Mindshaping the world can make mindreading tractable: Bridging the gap between philosophy and computational complexity analysis

Andrea Zeppi (azeppi@unime.it)
Department of Cognitive Science, University of Messina
Messina, Italy

Mark Blokpoel (m.blokpoel@psych.ru.nl)
Department of Artificial Intelligence, Radboud University
Nijmegen, The Netherlands

Abstract

It is often assumed that the socio-cultural context positively influences mindreading performances. Among the available theories, mindshaping is proposed to consist of cultural mechanisms that make the social domain homogeneous and, hence, easier to interpret. Proponents of the mindshaping hypothesis claim that homogeneity is responsible for the computational tractability of mindreading, which is otherwise intractable. In this paper, we examine this core claim of mindshaping and investigate how homogeneity influences mindreading tractability. By taking action understanding as a case-study for mindreading, we formally operationalize mindshaping homogeneity in different ways with the goal of bridging the gap between informal claims and formal (in)tractability results. The analysis shows that only specific combinations of homogeneity may lead to tractable mindreading, whilst others do not. Additionally, the analysis reveals the possibility of a yet undiscovered mindshaping mechanism.

Keywords: mindshaping; mindreading; computational intractability; culture; goal inference; conceptual/philosophical analysis; computational modeling

Introduction

The ability to understand what motivates other people's behavior is often considered a defining capacity of human cognition. Theories of this *mindreading* capacity, however, are challenged with explaining how humans can interpret behaviors in a timely manner, because the available theories are often computationally intractable (Alechina & Logan, 2010; Apperly, 2010; Zawidzki, 2013). As Gigerenzer and colleagues proposed:

> *"The computations postulated by a model of cognition need to be tractable in the real world in which people live, not only in the small world of an experiment with only a few cues. This eliminates NP-hard models that lead to computational explosion, ..." Gigerenzer (2008)*

Given that mindreading is performed in a complex, real-world socio-cultural environment (Adams et al., 2010; Perez-Zapata, Slaughter, & Henry, 2016; Tomasello, Carpenter, Call, Behne, & Moll, 2005), it stands to reason that intractable (NP-hard) theories of mindreading cannot explain how humans can perform the computations postulated by the theory quickly.

In an attempt to address this theoretical paradox, Zawidzki has proposed that the socio-cultural environment plays a key role. Zawidzki proposes that this environment is shaped by agents themselves so that mindreading can be tractable (Mameli, 2001; Zawidzki, 2008, 2013). Introduced as the *mindshaping hypothesis*, this claim entails a collection of (social) cognitive and evolutionary mechanisms that bring structure to the environment.

In this paper, we assess the potential that the mindshaping hypothesis has to solve the intractability paradox of mindreading. Given that computational (in)tractability is a well-defined mathematical property of computational-level theories (Marr, 1982; van Rooij, 2008), a bridge will have to be built between Zawidzki's informal theoretical contributions and formal complexity-theoretic results. We propose that such a bridge can be built by taking action understanding as a special-case proxy for evaluating how mindshaping mechanisms may pave the way for tractable mindreading.

In order to do this, we take two steps. First, we analyze the mindshaping hypothesis and extract specific claims about the effects that mindshaping mechanisms may have on the structure of the socio-cultural environment and consequently the (in)tractability of mindreading. Second, we assess the plausibility of the claims identified in the first step by operationalizing the mindshaping structuring effects in a computational-level model of action understanding (viz., Bayesian inverse planning; Baker, Saxe, & Tenenbaum, 2009; Baker, Tenenbaum, & Saxe, 2008). This allows us to relate mindshaping effects to formal (in)tractability results (Blokpoel, Kwisthout, van der Weide, Wareham, & van Rooij, 2013). The analysis will show that only certain combinations of mindshaping effects lead to tractable mindreading, whilst other effects do not. Furthermore, the analysis also suggests the possibility of a novel effect that is necessary for tractability, which may lead to the discovery of new mindshaping mechanisms.

Mindreading as abductive inference

Several theoretical accounts of mindreading have been proposed. Fast and frugal heuristics theories (Chater, Oaksford, Nakisa, & Redington, 2003) conjecture that humans can understand what motivates other people's behavior through simple cue-based rules. Simulation theory

(Goldman, 2006) conjectures that we understand external behaviors by the means of mental simulations. In this paper, we focus on a third hypothesis proposed by Zawidzki. By rejecting modularity, acknowledging domain-generality and the relevance problem (Fodor, 1983; Heal, 1996) for social reasoning, Zawidzki implicitly accepts isotropy and with it a mindreading account that is inferential in nature (Zawidzki, 2013).[1] Under this view, mindreading can be construed as a mapping from an observed *socio-cultural environment* (consisting of observed behaviors, actions and context) and *social knowledge* to the *intentional attributions* that best explain the observed social environment. Unfortunately, with the notion that mindreading is inferential also comes intractability.

