Trends in Cognitive Sciences

CellPress

Forum

Cooperative communication as belief transport

Patrick Shafto,[1,*] Junqi Wang,[1] and Pei Wang[1]

Recent research formalizes cooperative communication as belief transport using the mathematical theory of optimal transport. This formalization allows rigorous *a priori* analysis of the statistical and ecological properties of models of cooperative communication, unification of prior models and analysis of their differences, and promising directions for future research.

Cooperative communication

Cooperative communication is mutual theory of mind reasoning between a pair of agents, one who selects data to convey a hypothesis and a second who infers a hypothesis given data, in which both agents have the shared goal of successfully transmitting beliefs (Figure 1). Cooperative communication plays a central role in theories of cognition, culture, and human–machine interaction. For cognition, cooperative communication is central to theories of effectiveness of language and the efficiency of learning [1,2]. For culture, cooperative communication is invoked to explain accumulation of knowledge over generations [2]. For human–machine interaction, cooperative communication represents the frontier by which humans and machines may work together more seamlessly toward societal goals [3]. Although central, whether cooperative communication can live up to these promises remains unanswered.

Answering theoretical questions about the implications of cooperative communication requires making statements that hold over the many possible contexts in which cooperative communication takes place. Computational models of cooperative communication, however, are developed and tested within specific experimental paradigms. For example, experimenters choose which and how many objects there are and the possible words (or actions) one may use to communicate, choices that tend toward small numbers of simple objects, referred to by a single word about which there can be little confusion or disagreement. For instance, one classic study asked participants whether they would use the word blue or circle to pick out a blue circle from an array including a blue square and a green square [1]. These models are tied to specific methodological choices related to which and how many objects and words, the distinctiveness of their perceptual features, and the shared understanding of these choices between participants. As a result, such models are ill-suited to making general statements, which would require simulating possibilities over a prohibitively large space of possibilities. Thus, there are fundamental limitations to the conclusions that one can draw, regarding whether cooperative communication is effective for transmitting information, efficient for learning, can explain knowledge accumulation or support human–machine cooperation, from current models. In this paper, we present an informal introduction to **belief transport** (see Glossary), a recently proposed mathematical theory that addresses these challenges, discuss recent advances enabled by the theory, and broader implications.

Belief transport

Optimal transport is a field within mathematics concerned with finding transport plans for moving resources. Suppose you are a delivery company owner who has trucks that pick up bread from bakeries and deliver it to cafes. Bob's bakery has ten loaves and Bonnie's bakery has five and you need to deliver ten to Cara's cafe and five to Carl's. If the cost of transporting between the bakeries and cafes is equivalent, then it is clear that you could simply deliver Bob's ten loaves to Cara, and Bonnie's five loaves to Carl. However, if instead Cara and Carl need seven and eight loaves, respectively, and the cost of delivery between the bakeries and cafes differs, finding the best plan for moving loaves from bakeries to cafes becomes complicated. Optimal solutions to resource allocation problems such as this are the subject of the mathematical theory of optimal transport.

Recent research [4] has shown how this framework can be used to understand transport of beliefs via data (Box 1) by proving equivalence between cooperative communication and optimal transportation plans [4]. Using this formulation, past cognitive models of cooperative communication [1,3,5,6] are approximate solutions to the problem of belief transport [4]. Indeed, belief transport can be extended to analyze sequential interactions [7] such as are typical in human conversation and in cultural transmission. The results have implications for understanding statistical and ecological validity of models of human cooperative communication, unifying and comparing models, and suggest new research directions.

Glossary

Belief transport: applies the idea of optimal transport to cooperative communication by viewing beliefs as a resource to be moved to observable data, where the costs are given by a probabilistic model.
Naïve utility calculus: a computational theory in cognitive development that describes commonsense psychological reasoning [5].
Optimal transport: a mathematical theory that describes optimal plans for moving resources from one configuration to another, given the cost of the moves.
Pedagogic-pragmatic value alignment: a computational theory in robotics for ensuring that robots' objectives match those of their human users [3].
Rational speech act: a computational theory in linguistic pragmatics describing how speakers and listeners choose words and interpret their referents [1].
Sinkhorn scaling: an algorithm for computing solutions to optimal transportation problems, which is equivalent to cooperative theory of mind reasoning.

