frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 30 June 2011
doi: 10.3389/fnhum.2011.00052

Intentional communication: Computationally easy or difficult?

Iris van Rooij[1], Johan Kwisthout[2], Mark Blokpoel[1], Jakub Szymanik[3], Todd Wareham[4] and Ivan Toni[1]*

[1] *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands*
[2] *Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, Netherlands*
[3] *Department of Artificial Intelligence, University of Groningen, Groningen, Netherlands*
[4] *Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada*

Edited by:
Harold Bekkering, University of Nijmegen, Netherlands

Reviewed by:
James Kilner, Institute of Neurology, UK
Patrick Shafto, University of Louisville, USA

****Correspondence:***
Iris van Rooij, Centre for Cognition, Donders Institute for Brain, Cognition, and Behavior, Radboud University Nijmegen, Montessorilaan 3, 6525 HR Nijmegen, Netherlands.
e-mail: i.vanrooij@donders.ru.nl

Human intentional communication is marked by its flexibility and context sensitivity. Hypothesized brain mechanisms can provide convincing and complete explanations of the human capacity for intentional communication only insofar as they can match the computational power required for displaying that capacity. It is thus of importance for cognitive neuroscience to know how computationally complex intentional communication actually is. Though the subject of considerable debate, the computational complexity of communication remains so far unknown. In this paper we defend the position that the computational complexity of communication is not a constant, as some views of communication seem to hold, but rather a function of situational factors. We present a methodology for studying and characterizing the computational complexity of communication under different situational constraints. We illustrate our methodology for a model of the problems solved by receivers and senders during a communicative exchange. This approach opens the way to a principled identification of putative model parameters that control cognitive processes supporting intentional communication.

Keywords: communication, computational complexity, computational modeling, intractability, Bayesian modeling, goal inference

INTRODUCTION

This paper introduces a formal methodology for analyzing the computational complexity of intentional communicative actions, i.e., actions designed to modify the mental state of another agent. The need for such a methodology is evident from the obvious discrepancies between intuitions on the complexity of human communication. Some neuroscientists have argued that intentional communication is easy: we have neural mechanisms that can directly extract communicative intentions from detectable sensory events through the filter provided by our motor abilities (Iacoboni et al., 1999, 2005; Rizzolatti et al., 2001). However, this non-inferential view of intentional communication seems at odds with a core feature of intention attributions, i.e., their context-dependency (**Figure 1** presents a graphical illustration; see also Toni et al., 2008; Uithol et al., 2011). The animal world is bursting with examples of communicative phenomena in which a physical event conveys information by virtue of an interpretation made by a receiver according to the dominant statistical association, without any need for the interpreter to postulate a sender. Highly conventionalized forms of human communication may appear to fall into this category, and thereby within the explanatory power of neuronal mechanisms like those afforded by the mirror neuron system (Jacob and Jeannerod, 2005). Yet, even highly conventionalized signals – such as a thumbs up action – may take on different meanings (e.g., "move up" or "let him live") depending on the communicative context in which they occur. Moreover, even in cases where a thumbs up action does signal the conventional "OK," this apparently unequivocal mapping between sign and signified remains highly ambiguous. For instance, it could signal "OK, let's have dinner," or "OK, it is a deal," or "OK, you can go ahead with what you were planning to do," or an indefinite number of other things which could all be said to be "OK." Accordingly, some theoreticians have argued that, in principle, human intentional communication forms an intrinsically intractable problem, as there is an indefinite number of possible intentions a communicator may entertain at any time and it is logically impossible for a recipient of a communicative signal to determine which intention motivated it (Levinson, 1995, 2006).

Evidently, human communication falls in between these two extremes. Trivial as well as intractable views of human communication fail to adequately characterize the complexity of communication problems solved by humans in everyday situations. After all, humans are often capable of communicating with each other with little or no error, and even when errors occur communicators are often quick to adapt their behaviors so as to resolve any ambiguities. This behavioral success suggests that humans are somehow able to quickly take contextual factors into account and use them to estimate the likely meanings of communicative behaviors in context. Yet, context-sensitive computations are notorious in cognitive computational science for the astronomical demands that they make on computation time (Pylyshyn, 1987; Haselager, 1997; Fodor, 2000; Lueg, 2004). It thus remains a real scientific challenge to explain the success of human communication in combination with its speed. We propose that a more fruitful research agenda is to characterize the conditions that mark boundaries of high and low computational complexity in human communication. This paper presents a conceptual framework and analytic methodology for fleshing out this research agenda.

The framework we propose builds explicit models of communicative problems in order to assess the computational (in)tractability of those problems under different situational constraints.