Zawidzki attributes intractability of mindreading to the problem of holism/isotropy. Because in principle any information that a person has might be relevant for any inference that is made, every possibility must be considered (Fodor, 2001). The intractability of inferential mindreading is corroborated by the fact that many of our best theories of inferential cognitive capacities are computationally intractable (NP-hard or worse) (Cherniak, 1986; Frixione, 2001; Levesque, 1988; Thagard & Verbeurgt, 1998; Tsotos, 1990; van Rooij, 2008; van Rooij & Wareham, 2012).

An intractable theory makes unrealistic (exponential or worse) demands on computational resources (van Rooij, 2008). Hence, such a theory cannot satisfactorily explain why people can 'mindread' as quick as they do (see Table 1). This leads to the paradox mentioned in the introduction. However, rather than rejecting the inferential mindreading account altogether, the paradox may be resolved if the effects of mindshaping mechanisms are adequately fleshed out. In the next section, we discuss the different possible structuring effects that mindshaping mechanisms may have on the environment.

Deconstructing mindshaping

First proposed by Mameli (2001) and later developed by Zawidzki (2009; 2008; 2013), the mindshaping hypothesis proposes that the success of human social cognition is explained by the (evolutionary) development of behavioral mechanisms that "shape our socio-cultural environment in ways that make coordination exponentially more tractable" (Zawidzki, 2013). Examples of mindshaping mechanisms are: imitation, over-imitation, the chameleon effect (Chartrand & Bargh, 1999), pedagogy, norm following and self-constituting narratives (Zawidzki 2013). Although these mechanisms are individually quite different, they all implement a form of social expectancy and conformity mechanism that 'mindshape' both the socio-cultural environment and social knowledge through biased transmission and selection of behaviors. Under the assumption that mindreading is a capacity that operates on the social environment and social knowledge, mindshaping may potentially have a positive effect on the tractability of mindreading. To assess this claim, however, it is necessary to characterize in more detail the different kinds of effects that mindshaping may have on the input of mindreading.

Table 1: An illustration of polynomial and exponential time requirements. The input size corresponds to the size of the representation of the input (e.g., the observed social environment and social knowledge encoded in a Bayesian network). The other columns illustrate the difference between tractable (i.e., polynomial time) intractable (i.e., exponential time or worse). The time required to compute intractable theories quickly outgrows the age of our universe.

Input size *n*	**Polynomial time required n^2**	**Exponential time required 2^n**
5	0,25 msec.	0,32 msec.
10	1 msec.	10 msec.
20	4 msec.	10,5 sec.
50	25 msec.	130312 days
100	0,10 sec.	4×10^{17} years
250	0,63 sec.	$5{,}7 \times 10^{62}$ years
500	2,50 sec.	$1{,}0 \times 10^{138}$ years

Unfortunately, the literature only vaguely provides such a characterization of the effects of mindshaping, which is thought to 'homogenize' the social environment and knowledge to make mindreading easier. Such a characterization is insufficient as it states only the ultimate effect of mindshaping (i.e., tractability of mindreading), which is exactly that which needs to be explained. In order to unravel if and how mindshaping can render mindreading tractable, we need to understand two things. First, we need to understand that by putting constraints on the input of a mindreading, those constraints may render mindreading tractable. Second, we need to understand how mindshaping can implement such constraints through homogeneity.

How homogeneity should affect mindreading

The main claim put forth by mindshaping is that mindshaping mechanisms positively affect the *reliability* of mindreading (Zawidzki, 2013). The term reliability, however, conflates two different meanings: accuracy and tractability. In order for mindreading to be reliable, inferred propositional attitudes need to be good and the computations need to be performed in a short amount of time (tractability). Perhaps counterintuitively, accuracy does not always cause intractability. An intractable function can be extremely inaccurate and it is also possible for a tractable function to,

[1] In a recent book chapter (Zawidzki, *forthcoming*), Zawidzki takes a stronger stance on mindshaping. Although Zawidzki seemingly proposes mindshaping as a separate alternative to inferential mindreading, in doing so one may throw out the baby with the bathwater. If, as originally proposed, we can show how Mindshaping mechanisms can render inferential mindreading tractable, it would strengthen the plausibility of the mindshaping account whether it is separate of mindreading or not.

instead, be accurate. Even approximate accuracy (compared to e.g., optimality) does not necessarily grant tractability (van Rooij & Wareham, 2012). It seems, therefore, that the reason mindreading is tractable for humans lies not in mindreading trading off accuracy, but in the homogeneity effect that mindshaping has.