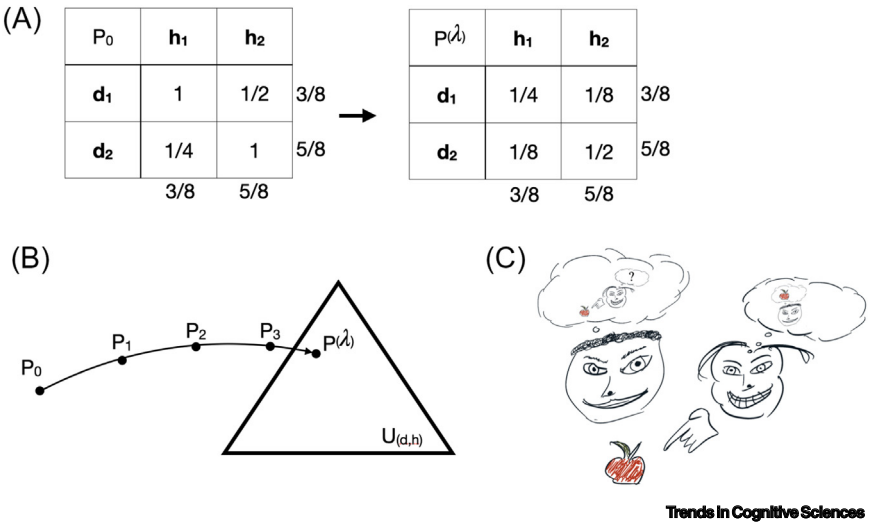

Figure 1. Schematic depiction of belief transport. Belief transport computes an optimal plan for transmitting beliefs via data. (A) Belief transport plans can be computed using Sinkhorn scaling, which finds the optimal plan by alternating between adjusting rows and columns to meet the desired distributions (see Box 1 for further details). (B) The optimal plan satisfies the desired distributions on data and hypotheses, and is closest subject to the choice of λ. (C) Belief transport is a mathematical formalization of cooperative communication as a social inference, encompassing prior models of communication. One agent (right) reasons about which data to select to induce the intended hypothesis in the second agent (left). The second agent reasons about which hypothesis would be best induced by which data in order to infer the intended hypothesis. Communication is successful when the second agent correctly infers the intended belief.

Consistency of inference and robustness to violations of common ground

A statistical model is said to be consistent when, for every hypothesis, the model converges to the target hypothesis as more and more data are sampled from it. Intuitively, consistency means that given enough data the model does not confuse one hypothesis for another. The cultural ratchet hypothesis suggests that people accumulate knowledge over generations by cooperatively communicating. Minimally, this requires that, over generations, learners would accumulate knowledge and eventually converge to the truth, given one had a teacher with knowledge and perfect transmission of knowledge between successive learners. This minimal analysis of cultural accumulation of knowledge is precisely a question about consistency.

Box 1. Formalization of belief transport

Belief transport seeks the optimal plan, $P^{(\lambda)}$, for transmitting beliefs, **h**, via data, **d**. Optimal plans are those that minimize the cost of transportation, C, and depending on the parameter λ, uncertainty about the plan, $H(P)$. $U(\mathbf{d}, \mathbf{h})$ represents the set of valid plans, which couple **h** and **d**. Formally,

$$P^{(\lambda)} = \underset{P \in U(\mathbf{d},\mathbf{h})}{\arg\min} \left\{ \langle C, P \rangle - \frac{1}{\lambda} H(P) \right\}, \quad (1)$$

where $\langle C, P \rangle$ is the sum of the elementwise product, λ is a parameter that controls the strength of entropy regularization, and $H(\cdot)$ is entropy. For communication, a natural way of setting the cost is via a generative model. For the teacher, $C^{T}_{i,j} = -\log P_L(h_j \mid d_i) - \log P_T(d_i)$, and symmetrically for the learner.

Sinkhorn scaling [8] is a method of computing optimal plans [9], which involves iterative row and column normalization of the initial matrix $P_0 = e^{-\lambda C}$ to arrive at the solution $P^{(\lambda)}$ [9]. Sinkhorn scaling is equivalent to cooperation between probabilistic agents [10] for $\lambda = 1$ and closely related to **rational speech act** [1], **naïve utility calculus** [5], machine teaching [6], and **pedagogic-pragmatic value alignment** [3] (see [4] for derivation).

Belief transport is provably consistent [7] and therefore cooperative communication can theoretically explain how people communicate successfully and accumulate knowledge over ontogeny and phylogeny.

Common ground is the assumption that the communicating agents know the other's beliefs and share an understanding of how data are related to hypotheses. Intuitively, it is clear that in ecologically valid settings common ground can never be exactly met; one can never know another's mind exactly. Prior models of cooperative communication assume common ground because there are infinite ways by which it may be violated. Yet, for this same reason, prior models are ill-suited for ensuring that predictions are robust to violations of common ground; this would require instantiating and testing the myriad possibilities. Belief transport is provably robust to violations of common ground and has analyzed ways in which approximate models introduce vulnerabilities using mathematical properties of Sinkhorn scaling [4]. Thus, belief transport provides a mathematical understanding of when and why cooperative communication can ensure reliable belief transport in ecologically valid settings.

Unification and comparison of models

Recursive theory of mind reasoning is a common thread across models of cooperative communication. Models differ along two primary directions: the number of steps of theory of mind recursion they propose, and the degree to which data selection by the communicating agent minimizes uncertainty. Belief transport unifies these approaches and allows detailed, general comparison of their implications.