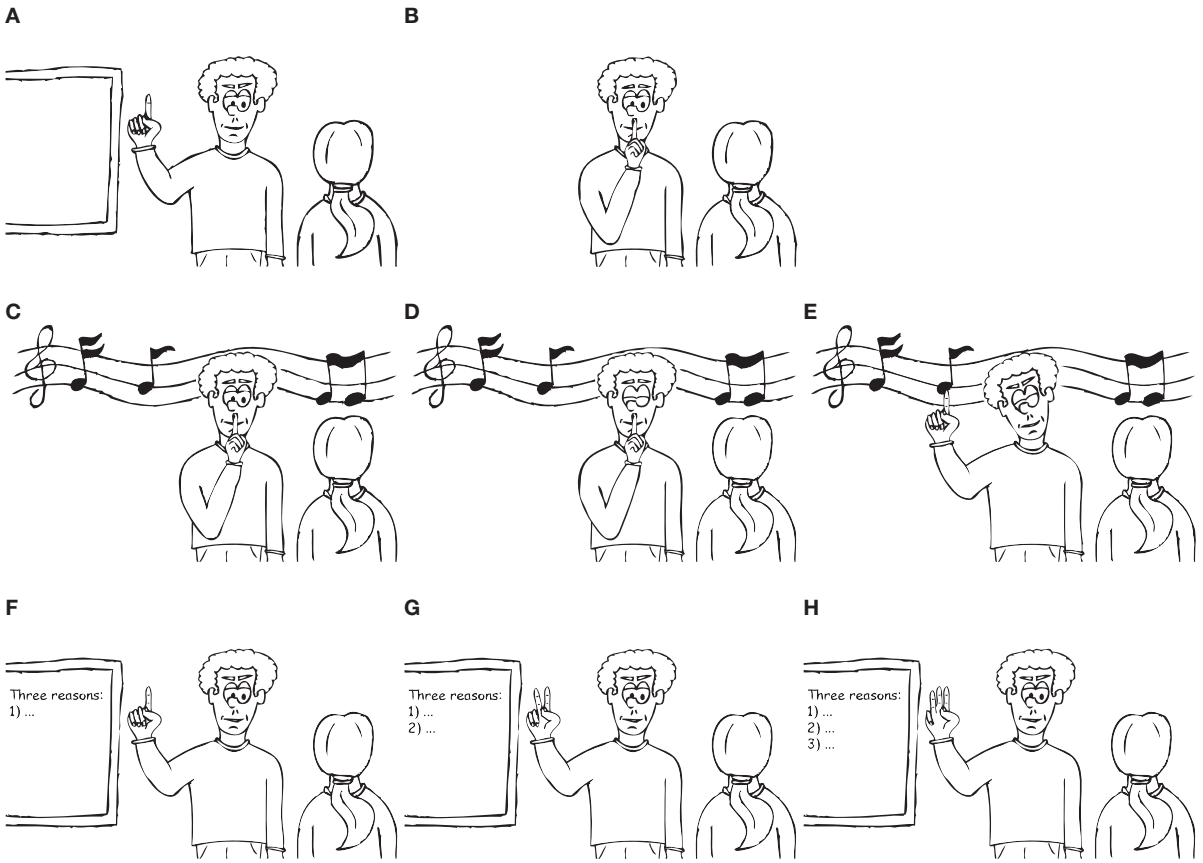

FIGURE 1 | Motor acts (such as raising of the arm and/or extension of the index finger) can suggest different communicative intentions depending on the context in which they are performed. Context information can pertain to information about the concurrent state of the world. For instance, the presence of a school board may make it more likely (compared to other contexts) that a person standing in front of it (likely a teacher) intends to signal "pay attention" to a recipient [see e.g., **(A)**]. Context information can also pertain to information about other concurrently performed motor acts. For instance, a person raising his arm, extending his index finger, *and* moving the index finger to his (closed) lips will more likely (compared to other combinations of actions) intend to signal "be silent" [see e.g., **(B)**]. When this combination of motor acts is additionally combined with a context in which music is played and furthermore additionally acts are made such as closing the eyes, tilting the head and moving the finger away from the lips a *combination* of communicative intentions may be signaled [e.g., "be silent, pay attention, *and* listen" see **(C–E)**]. Importantly, identical actions, performed in different contexts, can signal qualitatively different communicative intentions. For instance, the act and posture of the communicator in **(A)** and **(F)** are identical, but the finger gesture in **(F)** likely does not signal (just) "pay attention" but instead (or also) "let's consider the first reason …", especially if this action is followed by those depicted in **(G)** and **(H)**. (Illustrations by Bas Maes)

Situational constraints that render communication tractable under the given models are then candidate explanations of the speed of everyday human communications. Crucially, these candidate explanations can then be empirically tested, assessing whether the same relation between situational constraints and computational demands predicted by the model fit with the behavior and cerebral processes observed in participants in the lab while solving the same communicative problems. This type of model-driven approach offers several benefits for cognitive neuroscience. First, it identifies putative model parameters that control the cognitive processes supporting intentional communication. Second, it provides a rigorous ground for empirical tests of those models. Third, it offers the possibility to test the neurophysiological plausibility and cognitive relevance of the model by comparing the predicted dynamics of relevant model parameters with the observed dynamics of communicators' internal states (i.e., cerebral signals measured with neuroimaging methods).

OVERVIEW OF THE PAPER

We will illustrate our proposed approach using one particular model of intentional communication as a case study. The model is an extension of the *Bayesian inverse planning* (BIP) model of goal inference proposed by Baker et al. (2007, 2009). In Section "A Probabilistic Model of Intentional Communication," we will describe in detail the original BIP model of goal inference and our adaptations of this model in the form of a Receiver model and Sender model. In Section "Computational Complexity Results," we will study the computational complexity of these Receiver and Sender models, observing that both are computationally intractable unless properly constrained. We then set out to identify constraints that render the models tractable. In Section "Implications and Predictions for Cognitive Neuroscience" we will discuss how the conditions of (in)tractability reported in Section "Computational Complexity Results" can inform critical empirical tests in cognitive neuroscience. In Section "Open Theoretical Questions and Future directions" we

present a set of open questions and suggestions for future theoretical research. We close, in Section "Conclusion," with a reflection on the general relevance of this analytic methodology for cognitive neuroscience of communication and other forms of social interaction.

A PROBABILISTIC MODEL OF INTENTIONAL COMMUNICATION

Our model of communication builds on an existing model of how humans infer *instrumental* goals from observing someone's actions, called the BIP model (Baker et al., 2007, 2009; see also Blokpoel et al., 2010). We extend this model to the domain of communication by making it apply to *communicative* goals as well. We first explain the BIP model (see Preliminaries: The BIP Model of Goal Inference), and then we explain how it can be adapted to the domain of intentional communication (see A BIP Model of Sender and Receiver).

PRELIMINARIES: THE BIP MODEL OF GOAL INFERENCE

According to the BIP model, observers assume that actors are "rational" in the sense that they tend to adopt those actions that best achieve their goals. Here "best" may, for instance, be defined in terms of (expected or believed) efficiency of a set of actions for achieving a given (combination of) goal(s). Say, a person has a single goal of tying his shoe laces. Then this person could make the necessary moves of the fingers, or he could start the finger movements, pause to scratch his chin, and then continue making the finger movements until his shoe laces are tied. If "rationality" is defined in terms of efficiency then the latter sequence of actions would be considered less rational for the goal of "tying one's laces" than the former sequence of actions. Be that as it may, the latter sequence of actions *can* of course be rational for a *different* goal, e.g., if the actor has two simultaneous goals "to tie the shoe laces" *and* "to get rid of that itch on the chin" (and an observer of the latter sequence of action will also likely attribute this combination of goals, over a single goal to the actor).

Given the assumption of rationality, (probabilistic) knowledge of the world and how actions are affected by it, and a measure of relative rationality of action–goal pairs, one can compute the probability that an agent performs an action given its goals, denoted

$$\Pr(\text{action}_1, \text{action}_2, \ldots, \text{action}_k \mid \text{goal}_1, \text{goal}_2, \ldots, \text{goal}_m, \text{context}) \quad (1a)$$

or in shorter format:

$$\Pr(a_1, a_2, \ldots, a_k \mid g_1, g_2, \ldots, g_m, c) \quad (1b)$$

An important insight of researchers such as Baker et al. (2007, 2009) is that observers can use this probabilistic model of planning to make inferences about the most likely goals that actors are pursuing. The reason is that the probability in Eq. 1a/b can be inverted using Bayes' rule to compute the probability of a given combination of goals, given observations of actions that an actor performs:

$$\Pr(g_1, g_2, \ldots, g_m \mid a_1, a_2, \ldots, a_k, c) \propto \Pr(a_1, a_2, \ldots, a_k \mid g_1, g_2, \ldots, g_m, c)\ \Pr(g_1, g_2, \ldots, g_m \mid c) \quad (2)$$

Of all the possible combinations of goals that an observer can (or does) entertain, the goal combination that maximizes the probability in Eq. 2 best *explains* why the observed actions were performed, and according to the BIP model it is this goal combination that an observer will infer[1]. In sum, goal inference under the BIP model equals the computation of the following input–output mapping, informally stated.

Goal Inference (informal)

Input: A representation of the probabilistic dependencies between actions, goals, and states and how these dependencies change over time, and a sequence of observed actions and world states.

Output: A combination of goals that best explains the sequence of actions and world states against the background of the probabilistic dependencies between actions, goals, and world states and how these dependencies change over time.

To be able to analyze the computational complexity of the Goal Inference problem we need to define formal counterparts of all the notions and constructs introduced in the informal problem definition above. Following Baker et al. (2007, 2009), we use BIP-*Bayesian networks* to formally model the input representations in the Goal Inference problem (for a mathematical definition see Preliminaries from Bayesian Modeling in the Appendix). In these networks actions, goals, and world are modeled as value assignments to nodes in the network, and probabilistic dependencies are indicated by directed arcs connecting such nodes (see **Figure 2** for an illustration). For each node in the network there is an associated prior probability of the node taking on a value (e.g., "true" or "false"). In addition, each node has an associated probability distribution, coding the conditional probability of the node taking on a value as a function of the probability of different possible value assignments for the nodes that are connecting to it (called, its "parents"). A set of observed actions and world states is modeled by a value assignment to a sequence of action and state nodes. A combination of goals is modeled as a truth assignment for the goal nodes. A combination of goals is said to "best" explain the observations of actions and world states if it maximizes the probability defined in Eq. 2.