Research has focused on characterizing the phylogenetic, ontogenetic and cultural evolution of human social cognition (Mameli, 2001; Zawidzki, 2008, 2013). Although the computational details are underspecified, the mindshaping hypothesis clearly states that the tractability of mindreading is not obtained by altering the mindreading capacity, but by changing the socio-cultural environment and social knowledge on which mindreading operates. Mindshaping mechanisms are hence not modules (Zawidzki, 2009, 2013). Therefore, for the purpose of assessing the claim about mindreading tractability, they can be considered complementary to mindreading. This allows us to abstract away from the evolutionary mechanisms that underlie mindshaping and focus on their homogeneity effect.

The solution for tractable mindreading lies "not *within* human mind readers, but, rather, *outside* of them" (Zawidzki, 2009). The idea is promising, since it is known that some intractable functions $f{:}I \rightarrow O$ can be tractable when their input domain is constrained $f'{:}I' \rightarrow O$, where $I' \subset I$ (Downey & Fellows, 1999; van Rooij, 2008). These constraints can be the result of naturally occurring or 'mindshaped' structure in the world. They are formally defined as restrictions on properties of the input of computational-level models, called parameters (e.g., see Figure 1). When the tractability of a function is obtained through such restrictions it is said to be fixed-parameter tractable for that subset of the input. Fixed-parameter tractability, however, is a formal, mathematical property of computational-level models. In order to assess if homogeneity can render mindreading tractable, we need a formal computational-level model of mindreading. Although such an account does not yet exist for mindreading in general, we can investigate the tractability claim using a special-case capacity for mindreading. In the next section, we take (inferential) action understanding as a special case of (inferential) mindreading and present possible ways of operationalizing homogeneity in Bayesian inverse planning (Baker, Saxe, Tenenbaum 2009). We then show that only certain combinations of homogeneity effects render Bayesian inverse planning tractable, whilst other do not.

Bridging homogeneity to tractability of Bayesian inverse planning

The ability to understand what goals underlie the actions of others is a prime example of the human capacity to mindread. In the Bayesian inverse planning model (Baker et al. 2009), action understanding is characterized as *inferring the most probable goal* given *observed social behavior*. In other words, a mapping from an observed socio-cultural environment (consisting of observed behaviors, actions and context) and social knowledge (about planning) to the intentional attributions (goals) that best explain the observations. Table 2 compares the input and output domains of mindreading and Bayesian inverse planning and Figure 1 illustrates the Bayesian network that underlies Bayesian inverse planning.

Figure 1. In Bayesian inverse planning knowledge about planning is represented by state, action and goal variables (circles) and the probabilistic dependencies between them (arrows). Each variable has a domain (boxes). The input of the model consists of such a network and observed states and actions (gray variables). The output is the most likely value assignment to the goal variables. Several parameters are: The number of goals $|G|$, the number of observed actions $|A|$, the maximum number of values a goal variable can have g and the maximum number of values an action variable can have a.

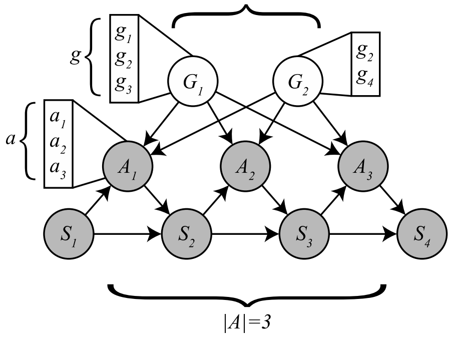

Table 2. Comparing the input and output of the mindreading capacity with those of the special case Bayesian inverse planning model.

