and trusting to learn represent the reciprocal nature of these processes.

makes several predictions that specific episodic information should generalize beyond the specific information provided and beyond the specific person providing the information.

Learners recognize the importance of an informant's intentions

First, our theoretical framework predicts that learners consider not only knowledgeability, but also an informant's intent, when considering evidence presented by that informant. This includes an informant's intentions to be helpful or deceptive. For instance, recent evidence suggests that children heavily consider niceness, even when it conflicts with knowledgeability, when deciding whose information to endorse (e.g., [5,6]). Similarly, older preschoolers prefer to accept advice from helpful informants over ones described as liars or trickers (e.g., [7]). Moreover, recent computational modeling of epistemic trust suggests that informants' knowledgeability alone cannot account for the empirical results in standard epistemic trust tasks. Instead, changes in how learners behave on some tasks are best explained by changes in their beliefs about the informant's intent [4].

Learners will interpret information as representative

Second, our theoretical framework predicts that learners' inferences go beyond evaluating the veracity of information;

they also will infer that the information from knowledgeable and helpful informants is purposefully chosen and, therefore, representative of the true concept (see [8]). This prediction, too, has been supported by literature on play and learning. For example, studies show that children will restrict their play to the one demonstrated function of a multifunction toy, believing a knowledgeable teacher's instruction to be representative [9]. Similarly, children will repeat unnecessary actions to activate a causal mechanism (e.g., a light), believing that all actions were necessary because all were demonstrated [10].

Learners will generalize beliefs about informants

Third, our framework predicts that learners will generalize inferences about knowledge and intent of one individual to other, unfamiliar members of a social category with whom the learners have had no prior experience. The literature illustrates such generalizations based on at least two types of social category. One such social category is expertise. Learners can use information about an unfamiliar individual's domain of expertise to make inferences about what he or she is likely to know. Recent research shows that children will attribute knowledge to, and endorse information from, people described as having relevant expertise [5].

A second social category is group membership. Group membership is a social category wherein people are grouped by inherited and selected features, such as ethnicity and team membership. Recent research demonstrates that children's experience with in-group members affects their decisions about whom to trust for new information, even when group membership is based on a seemingly superficial characteristic (e.g., a minimal group paradigm [11]). This shows that children are generalizing their trust to completely unfamiliar informants, based solely on their experience with informants of the same group.

An extreme example of social category-based inferences can be seen in generalizations from experiences with specific informants to all informants. Given that children have so much experience with trustworthy informants, the expectation that informants are trustworthy may become so strong and stable that it is difficult to overcome. Recent research has shown that children continue to trust in information provided by informants who have been wrong on several consecutive occasions when the children have no alternative informants or opportunities to access information independently [12].

Superficial characteristics are cues to trustworthiness

Lastly, our framework predicts that, as learners engage with trustworthy and untrustworthy informants over time, they may infer that certain observable cues are indicative of the unobservable psychological properties of knowledgeability and intent. For instance, recent research has found that children seem to use observable informant features, such as attractiveness [13], accent [14], and age [15], as well as many others, when deciding whom to trust. Although these features may not be causally or logically related to an informant's knowledge or intent, they may be statistically related to these otherwise unobservable psychological properties of interest and, therefore, are useful tools in predicting who is trustworthy.

Concluding remarks

Recent research has shown a marked shift from a nonsocial view of learning to a view that is deeply social. The critical challenge in creating a broad theoretical framework lies in explaining how learning in social situations differs from learning in asocial situations. Our theoretical framework explains how the automatic psychological reasoning involved in trust and learning creates a dynamic process of social learning that evolves over time. Implications of this work include testable claims about the effects of reasoning about other people for learning, a unified framework for understanding how beliefs about people affect learning and how learning affects beliefs about people, and a dynamic perspective on learning from and about people that can be used to model effects of experience on learning and development. Research is ongoing, but it is clear that any complete account of learning must explain how learners deal with the joint problems of learning to trust and trusting to learn.

Acknowledgments

The authors would like to thank Kelley Durkin, Nick Searcy, and Campbell Rightmyer as well as the CoCoSci Lab for their helpful feedback. This work was supported in part by a grant from the National Science Foundation (DRL 1149116) to P.S.

References

- 1 Csibra, G. and Gergely, G. (2009) Natural pedagogy. *Trends Cogn. Sci.* 13, 148–153
- 2 Koenig, M.A. and Harris, P.L. (2005) Preschoolers mistrust ignorant and inaccurate speakers. *Child Dev.* 76, 1261–1277
- 3 Mills, C.M. (2013) Knowing when to doubt: developing a critical stance when learning from others. *Dev. Psychol.* 49, 404–418
- 4 Shafto, P. et al. (2012) Epistemic trust: modeling children's reasoning about others' knowledge and intent. *Dev. Sci.* 15, 436–447
- 5 Landrum, A.R. et al. (2013) When do children trust the expert? Benevolence information influences children's trust more than expertise. *Dev. Sci.* 16, 622–638
- 6 Lane, J.D. et al. (2013) Informant's traits weigh heavily in young children's trust in testimony and their epistemic inferences. *Child Dev.* 84, 1253–1268
- 7 Mascaro, O. and Sperber, D. (2009) The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition* 112, 367–380
- 8 Tenenbaum, J.B. and Griffiths, T.L. (2001) The rational basis of representativeness. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (Moore, J.D. and Stenning, K., eds), pp. 1036–1041, Lawrence Erlbaum Associates
- 9 Bonawitz, E. et al. (2011) The double-edged sword of pedagogy: instruction limits spontaneous exploration and discovery. *Cognition* 120, 322–330
- 10 Buchsbaum, D. et al. (2011) Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition* 120, 331–340
- 11 Elashi, F.B. and Mills, C.M. (2014) Do children trust based on group membership or prior accuracy? The role of novel group membership in children's trust decisions. *J. Exp. Child Psychol.* 128, 88–104
- 12 Vanderbilt, K.E. et al. (2014) In the absence of conflicting testimony young children trust inaccurate informants. *Dev. Sci.* 17, 443–451
- 13 Bascandziev, I. and Harris, P.L. (2014) In beauty we trust: children prefer information from more attractive informants. *Br. J. Dev. Psychol.* 32, 94–99
- 14 Kinzler, K.D. et al. (2011) Children's selective trust in native-accented speakers. *Dev. Sci.* 14, 106–111
- 15 VanderBorgh, M. and Jaswal, V.K. (2009) Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant Child Dev.* 18, 61–71