Having formalized all the relevant notions in the Goal Inference problem, we can now state the formal counterpart of the problem (for explanation of notation and mathematical concepts, refer to Preliminaries from Bayesian Modeling in Appendix).

Goal Inference (formalized)

Input: A BIP-Bayesian network $\mathbf{B} = (\mathbf{D}, \mathbf{G})$, where $\mathbf{D}$ is a dynamic Bayesian network with time slices $\mathbf{D}_1 \ldots \mathbf{D}_T$, where each time slice contains an action variable A and a state variable S, and $\mathbf{G} = \mathbf{G}_1 \ldots \mathbf{G}_k$ denotes a set of instrumental goals. Further, a set of observed actions and states $\mathbf{a} \cup \mathbf{s}$.

Output: The most probable joint value assignment $\mathbf{g}$ to the goals in $\mathbf{G}$, i.e., $\text{argmax}_{\mathbf{g}} \Pr(\mathbf{G} = \mathbf{g} \mid \mathbf{A} = \mathbf{a}, \mathbf{S} = \mathbf{s})$, or $\varnothing$, if $\Pr(\mathbf{G} = \mathbf{g} \mid \mathbf{A} = \mathbf{a}, \mathbf{S} = \mathbf{s}) = 0$ for all joint value assignments $\mathbf{g}$ (here the output $\varnothing$ can be read as meaning "no plausible goal attribution can be made").

[1] In other words, in the BIP model, goal inference is conceptualized as a form of probabilistic inference to the best explanation, a.k.a. *abduction* (Charniak and Shimony, 1990). As this model is situated at Marr's (1982) computational level, the theory is in principle consistent with a variety of hypotheses about the architecture that implements the postulated computations, ranging for instance from predictive coding or forward models (Miall and Wolpert, 1996; Oztop et al., 2005; Kilner et al., 2007a,b) to neural network models (Paine and Tani, 2004) or random sampling models (Vul et al., 2009).

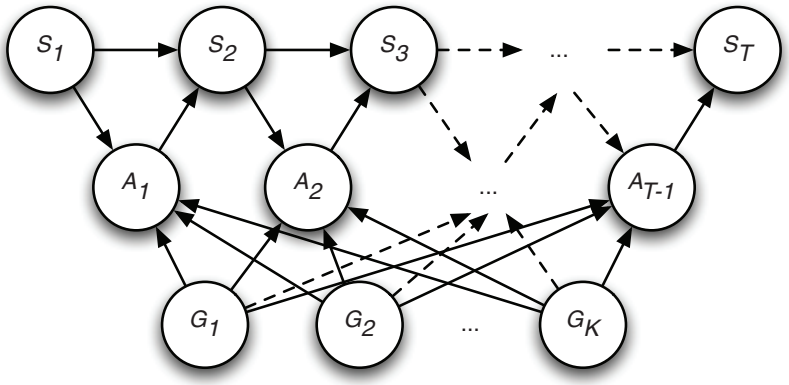

FIGURE 2 | The BIP-Bayesian network with *T* time slices: S_t with $t = 1, 2, \ldots T$, is a state variable at time *t*; A_t with $t = 1, 2, \ldots T$, is an action variable at time *t*. The goal variables, G_1, G_2, … G_K are fixed over the different time slices. Arrows indicate probabilistic dependencies between variables in the network. Note that no direct dependencies exist between states and goals, but that any indirect dependencies between states and goals are mediated by action variables.

The computational complexity of this model has previously been analyzed by Blokpoel et al. (2010). Specifically, these authors presented a mathematical proof that the problem is computationally intractable (NP-hard) if no constraints are imposed on the input representations. They furthermore found that the problem is tractably computable under the constraint that the set of candidate goals (**G**) is small and/or the probability of the most probable goal assignment (p) is large. Here, we will use the GOAL INFERENCE problem as an inspiration for similarly explicit characterizations of the SENDER and RECEIVER problems solved by human communicators. As we will see, some (though not all) of the computational complexity results for GOAL INFERENCE apply also to the new SENDER and RECEIVER problems.

A BIP MODEL OF SENDER AND RECEIVER

If indeed – as the BIP-model postulates – observers of actions infer goals from actions by means of an inference to the best explanation, then senders can use this knowledge to predict how a receiver will interpret their actions ahead of time. For instance, they could engage an internal simulation of a receiver's inferences to the best explanation when considering the suitability of candidate actions (Noordzij et al., 2009). Such a simulation subroutine could be called multiple times during the planning of instrumental and communicative behaviors, allowing a sender to converge on an action sequence which is both efficient and is likely to lead a receiver (if properly simulated) to attribute the correct (i.e., intended) communicative goal to the sender. This conceptualization of the computational bases of sender signal creation can be summarized by the following informal input–output mapping[2].

[2]We remark that the computational-level SENDER and RECEIVER models that we propose are in one respect a simplification and in another an enrichment of the Shafto and Goodman (2008) model (see also Frank et al., 2009). Our models are a simplification in the sense that they do not assume infinite recursive reasoning by sender and receiver about each other, in contrast to the Shafto and Goodman model. Our model is an enrichment in the sense that we adopt the Bayesian network structure of the Baker et al. (2007, 2009) BIP model to have a more general characterization of the assumed dependencies between states, goals, and actions.

Sender (informal)

Input: A representation of the probabilistic dependencies between actions, goals, and states and how these dependencies change over time, and one or more communicative and instrumental goals.

Output: A sequence of actions that will lead to the achievement of the instrumental goals and will lead a receiver to attribute the correct communicative goals to the sender.

Note that our model allows for the possibility that senders can have simultaneous instrumental and communicative goals as seems required for ecological validity. After all, in everyday settings communicative behaviors are typically interlaced with several sequences of instrumental actions – think, for instance, of a car driver signaling to another driver at night that his lights are off, while at the same time trying to drive safely and stay on route.

Building on the formalisms used in the GOAL INFERENCE model, the informal SENDER model yields the following formal input–output mapping (see **Figure 3** for an illustrative example of a BIP-network for communicative goals).

Sender (formalized)

Input: A BIP-Bayesian network $\mathbf{B} = (\mathbf{D}, \mathbf{G})$, where $\mathbf{D}$ is a dynamic Bayesian network with time slices $\mathbf{D}_1 \ldots \mathbf{D}_T$, such that each time slice contains an action variable A and a state variable S, $\mathbf{G}_I = \mathbf{G}_1 \ldots \mathbf{G}_k$ denotes a set of instrumental goals, and $\mathbf{G}_C = \mathbf{G}_{k+1} \ldots \mathbf{G}_n$ denotes a set of communicative goals, and a set of goal assignments $\mathbf{g}_I \cup \mathbf{g}_C$ to $\mathbf{G}_I \cup \mathbf{G}_C$.

Output: A joint value assignment $\mathbf{a}$ to the action variables such that $\mathbf{a} = \text{argmax}_{\mathbf{a}} \Pr(\mathbf{A} = \mathbf{a} | \mathbf{G}_I = \mathbf{g}_I)$ and $\text{RECEIVER}(\mathbf{B}, \mathbf{a}, \mathbf{s}) = \mathbf{g}_C$, or $\varnothing$ if no such joint value assignment $\mathbf{a}$ is possible. Here $\mathbf{s}$ is $\text{argmax}_{\mathbf{s}} \Pr(\mathbf{S} = \mathbf{s} \mid \mathbf{A} = \mathbf{a})$, i.e., the most likely state $\mathbf{s}$ to follow from the actions.