	Mindreading	**Bayesian inverse planning**
Input	*observed socio-cultural environment* and *social knowledge*	*observed states, actions* and *knowledge about planning encoded in a Bayesian network*
Output	*intentional attributions*	*most probable goals, given the input*

Like Bayesian inference (Chater, Tenenbaum, & Yuille, 2006; Martignon & Hoffrage, 2002), Bayesian inverse planning is computationally intractable in general (Blokpoel et al., 2013). This is consistent with the idea that (inferential) mindreading is computationally intractable too. Blokpoel et al. (2013), however, used formal analysis to prove that when certain constraining assumptions are made on the input of Bayesian inverse planning, it becomes tractable. To investigate whether or not a more homogeneous socio-cultural environment and more homogeneous social knowledge may lead to tractable mindreading, we have to build a bridge all the way to Bayesian inverse planning. We first start by illustrating possible interpretations of homogeneity. We then operationalize these interpretations in Bayesian inverse planning, and finally relate the operationalized homogeneity effects with known computational tractability results.

Interpreting homogeneity effects

There are two types of homogeneity that are consistent with the mindshaping literature: *Cognitive homogeneity* and *mindshaping homogeneity*. This analysis is a first attempt and by no means exhaustive, i.e., more homogeneity effects may be postulated/discovered in the future. For example, our analysis will point to possible novel homogeneity effect that is not yet covered by a mindshaping mechanism.

Cognitive homogeneity

Since all humans have similar biological and cognitive systems, one could argue that humans also share the majority of their propositional attitudes. For example, if all humans behave rationally, they all have the same (rational) bias when deciding how to act to achieve a goal.

Cognitive homogeneity mechanisms may result in a population where sets of intentions overlap a lot (i.e., most of people's possible intentions are shared; CH-S). This, however, does not necessarily restrict the number of possible intentions, as the shared set can still be very large.

Furthermore, Zawidzki argues that tractability of mindreading cannot be achieved only by cognitive homogeneity. This seems to make sense from a computational perspective as well. Even if, for example, all humans mindread 'rationally' this does not explain why mindreading is tractable. And even if all humans share most of their intentions, that set can still be extremely big. Other restrictions are needed to provide any computational benefits.

Mindshaping homogeneity

Mindshaping homogeneity is effected by a set of mechanisms either cognitive, cultural or evolutionary that decrease the heterogeneity of the socio-cultural environment and social knowledge. For example, by having a culture that keeps reinforcing the same knowledge and behaviors through pedagogy, norms enforcement and imitation (Zawidzki 2013), the knowledge in that population can become more and more homogeneous over generations.

Mindshaping homogeneity can be conjectured to result in the following restrictions as a consequence of biased transmission of knowledge over generations:

- Biased transmission can restrict on the number of available behaviors in a population (MH-B);
- Biased transmission can make social knowledge (i.e., the relations between behaviors and intentions) less ambiguous (MH-A).
- Mindshaping mechanisms resulting in ritualization phenomena may limit, regardless of the total number of possible actions available, the number of executed and observed actions (MH-R).
- Habitualization and culture codification may further limit the complexity of what people can achieve to facilitate social understanding (MH-C).

In the mindshaping literature, the focus has been on identifying the nature of the mindshaping mechanisms that lead to homogeneity. Here instead, we focus on the actual contribution of mindshaping mechanisms and the relative homogeneity. Hence, we assume the validity of mindshaping as a starting point, together with homogeneity, and investigate if their effect can render an intractable model of a mindreading capacity tractable.

Operationalizing homogeneity effects in Bayesian inverse planning

Taking Bayesian inverse planning as our case-study we can now investigate the possible input-restrictions that can result from the mindshaping hypothesis for action understanding, and show which of the homogeneity-based restrictions may lead to tractability of action understanding. To this end, we build the final part of our bridge by linking the homogeneity-based restrictions to input restrictions of the Bayesian inverse planning model. The effect of these restrictions on the (in)tractability has been investigated by Blokpoel et al. (2013). Table 3 and Figure 1 provide an overview of the five parameters they analyzed.

The above-mentioned parameterizations of Bayesian inverse planning and the previously given interpretations of homogeneity make it possible to operationalize homogeneity. Our contribution is to provide these operationalizations as restrictions on input parameters for Bayesian inverse planning. We go beyond what is currently in the literature by fleshing out more detailed effects that mindshaping might have.

Table 3. Possible parameters for the Bayesian inverse planning model (taken from Blokpoel et al. 2013) and their associated homogeneity hypothesis.

	Homogeneity	Description
|A|	MH-R (partly)	The number of actions that are observed by an interpreter.
|G|	MH-C	The number of goal variables that are inferred by an interpreter.
a	MH-B	The number of available actions values per action variable.
g	unknown	The number of available values per goal variable.
1-p	MH-A	The probability of the most likely goal inference, dependent on the probabilistic knowledge encoded in the Bayesian network.