Sinkhorn scaling is an algorithm that solves for optimal belief transport plans alternating between making the rows add to the desired total and making the columns sum to the desired total by dividing by the respective sum. Sinkhorn scaling is provably equivalent [4] to the model of intuitive

pedagogical reasoning [10]. Further, rational speech act theory [1], a model that uses one or a few steps of recursive reasoning to explain pragmatic inferences by speakers and listeners, approximates belief transport in a precise statistical sense and pedagogic-pragmatic inference, which proposes robots that select data that maximize the probability of the desired hypothesis [3], approximates unregularized belief transport when uncertainty is minimized (i.e., λ is large) [4]. Indeed, models across literatures [5,6,11] can be viewed as approximations of belief transport, opening up the possibility of systematically comparing theories in a single framework to better understand their relative predictions.

Indeed, one can investigate whether, how, and to what degree existing models differ in their predictions [4]. Increasing the number of recursive steps has modest effects on increasing the effectiveness of communication in the presence of perfect common ground; however, with violations of common ground, increasing the number of recursive steps yields greater effectiveness. They also find that more greedy selection of data, as represented by large values of the λ parameter, yields qualitative decreases in robustness to violations of common ground. Specifically, λ=1, which represents a form of probability matching, is ideal in some sense: values below 1 lose information and values above 1 introduce sensitivity to violations of common ground. The mathematical tools provided by belief transport allow unification and comparison of models, including never before proposed approaches.

New directions and implications

There are many algorithms for obtaining belief transport plans. We previously noted that the Sinkhorn scaling algorithm is equivalent to using theory of mind recursions in cooperative communication. Another algorithm for obtaining belief transport plans is based on gradient descent. Thus, at the algorithmic level, belief transport highlights potential connections between probabilistic models and neural networks and indicates exciting theoretical possibilities and alternative testable models of approximate inference.

Beyond basic research, cooperative communication between humans and machines is increasingly important in society. Explainable artificial intelligence (AI) is one such example, in which the goal is to explain inferences of AI to a human user. This is challenging because the most high-performing classifiers, deep neural networks, are opaque even to AI experts, which limits applicability to high stakes domains where ethical and legal considerations require auditability. Explainable AI can be formalized as approximation of cooperative communication, in which the goal is for the AI to explain itself to the human by selecting important or influential data [12]. By connecting explainable AI to models of cooperative communication in cognitive science, they also open the door to theoretical analysis of which AIs are likely to be more or less explainable, and why, through the lens of belief transport. In this and other practical domains, theoretical guarantees are an important tool for ensuring that whatever methods we use to explain AI are robust.

Concluding remarks

Advances in science occur when new tools shed light on classic questions. Cooperative communication is foundational across many domains of cognition and is increasingly important in society, yet we have lacked tools to systematically analyze theoretical claims and guarantee robust performance. By unifying existing proposals, belief transport enables the systematic comparison of strengths and limitations, sheds light on connections across theories, and suggests paths toward empirical and theoretical progress in understanding cooperative communication and its role in cognition, language, and culture, and practical progress in designing and implementing AI that are effective partners in solving important societal problems.

Declaration of interests

No interests are declared.

[1]Rutgers University, Newark, NJ, USA

*Correspondence:
patrick.shafto@rutgers.edu (P. Shafto).

https://doi.org/10.1016/j.tics.2021.07.012

References

1. Frank, M.C. and Goodman, N.D. (2012) Predicting pragmatic reasoning in language games. *Science* 336, 998
2. Tomasello, M. *et al.* (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691
3. Fisac, J.F. *et al.* (2020) Pragmatic-pedagogic value alignment. In *Robotics Research*, pp. 49–57, Springer
4. Wang, P. *et al.* (2020) A mathematical theory of cooperative inference. In *Advances in Neural Information Processing Systems Pre-proceedings (NeurIPS 2020)* (Larochelle, H. *et al.*, eds), PMLR
5. Jara-Ettinger, J. *et al.* (2016) The nave utility calculus: Computational principles underlying commonsense psychology. *Trends Cogn. Sci.* 20, 589–604
6. Zhu, X. (2015) Machine teaching: An inverse problem to machine learning and an approach toward optimal education, Proceedings of the AAAI Conference on Artificial Intelligence 29. Hilton New York, NY, USA
7. Wang, J. *et al.* (2020) Sequential cooperative Bayesian inference. In (DauméIII, H. and Singh, A., eds), pp. 10039–10049, Proceedings of Machine Learning Research (PMLR 2020), Vienna, Austria
8. Sinkhorn, R. and Knopp, P. (1967) Concerning nonnegative matrices and doubly stochastic matrices. *Pac. J. Math.* 21, 343–348
9. Cuturi, M. (2013) Sinkhorn distances: lightspeed computation of optimal transport. In *NIPS* (2), pp. 4
10. Shafto, P. *et al.* (2014) A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cogn. Psychol.* 71, 55–89
11. Eaves, B.S., Jr *et al.* (2016) Infant-directed speech is consistent with teaching. *Psychol. Rev.* 123, 758
12. Yang, S.C.H. *et al.* (2021) Mitigating belief projection in explainable artificial intelligence via Bayesian teaching. *Sci. Rep.* 11, 9863