Here the receiver function RECEIVER(.,.,.) is modeled after the GOAL INFERENCE function studied by Baker et al. (2007, 2009) and Blokpoel et al. (2010), but with an important difference: the receiver is presented with a sequence of actions that was purposely designed by

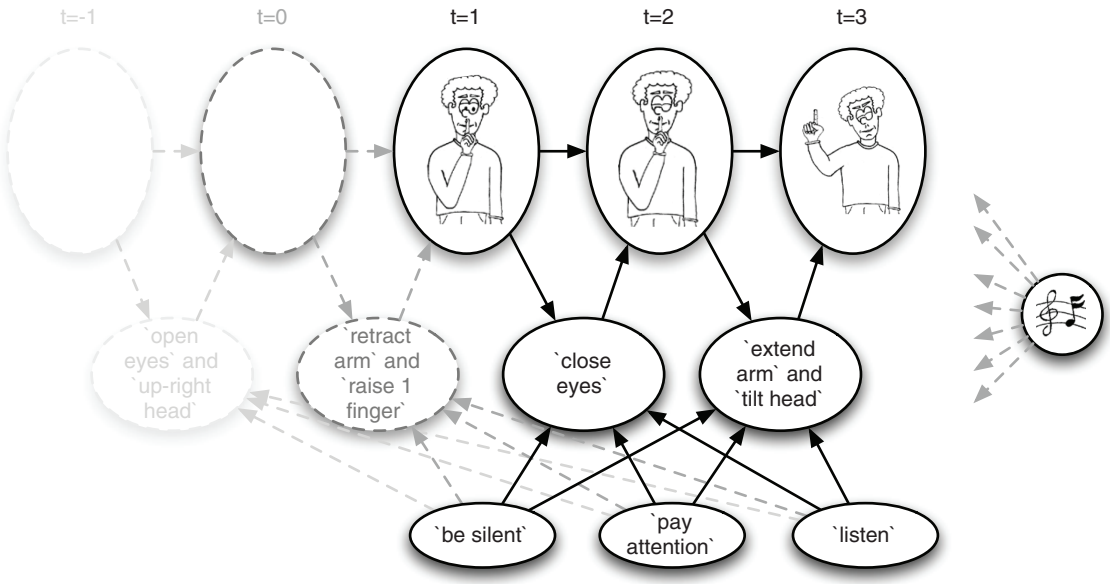

FIGURE 3 | An illustration of how probabilistic knowledge about the scenarios depicted in Figures 1C–E can be encoded in a BIP-Bayesian network as probabilistic dependencies between goals, actions, and states.

a sender with the aim of making her communicative goals interpretable by a receiver, while concurrently trying to achieve one or more instrumental goals unknown to the receiver. This difference between the GOAL INFERENCE problem and the RECEIVER problem can have non-trivial consequences for the computational complexity of the RECEIVER problem, which is why we need to analyze it anew in this paper.

For completeness and clarity, we state the RECEIVER problem in both informal and formal format below.

Receiver (informal)

Input: A representation of the probabilistic dependencies between actions, goals, and states and how these dependencies change over time, and a sequence of observed actions and world states.

Output: A combination of communicative goals that best explains the sequence of actions and world states against the background of the probabilistic dependencies between actions, goals, and world states and how these dependencies change over time.

Receiver (formalized)

Input: A BIP-Bayesian network $\mathbf{B} = (\mathbf{D}, \mathbf{G})$ as in the SENDER problem; a set of observed actions and states $\mathbf{a} \cup \mathbf{s}$.

Output: The most probable joint value assignment $\mathbf{g}_C$ to the communicative goals in $\mathbf{G}_C$, i.e., $\text{argmax}_g \Pr(\mathbf{G}_C = \mathbf{g}_C \mid \mathbf{A} = \mathbf{a}, \mathbf{S} = \mathbf{s})$, or $\varnothing$, if $\Pr(\mathbf{G}_C = \mathbf{g}_C \mid \mathbf{A} = \mathbf{a}, \mathbf{S} = \mathbf{s}) = 0$ for all joint value assignments $\mathbf{g}_C$.

Having defined both the generic SENDER and RECEIVER problems, we are now in a position to analyze their computational complexity. We report on our analyses in the next section.

COMPUTATIONAL COMPLEXITY RESULTS

In this section we state all computational complexity results. Readers interested in full details on the mathematical proofs are referred to the Section "Computational Models and Computational Complexity Analyses" in the Appendix. This section assumes familiarity with concepts and methods from the field of computational complexity analysis (for a primer see Preliminaries from Complexity Theory in the Appendix). For instance, contrary to the suggestion that computationally intractable functions can be approximately computed efficiently (e.g., Love, 2000; Chater et al., 2003, 2006), computational complexity theory (cf. Kwisthout et al., 2011) has clearly shown that many intractable (e.g., NP-hard) functions cannot be efficiently approximated (Arora, 1998; Ausiello et al., 1999), and almost all are intractable to approximate if only a constant sized error is allowed (Garey and Johnson, 1979; van Rooij et al., 2010)[3]. Accordingly – although the analyses we provide work under the assumption that the models exactly predict the outputs of human computations – we conjecture that the obtained results also apply under the assumption that the models approximately predict the outputs of human computations. Here "approximate" means that the approximate prediction does not differ from the exact prediction by more than some small constant factor. For ideas on how approximation can be explicitly taken into account in fixed-parameter tractability analyses we refer the reader to Hamilton et al. (2007) and Marx (2008).

[3]We are aware of computer simulation studies in the cognitive science literature that seem to suggest that Bayesian models can be efficiently approximated (Vul et al., 2009; Sanborn et al., 2010). However, the intractability results for approximating Bayesian computations available in the computer science literature show that these "simulation-based approximability findings" cannot generalize to all types of inputs, but only to some that have special properties (cf. "parameters") that can be exploited for tractable (exact or approximate) computation (Kwisthout et al., in 2011).

INTRACTABILITY OF GENERIC COMMUNICATION

Our two main computational intractability results are:

Result 1. The RECEIVER problem is NP-hard.
Result 2. The SENDER problem is NP-hard.

These results establish that both RECEIVER and SENDER problems are computationally as hard to compute as a whole class of problems that are strongly conjectured to be intractable, called the NP-complete problems, and possibly even harder. This means that SENDER and RECEIVER problems cannot be computed efficiently (more precisely, in polynomial-time) unless all NP-complete problems can be (Garey and Johnson, 1979; Aaronson, 2005; Fortnow, 2009). As there is both theoretical and empirical evidence that no NP-complete problem allows for an efficient solution procedure we conclude that RECEIVER and SENDER problems, as we have defined them, are intractable. We note that our Results 1 and 2 are, to the best of our knowledge, the first formal proofs that are consistent with intuitive claims of intractability in the communication literature (Levinson, 1995, 2006; Pickering and Garrod, 2004; Barr and Keysar, 2005).

IDENTIFYING CONDITIONS FOR TRACTABILITY

Importantly, our analyses do not stop at the intractability results (Results 1 and 2). On the contrary, we view such results as merely the fruitful starting point of rigorous analyses of the sources of complexity in human communication. For these further analyses we adopt a method for identifying sources of intractability in cognitive models developed by van Rooij and Wareham (2008; see also van Rooij, 2008; van Rooij et al., 2008). The method builds on concepts and techniques from the mathematical theory of parameterized complexity (Downey and Fellows, 1999), and works as follows.