Restricting the number of observed actions |*A*|

Parameter |*A*| defines the number of actions that an interpreter observes in order to infer the underlying goal.

Mindshaping mechanisms at work in phenomena like ritualization may limit the number of executed and observed actions. However, ritualized behavior may only explain why $|A|$ may be small in those codified situations. Action understanding transcends those cases. If MH-R proponents are committed to small $|A|$ in general, the account needs to be strengthened. Regardless, restricting $|A|$ does not lead to any known tractability result.

Restricting the number of inferred goals $|G|$

If action understanding is to be tractable, then one option is for $|G|$, the number of possible goals that an interpreter actually pursues, to be small (together with g). If, within a social community, an actor would like his/her actions to be timely interpretable to others, then this actor might pursue few goals at a given time so as to make $|G|$ small. This behavior might be the result of mindshaping mechanisms such as habitualization, culture and phenomena like ritualization (MH-C).

Restricting the domain of actions a

Parameter a can be seen as the maximum number of possible actions that are available at any point in time. This number is upper-bounded by the total number of possible actions that are available to a person (MH-B).

Restricting the domain of goals g

Parameter g can be seen as the maximum number of possible goal attributions available for the given inference. This number is upper-bounded by the total number of intentions available to an agent. One might argue that sharing the same intentions (CH-S) may lead to a restriction on g, but it does not as humans may in principle share even an infinitely large set. If anything restricts g, it seems there must be some not yet discovered mindshaping process that does so, or another cognitive process that selects the relevant intentions from the set of all possible intentions. The latter, however, would imply solving the relevance problem (Fodor, 1983, 2001; Pylyshyn, 1989) which is notoriously hard but perhaps the solution lies in a combination of mindshaping and cognitive relevance selection. Due to the ubiquity of g in tractability results discovering these processes would be paramount for having a complete picture of the relation between homogeneity and tractability of mindreading.

Restricting the ambiguity of behavior to make 1-p low

The relational probabilities between variables can be seen as encoding the social knowledge that is brought to bear when inferring the most probable goal. The prior probabilities of variables can be seen as the disposition a person has towards particular unobserved variables (such as goals) at the time of the inference. Together, these probabilistic relations between variables and the prior probability of variables may be shaped by pedagogy, norm following and imitation such that, 1-p is low (MH-A).

Computational complexity of Bayesian inverse planning in 'mindshaped' worlds

Blokpoel et al. (2013) proved several computational complexity results for Bayesian inverse planning.[2] These results show that tractability is not easily achieved. Even restricting multiple parameters simultaneously does not necessarily render the model computationally tractable. The following two intractability results prove that either by themselves or in combination, these restrictions do not make Bayesian inverse planning tractable:

1. Restricting $|A|$, a, and $|G|$ simultaneously, or
2. Restricting $|A|$, a and g simultaneously

Importantly, none of the parameters by themselves render Bayesian inverse planning tractable.[3] Only when the right combination of parameters is restricted, i.e., only when the world is mindshaped in the right way, Bayesian inverse planning does become tractable. The following two results show that if either (3) or (4) or both conditions hold, then action understanding is tractable.

3. Restricting $|G|$ and g, and/or
4. Restricting *1-p* and g

These two tractability results show that in principle, under the correctly (mind)shaped conditions, Bayesian inverse planning can be tractable. However, the results also reveal (at least for the restricted case of action understanding) a gap in the mindshaping theory. While one of the main claims of mindshaping is the importance of homogeneity for the tractability of mindreading, homogeneity alone cannot (yet) fully explain tractability. All known tractability results show that a restriction on g is always necessary for tractability, but no known mindshaping process leads to that restriction.

Discussion

Explaining why people can understand what motivates other people's behavior quickly is, at least from a computational perspective, not trivial. Often, culture or evolution are used to trivialize the paradox of mindreading intractability and explain the speed at which people understand the social world around them. These ideas are embodied by Zawidzki's mindshaping hypothesis. In our analysis, we have shown that, a bridge can be built between mindshaping and a special-case capacity for mindreading, a lot of ground still needs to be covered if these ideas are to fully deal with the intractability paradox.

By detailing possible interpretations of mindshaping effects and relating those to known (in)tractability results for a computational model of action understanding, we have

[2] Blokpoel et al. proved that Bayesian inverse planning can encode and 'solve' computational problems that are amongst some of the hardest problems known in computer science. For details (and a full tutorial) see Blokpoel et al. (2013).