First, one identifies a set of potentially relevant problem parameters $K = \{k_1, k_2, \ldots, k_m\}$ of the problem P under study (for us, the RECEIVER and SENDER problems). Then one tests if it is possible to solve P in a time that can grow excessively fast (more precisely: exponential or worse) as a function of the elements in the set $K = \{k_1, k_2, \ldots, k_m\}$ yet slowly (polynomial) in the size of the input[4]. If this is the case, then P is said to be *fixed-parameter (fp-) tractable* for parameter set K, and otherwise it is said to be *fp-intractable* for K. Observe that if a parameter set K is found for which P is fp-tractable then the problem P can be solved quite efficiently, even for large inputs, provided only that the members of K are relatively small. In this sense the "unbounded" nature of K can be seen as a reason for the intractability of the unconstrained version of P. Therefore, we also call K a *source of intractability* of P.

The RECEIVER and SENDER models have several evident parameters, each of which may be a source of the intractability inherent in the general problems postulated by these models. **Table 1** gives an overview of the parameters that we consider here. Using the abovementioned methodology for fp-tractability analysis we have been able to derive a set of fp-(in)tractability results, which are summarized in **Table 2**. We will discuss these results and what they imply for the tractability of the RECEIVER and SENDER under different situational constraints.

[4]More formally, this would be a time on the order of $f(k_1, k_2, \ldots, k_m)n^c$, where f is an arbitrary computable function, n is a measure of the overall input size, and c is a constant.

We start by considering the fp-(in)tractability of the RECEIVER problem. We have two main fp-tractability results:

Result 3. The RECEIVER problem is fp-tractable for parameter set $\{|\mathbf{G_I}|, |\mathbf{G_C}|\}$.
Result 4. The RECEIVER problem is fp-tractable for parameter set $\{|\mathbf{G_I}|, 1 - p\}$.

Note that, by definition, if a problem is fp-tractable for a parameter set K then it is also fp-tractable for any superset $K' \supseteq K$. Hence, Results 3 and 4 imply all other fp-tractability results listed in **Table 2** for the RECEIVER problem, including the fp-tractability of the problem for the combined parameter set $\{|\mathbf{G_I}|, |\mathbf{G_C}|, 1 - p\}$.

Result 3 establishes that it is possible to compute the input–output mapping defined by the RECEIVER problem in a time which grows fast (read: exponential) only in the two parameters $|\mathbf{G_I}|$ and $|\mathbf{G_C}|$, and slow (read: polynomial) in the remainder of the input, regardless the size of other input parameters. Result 4 establishes that it is

Table 1 | Overview of parameters considered in our tractability analyses.

Parameter	Description
$\|\mathbf{G_C}\|$	The size of the set of communicative goals
$\|\mathbf{G_I}\|$	The size of the set of instrumental goals
$\|\mathbf{A}\|$	The size of the set of action nodes
$\|\mathbf{S}\|$	The size of the set of states
T	The number of time slices
$1 - p$	Here p is the probability of the most probable communicative goal attribution

Table 2 | Overview of complexity results and open questions for RECEIVER problem (top) and SENDER problem (bottom).

	–	$\|\mathbf{G_I}\|$	$\|\mathbf{G_C}\|$	$\|\mathbf{G_I}\|, \|\mathbf{G_C}\|$
RECEIVER				
–	NP-hard	fp-intractable	fp-intractable	fp-tractable
$\|\mathbf{A}\|$	fp-intractable	fp-intractable	fp-intractable	fp-tractable
$1 - p$	fp-intractable	fp-tractable[a]	fp-intractable	fp-tractable
$\|\mathbf{A}\|, 1 - p$	fp-intractable	fp-tractable	fp-intractable[a]	fp-tractable
SENDER				
–	NP-hard	fp-intractable	fp-intractable	fp-intractable
$\|\mathbf{A}\|$	fp-intractable	fp-intractable	fp-intractable	fp-tractable
$1 - p$	fp-intractable	fp-intractable	fp-intractable	fp-intractable[a]
$\|\mathbf{A}\|, 1 - p$	fp-intractable	?	fp-intractable[a]	fp-tractable

The table lists all possible combinations of parameters from **Table 1:** *each cell stands for the parameter set consisting of its row and column labels. In the special case of the empty parameter set classical (non-parameterized) complexity applies, in this case NP-hardness. Note that in our model $|\mathbf{A}| = |\mathbf{S}| = T$. Hence for all listed results including parameter $|\mathbf{A}|$ in the parameter set the other two parameters $|\mathbf{S}|$ and T may be replaced (or even added) with no change to the listed result. Full details on the proofs of theorems and explanations of corollaries and conjectures can be found in Section "Computational Models and Computational Complexity Analyses" in the Appendix.*

[a]*Technically, as a consequence of probabilities being constrained to a value between 0 and 1, these results are not fp-(in)tractability results in the formal sense of the term. Yet, we do have formal proofs that show 1 – p is (or respectively, is not) a source of complexity in the same sense that the formal notion of fp-(in)tractability intends to capture.*

also possible to compute the input–output mapping defined by the RECEIVER problem in a time which grows fast only in the two parameters $1-p$ and $|\mathbf{G}_I|$, and slow in the remainder of the input, regardless the size of other input parameters (including $|\mathbf{G}_C|$). Informally, this means that the inference task modeled by the RECEIVER problem can be performed fast, even for large networks of beliefs, if the number of possible instrumental goals that the receiver believes the sender may have ($|\mathbf{G}_I|$) is relatively small *and at the same time* either the number of communicative goals is relatively small or the probability of the most likely communicative goal attribution (p) is relatively large.

These mathematical results lead to a clear prediction. A receiver is able to quickly attribute communicative intentions to a sender's actions if the number of instrumental goals that the receiver assumes that the sender is pursuing in parallel to her communicative goals is small *and* the sender does not have many communicative goals she wishes to convey simultaneously, or otherwise she was able to construct a sequence of actions that leads to a signal of low ambiguity (captured by small $1-p$).

Importantly and perhaps counter intuitively, low ambiguity of the signal is by itself not sufficient for tractability, as we have the following intractability result.

Result 5. The RECEIVER problem is fp-intractable for parameter set $\{1-p\}$.

Similarly, a small number of communicative goals is by itself not sufficient for tractability, as we also have the following fp-intractability result.

Result 6. The RECEIVER problem is fp-intractable for parameter set $\{|\mathbf{G}_C|\}$.

However, if senders were to focus solely on communicating (and forgetting for the moment about any other instrumental goals they may also want to achieve), then $|\mathbf{G}_I|=0$, and hence low ambiguity (or small number of communicative goals) would by itself suffice to make the inference task easy for the receiver. The reason that low ambiguity (or small number of communicative goals) is not enough for tractability for $|\mathbf{G}_I|>0$ is that, in order to compute the probability of the most probable communicative goal assignment, the receiver needs to do this computation against the background of all possible instrumental goals a sender may have, as having one or more of those can affect the probability of the target communicative goals. The number of possibilities that the receiver needs to be taken into account grows, in this case, exponential in $|\mathbf{G}_I|$, and therefore the computation is too resource demanding to be done efficiently for large $|\mathbf{G}_I|$.

Having seen that neither large p nor small $|\mathbf{G}_C|$ are by itself sufficient for tractability of the RECEIVER problem, but that combining either with small $|\mathbf{G}_I|$ does yield tractability, one may ask if there are other *combinations* of constraints that are sufficient for tractability of the RECEIVER problem. As can be seen from the overview of fp-(in)tractability results in **Table 2**, no combination of parameters excluding $\{|\mathbf{G}_I|, |\mathbf{G}_C|\}$ or $\{|\mathbf{G}_I|, 1-p\}$ as subsets yields fp-tractability for the RECEIVER problem. This means that these two sets seem to fully characterize what makes the RECEIVER problem difficult or easy (at least, relative to the total set of parameters that we consider here and listed in **Table 1**).

As the parameter $|\mathbf{G}_I|$, $|\mathbf{G}_C|$ and $1-p$ all seem to be, to some extent, under the control of the sender, Results 3 and 4 raise the question of how computationally complex it is for senders to ensure these parameters (or pairs of them) are small, and thereby make the intention attribution task tractable for receivers. We will discuss a set of results that are relevant to this question, starting with the following:

Result 7. The SENDER problem is fp-intractable for parameter set $\{|\mathbf{G}_I|, |\mathbf{G}_C|\}$.

In other words, the RECEIVER problem being tractable is by itself not sufficient for the SENDER problem to be tractable (compare Result 3 and Result 7). Additional constraints need to be assumed to explain the tractability of the sender's task. Result 8 shows that a low upper-bound on the length of the sequence of actions that needs to be planned can help make the sender's task easier.

Result 8. The SENDER problem is fp-tractable for parameter set $\{|\mathbf{A}|, |\mathbf{G}_I|, |\mathbf{G}_C|\}$.

Result 8 can be understood as a consequence of the ability of a sender to use the same fp-tractable algorithm that the receiver can use to tractably infer the sender's communicative goals to predict the receiver's interpretation of a given action sequence, and then search the space of action sequences (which is exponential only in $|\mathbf{A}|$) for a sequence that yields the right receiver inference. This algorithmic strategy does not yield fp-tractability for the Sender problem for the parameter set $\{|\mathbf{A}|, |\mathbf{G}_I|, 1-p\}$, because contrary to $|\mathbf{G}_C|$, which is constant for all action sequences the sender may consider, the value of $1-p$ is different for different candidate action sequences. As a result, $1-p$ cannot generally be assumed to be small for *all* possible action sequences. Whether or not a different strategy exists that can exploit this combination of parameters for solving the SENDER problem in fixed-parameter tractable time is currently unknown:

Open question. Is the SENDER problem is fp-tractable for parameter set $\{|\mathbf{A}|, |\mathbf{G}_I|, 1-p\}$?

What we do know is that no other combination of parameters excluding $\{|\mathbf{A}|, |\mathbf{G}_I|, |\mathbf{G}_C|\}$ as a subset suffices to render the SENDER problem fp-tractable. If future research were to reveal that the SENDER problem is fp-tractable for parameter set $\{|\mathbf{A}|, |\mathbf{G}_I|, 1-p\}$, then this would indicate that given that the RECEIVER problem is tractable (e.g., Result 3 and 4) parameter a bound on $|\mathbf{A}|$ suffices for tractability of the SENDER problem. If, on the other hand, it would turn out that the SENDER problem is fp-intractable for parameter set $\{|\mathbf{A}|, |\mathbf{G}_I|, 1-p\}$, then this would underscore the high complexity of the SENDER problem, as it is impossible for the input to the SENDER problem to be large when all three parameters $|\mathbf{A}|$, $|\mathbf{G}_I|$, and $|\mathbf{G}_C|$ are small at the same time.

DISCUSSION

We analyzed the complexity of two models of senders and receivers in a communicative exchange. In Section "Implications and Predictions for Cognitive Neuroscience" we will discuss the implications of our findings for cognitive neuroscience, including a set of empirical predictions testable by cognitive neuroscience methods. All predictions made in this section must, of course, be understood relative to the models that we studied, which inevitably make some basic assumptions about properties of the mental representations of receivers and senders in the communication task. In Section "Open Theoretical Questions and Future directions" we will explicitly discuss these assumptions and give pointers for possible extensions or

adaptations of the models. As such new model variants need not inherit all the (in)tractability results from our SENDER and RECEIVER models they naturally yield a set of open questions for future theoretical research.

IMPLICATIONS AND PREDICTIONS FOR COGNITIVE NEUROSCIENCE

We presented two new models of the tasks engaging senders and receivers during a communicative exchange: see the RECEIVER and SENDER models in Section "A BIP model of Sender and Receiver." As these models are situated at Marr's (1982) computational level, they are not bound to particular assumptions about a specific algorithm that human brains would use to perform the postulated computations, nor about how these algorithms are exactly implemented in neural mechanisms. They merely state the hypothesized input–output mappings assumed to underlie the receiver and sender tasks in communication respectively. This high level of abstraction has the benefit of generalizability, i.e., the sources of intractability identified in the models are inherent in the problems that they describe, and not specific to particular algorithms for solving those problems or implementations thereof. As a consequence, our intractability results are proofs of *impossibility*: *no* algorithmic implementation of any type, in any neural mechanism, can ever compute the problems quickly if the input domain is unconstrained.

It could be argued that, since human communication is evidently tractable for real humans in the real world, the present results are useless. However, the mismatch between model and reality is informative for improving formal models of human communication, and for rejecting claims that intentional communication is computationally trivial to explain (Rizzolatti and Craighero, 2004; Iacoboni et al., 2005). The results can help to avoid trying to account for situations that are in fact impossibilities as far as efficient human communication is concerned. In other words, the generic RECEIVER and SENDER models considered here are too powerful, i.e., they describe situations that preclude tractable computation. Many models in cognitive neuroscience which are powerful enough to cover a wide variety of domains, such as models of reinforcement learning (Gershman and Niv, 2010) or decision-making (Dayan, 2008), appear to have this property. This can be understood by the inherent computational complexity introduced by the (unconstrained) domain generality of computations, an issue that has long been known in artificial intelligence and cognitive science (Fodor, 1983, 2000; Pylyshyn, 1987; Haselager, 1997).

A major advantage of the current approach lies in its ability to indentify, from first principles, which situational constraints render the generic RECEIVER and SENDER models tractable. Our fp-tractability results show that there exist algorithms for efficiently computing the problems if the situations that arise are constrained in specific ways. For instance, we found that if receiver assumes that senders do not pursue many instrumental goals concurrently to their communicative goals, and if they ensure that the most probable communicative goal assignment has high probability, then the receiver task is tractable. Similarly, we found that if senders do not pursue many instrumental and communicative goals, and use short action sequences for communication, then the sender task is tractable. Conditional on the validity of our assumptions, these findings are novel and important insofar that they describe those situations in which human communication could proceed efficiently. These results are testable in the context of novel experimentally controlled empirical approaches to communication, in particular those rapidly evolving in the new exciting domain of experimental semiotics (Galantucci and Garrod, 2011). This domain is focused on studying interactions that occur in the absence of pre-established communicative conventions, thereby allowing researchers an unprecedented level of experimental control over the communicative means and goals of the communicators. Accordingly, communicative interactions are studied in the context of non-verbal tasks. This approach allows one to disambiguate the ability to solve genuine communicative problems *de novo*, from the implementation of empirical generalizations from past experience. This approach also excludes that communicators simply exploit a pre-existing communicative system powerful enough to mutually negotiate new communicative behaviors, an obvious non-sequitur for understanding how human communicative abilities come into place. The communicative games designed according to these principles appear amenable to manipulation of the crucial parameters that have emerged from the present analysis of computational complexity, namely the number of instrumental goals that a sender needs to accomplish during a communicative exchange, or the number of actions afforded by a sender. Furthermore, bringing these present results to the level of empirical research will be beneficial for highlighting some of the simplifications inherent in developing a computational-level account of human communicative abilities. For instance, translating the present results into empirical studies will require non-trivial operationalization of abstractions like "actions" and "goals."