[3] No results are known for *1-p* by itself. It is, however, prudent to assume that restricting *1-p* by itself also does not lead to tractability of Bayesian inverse planning.

shown that only very specific combinations of mindshaping effects have the potential to explain the performance of human mindreading.

The analysis has also revealed that a restriction on the number of available intentions (specifically, the maximum number of possible goal attributions) is a necessary condition for tractability. At the same time, no clear homogeneity effect leads to this restriction. Theoreticians interested in computationally explaining the speed of human mindreading through mindshaping may look for mindshaping mechanisms that specifically lead to this constraint.

Even for a restricted case of mindreading such as action understanding, some of these restrictions have an effect on the tractability of this capacity, while others do not. It stands to reason that caution is in order when claims about tractability are concerned. While not exhaustive, our analysis can be seen as a structured attempt at capturing philosophical and psychological claims about the influence of culture on mindreading into a systematic computational framework.

References

Adams, R., Rule, N., Franklin, R., Wang, E., Stevenson, M., Yoshikawa, S., … Ambady, N. (2010). Cross-cultural reading the mind in the eyes: an fMRI investigation. *Journal of Cognitive Neuroscience*, *22*(1), 97–108.

Alechina, N., & Logan, B. (2010). Belief ascription under bounded resources. *Synthese*, *173*(2), 179–197.

Apperly, I. (2010). *Mindreaders: the cognitive basis of‘ theory of mind’*. Psychology Press.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2008). Bayesian models of human action understanding. *Consciousness and Cognition*, *17*(1), 136–144.

Blokpoel, M., Kwisthout, J., van der Weide, T. P., Wareham, T., & van Rooij, I. (2013). A computational-level explanation of the speed of goal inference. *Journal of Mathematical Psychology*, *57*(3–4), 117–133.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893.

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, *90*(1), 63–86.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287–291.

Cherniak, C. (1986). Limits for knowledge. *Philosophical Studies*, *49*(1), 1–18.

Downey, R. G., & Fellows, M. R. (1999). *Parameterized Complexity*. New York, NY: Springer New York.

Fodor, J. (1983). *The Modularity of Mind*. (Z. W. Pylyshyn, W. Demopoulos, Z. W. E. Pylyshyn, & W. E. Demopoulos, Eds.)*Philosophical Review* (Vol. 94). MIT Press.

Fodor, J. (2001). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. *Representation and mind* (Vol. 10). MIT Press.

Frixione, M. (2001). Tractable competence. *Minds and Machines*, *11*(3), 379–397.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*(1), 20–29.

Goldman, A. I. (2006). *Simulating Minds*. *Philosophical Books* (Vol. 49).

Heal, J. (1996). Simulation, theory, and content. *Theories of Theories of Mind*, 75–89.

Levesque, H. J. (1988). Logic and the complexity of reasoning. *Journal of Philosophical Logic*, *17*(4), 355–389.

Mameli, M. (2001). Mindreading, Mindshaping, and Evolution, 597–628.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman and Company.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, *52*(1), 29–71.

Perez-Zapata, D., Slaughter, V., & Henry, J. D. (2016). Cultural effects on mindreading. *Cognition*, *146*, 410–414.

Pylyshyn, Z. W. (1989). The Robot's Dilemma: The Frame Problem in Artificial Intelligence. Norwood: Ablec Publishing Corporation.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, *22*(1), 1–24.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691.

Tsotos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, *Behavioral*(13), 423–469.

van Rooij, I. (2008). The Tractable Cognition Thesis. *Cognitive Science: A Multidisciplinary Journal*, *32*(6), 939–984.

van Rooij, I., & Wareham, T. (2012). Intractability and approximation of optimization theories of cognition. *Journal of Mathematical Psychology*, *56*(4), 232–247.

Zawidzki, T. (*forthcoming*). Mindshaping. In A. Newen, L. de Bruin, & S. Gallagher (Eds.), *Oxford Handbook of 4E Cognition*. Oxford University Press.

Zawidzki, T. (2008). The function of folk psychology: mind reading or mind shaping? *Philosophical Explorations*, *11*(3).

Zawidzki, T. (2009). Theory of mind, computational tractability, and mind shaping: 2009 Performance Metrics for Intelligent Systems Workshop. In *Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems* (pp. 149–154). ACM.

Zawidzki, T. (2013). *Mindshaping: A New framework for understanding human social cognition*. MIT Press.