The fp-tractability results are *possibility* proofs in the sense that there exist fp-algorithms that can perform the sender and receiver tasks quickly if certain situational parameters are constrained to take small values. Specifically, we proved the existence of an fp-tractable algorithm for the SENDER problem which exploits small values for parameters $|\mathbf{A}|$ (the number of to be planned sender actions), $|\mathbf{G_I}|$ (the number of sender instrumental goals), and $|\mathbf{G_C}|$ (the number of sender communicative goals). Furthermore, we proved the existence of an fp-tractable algorithm for the RECEIVER problem that can exploit the same set of parameters, or even a smaller set which does not include $|\mathbf{A}|$. In addition, we proved the existence of an fp-tractable algorithm for the RECEIVER problem that can exploit the parameters $|\mathbf{G_I}|$ and $1 - p$ (where p is the probability of the most likely communicative goal attribution). Whether or not human brains actually use and implement such fp-tractable algorithms is an empirical question. Note that posing this question is only made possible by the type of complexity analysis described in this paper. We believe this is a relevant question, as it stems from a principled analysis rather than occasional observations. The cognitive neuroscience literature is replete with examples of cerebral responses being attributed to "task difficulty" on the basis of introspections. Here, we show how task difficulty could be formally parameterized. This approach opens the way for using cognitive neuroscience methods to test whether and how human brains exploit similar fp-tractable computations to efficiently solve communicative problems. The prediction to be tested is that resource demands during human communication should be extremely sensitive to the parameters identified as critical for tractability of the model. Recall that an fp-tractable algorithm for parameters $\{k_1, k_2, \ldots, k_m\}$ runs fast if and only if all the parameters k_1, k_2, …, and k_m

are constrained to sufficiently small values in the current input (i.e., situation). If, in the lab, one were to create communication tasks in which the critical parameters can be systematically manipulated, then two critical predictions follow from the hypothesis that the senders use, say, the fp-tractable algorithm underlying Result 8, and receivers use the fp-tractable algorithm underlying Result 3 or 4:

(1) As long as $|\mathbf{A}|$, $|\mathbf{G_I}|$, $|\mathbf{G_C}|$, and $1 - p$ remain small the task of communication should proceed efficiently.
(2) As soon as two parameters in the set $\{|\mathbf{G_I}|, |\mathbf{G_C}|, 1 - p\}$ take on large values, the task of communication would start to consume an excessive amount of computational resources (for sender and/or receiver, depending in which the relevant parameters are large), which may be reflected in a decrease of speed, increase of communicative errors, and/or increased brain activities in certain brain areas underlying the relevant computations.

If these predictions were to be confirmed, then this result would provide evidence for both cerebral use of fp-tractable computations, and for the RECEIVER and SENDER models as defined above. A natural next question would then be how these algorithms are neurally implemented. If, on the other hand, the predictions were to be disconfirmed, then two interpretations are open. Either, the RECEIVER and SENDER models are not valid, and this would motivate new modeling directions; or RECEIVER and SENDER models are valid, but humans use different fp-tractable algorithms for computing them, exploiting parameters other than the ones that we have studied here. This latter option is not excluded, as there may be parameters of the problems outside the set described in **Table 1**. Constraining these additional parameters may render computation of these problems tractable as well. Both research directions appear highly informative.

We end this section by remarking that complexity results, such as we have derived here, also expand the ways in which computational-level models can be tested for input–output equivalence with human communicators. This form of testing involves presenting a participant with a situation that can be described as a particular input in the model's input domain, computing the output that according to the model corresponds to that input, and testing if the responses of the participant match the predicted outputs. This test is quite common in cognitive science, and was also the method used by Baker et al. (2007, 2009) to test their computational-level model of (instrumental) goal inference. However, this methodology is often used within particular "toy-scenarios," where the tested situations need to be exceedingly simple, such that inputs can be explicitly modeled and output computation is tractable for the researcher (see e.g., Oztop et al., 2005; Cuijpers et al., 2006; Erlhagen et al., 2006; Baker et al., 2007; Yoshida et al., 2008). Without a clear estimate of the complexity of the problem at hand, the results found within the boundaries of those toy-scenarios might easily fail to scale up to more complex and realistic domains. For instance, those computational models of goal inference that seem to work (i.e., that make plausible goal inferences without running into tractability issues) severely restrict the possible contexts and the number of possible actions and goals, keeping their application domain far removed from realistically complex situations. In contrast, the present results enable communication researchers to move away from trivial "toy domains," using the fp-tractable algorithms identified here to expand the domains in which to compute predicted input–output mappings (for details on the algorithms see the Appendix). In cases where it is difficult or impossible to get an explicit model of the entire communicative setting, our proposed methodology of testing sources of tractability in the computational-level models of interest can be used instead.

OPEN THEORETICAL QUESTIONS AND FUTURE DIRECTIONS

For our case study, we opted for analyzing the RECEIVER and SENDER models as defined in Section "A BIP model of Sender and Receiver." By doing so, we made several commitments that may have affected our complexity results. We will briefly highlight those commitments known to affect the results, and those which we believe are unlikely to have had an effect. We will distinguish between model-specific commitments (such as the nature of the connectivity of the input networks and assumptions about the amount and type of information available to senders and receiver) and our choice of formalism (i.e., a probabilistic formalism).

The models that we defined make the strong commitment that there is no direct (probabilistic) dependency between states and goals, as all such dependencies are assumed to be mediated by actions (see **Figures 2 and 3**). This assumption seems problematic if one considers that states of the world may make certain goals of an actor more likely (e.g., if a cup is present and within reach, then a grasping action toward that cup is more likely to have as goal "grasp the cup," as compared to a situation where the cup is not present, or out of reach). Although this commitment strictly speaking reduces the validity of the model, we know it has no effect on our complexity results (they would apply equally well to a model that allows for direct state-goal dependencies). The reason is that the model makes an additional assumption, namely the assumption that all relevant states and actions are observable by the receiver. This brings us to the point of validity of this assumption.

Although complete observability may hold true in some situations (and in those situations the models may thus apply without problem), it is likely that in many real-world situations not all relevant actions and states are observable. Think, for instance, of the possibility that part of a visual scene is occluded by nearby objects or even an eye blink. Then some actions and state changes may be observed directly, while others need to be inferred by an observer. Humans are often able to do this, and models of receivers should be able to explain how they can do so. Because our fp-tractability results reported in Section "Computational Complexity Results" heavily depend on the assumption of complete observability, we must admit that our analyses have not yet shed light on how communication can proceed efficiently under conditions of partial observability. Future research on the computational complexity of model variants that include the assumption of partial observability are thus called for.

Another important property of our models is that they assume that the sender and receiver know everything that is relevant to the context of the communicative exchange, including which action and goal variables are relevant, and their probabilistic interdependencies. Clearly, this assumption sweeps a considerable amount of computational complexity under the rug; after all, computing the

set of relevant information itself may well be computationally very resource demanding. Be that as it may, our analyses apply in those cases where humans do have complete knowledge of everything relevant to the communication task, and this is a condition that can certainly be met in laboratory settings used to investigate sources of complexity in communication as suggested in Section "Implications and Predictions for Cognitive Neuroscience." Also, our analyses show that even if there are not an indefinite number of intentions that senders and receivers may entertain (see our Introduction), communication under simpler models can still be intractable. In contrast, even in the case where sender and receiver have complete common ground on the set of possible sender goals, the planning biases of the sender (i.e., the probabilistic dependencies between sender's goals and actions, and world, as well as how those dependencies change over time) and all events that are relevant to the interpretation of the sender's actions, still the task of communication is computationally intractable. Future research may address the additional computational complexities introduced by establishing this common ground in the first place. This could be done by building an explicit model of that process and submitting it to the same sort of complexity analyses as we have performed here for the SENDER and RECEIVER models.

Last, we reflect on our choice of formalism: we opted for Bayesian networks as a formal representation of communicators' situational knowledge and we used the mathematical notion of posterior probability to define the intuitive notion of "inference to the best explanation." This choice of probabilistic formalism aligns our models with the current trend in cognitive neuroscience, which assumes that the brain implements its cognitive functions by means of probabilistic computations (Wolpert and Ghahramani, 2000; Ma et al., 2006), including the cognitive function of inferring people's goals and intentions from their actions (Cuijpers et al., 2006; Baker et al., 2007, 2009; Kilner et al., 2007a,b). Importantly, our complexity results should not be seen as bound to this formalism. It is well known that knowledge structures can be represented using other formalisms, such as (non-classical) logics or as constraint networks (van Ditmarsch et al., 2007; Russell and Norvig, 2009). In those cases, "Inference to the best explanation" could be defined in terms of logical abduction (e.g., Bylander et al., 1991; Eiter and Gottlob, 1995; Nordh and Zanuttini, 2005, 2008) or constraint satisfaction (e.g., Thagard and Verbeurgt, 1998; Thagard, 2000). The present results would apply to those alternative formalisms, provided that the same basic structural information is encoded in the representations of those alternative formal languages. If models identical or similar to those analyzed in this paper get expressed in the different formalisms, chances are that all such formal model variants can be transformed into each other via mathematical reductions not unlike the one we used to prove that the Sender and Receiver models are of the same computational complexity as the known intractable logic problem called Satisfiability (see Computational Models and Computational Complexity Analyses). One would not want complexity predictions to depend on the formal language chosen to express a model, in the same way that one would not want a verbal theory to lead to different predictions if it is expressed in English or in Dutch.

CONCLUSION

Recent accounts of human communication have focused on the motoric abilities of individual agents (Iacoboni et al., 2005), as if the meaning conveyed by a pointing finger were an intrinsic property of that action. Others have correctly highlighted the fact that, when unconstrained, human communication is intractable (Levinson, 1995, 2006). Here we have tried to move our knowledge forward, showing that those positions are case limits of a continuum as a function of the constraints applicable to the inferences involved in human communication. We believe the approach exemplified in this paper has the potential to formally define sources of computational complexity, opening the way to tackle some hard problems in intentional communication. Acquiring this knowledge appears fundamental to generate a neurobiologically grounded formal account of the computational mechanisms supporting our communicative abilities, which is a necessary step to understand the biological and cognitive bases of human sociality.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their valuable comments on an earlier version of this article. Johan Kwisthout was supported by the OCTOPUS project under the responsibility of the Embedded Systems Institute. Mark Blokpoel was supported by a DCC PhD grant awarded to Iris van Rooij and Ivan Toni. Todd Wareham was supported by NSERC Personal Discovery Grant 228104. Jakub Szymanik was supported by NWO (VICI grant #277-80-001). Ivan Toni was supported by NWO (VICI grant #453-08-002).

REFERENCES

Aaronson, S. (2005). NP-complete problems and physical reality. *SIGACT News Complex. Theory Column* 36, 30–52.

Arora, S. (1998). Polynomial time approximation schemes for Euclidean TSP and other geometric problems. *J. ACM* 45, 753–782.

Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A., and Protasi, M. (1999). *Complexity and Approximation: Combinatorial Optimization Problems and their Approximability properties*. Berlin: Springer.

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

Baker, C. L., Tenenbaum, J. B., and Saxe, R. R. (2007). "Goal inference as inverse planning," in *Proceedings of the 29th Annual Cognitive Science Society*, eds D. S. McNamara and J. G. Trafton (Austin, TX: Cognitive Science Society), 779–784.

Barr, D. J., and Keysar, B. (2005). "The paradox of egocentrism in language use," in *Figurative Language Comprehension: Social and Cultural Influences*, eds H. L. Colston and A. N. Katz (Mahwah, NJ: Erlbaum), 21–42.

Blokpoel, M., Kwisthout, J., van der Weide, T., and van Rooij, I. (2010). "How action understanding can be rational, Bayesian and tractable," in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (Austin, TX: Cognitive Science Society) 1643–1648.

Bylander, T., Allemang, D., Tanner, M. C., and Josephson, J. R. (1991). The computational-complexity of abduction. *Artif. Intell.* 49, 25–60.

Charniak, E., and Shimony, S. E. (1990). Probabilistic semantics for cost based abduction. In Proceedings of the eighth National conference on Artificial intelligence, AAAI Press 106–111.

Chater, N., Oaksford, M., Nakisa, R., and Redington, M. (2003). Fast, frugal and rational: how rational norms explain behavior. *Organ. Behav. Hum. Decis. Process* 90, 63–86.

Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: where next? *Trends Cogn. Sci.* 10, 292–293.

Cuijpers, R., Schie, H. T. V., Koppen, M., Erlhagen, W., and Bekkering, H. (2006). Goals and means in action observation: a computational approach. *Neural Netw.* 19, 311–322.

Dayan, P. (2008). "The role of value systems in decision making," in *Better Than Conscious? Decision Making, the Human Mind, and Implications For Institutions*, eds C. Engel and W. Singer (Frankfurt: MIT Press), 51–70.

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

Eiter, T., and Gottlob, G. (1995). The complexity of logic-based abduction. *J. ACM* 42, 3–42.

Erlhagen, W., Mukovskiy, A., and Bicho, E. (2006). A dynamic model for action understanding and goal-directed imitation. *Brain Res.* 1083, 174–188.

Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor, J. A. (2000). *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press.

Fortnow, L. (2009). The Status of the P versus NP Problem. *Commun. ACM* 52, 78–86.

Frank, M., Goodman, N. D., and Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychol. Sci.* 20, 579–585.

Galantucci, B., and Garrod, S. (2011). Experimental semiotics: a review. *Front. Hum. Neurosci.* 5:12. doi: 10.3389/fnhum.2011.00011

Garey, M. R., and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: Freeman.

Gershman, S. J., and Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* 20, 1–6.

Ghahramani, Z. (1998). "Learning dynamic Bayesian networks," in *Adaptive Processing of Temporal Information*, eds C. L. Giles and M. Gori (Berlin: Springer-Verlag). [Lecture Notes in Artificial Intelligence], 168–197.

Hamilton, M., Müller, M., van Rooij, I., and Wareham, T. (2007). "Approximating solution structure," in *Structure Theory and FPT Algorithmics for Graphs, Digraphs, and Hypergraphs*, eds E. Demaine, G. Z. Gutin, D. Marx, and U. Stege (Schloss Dagstuhl: Internationales Begegnungs- und Forschungszentrum fur Informatik (IBFI)). [Dagstuhl Seminar Proceedings no. 07281].

Haselager, W. F. G. (1997). *Cognitive Science and Folk Psychology: The Right Frame of Mind*. London: Sage.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J., and Rizzolatti, G. (2005). Grasping the intentions of others with one's owns mirror neuron system. *PLoS Biol.* 3, e79. doi: 10.1371/journal.pbio.0030079

Iacoboni, M. Woods, R. P., Brass, M., Bekkering, B., Mazziotta, J. C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science* 286, 2526–2528.

Jacob, P., and Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends Cogn. Sci. (Regul. Ed.)* 9, 21–25.

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007a). The mirror-neuron system: a Bayesian perspective. *Neuroreport* 18, 619–623.

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007b). Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166.

Kwisthout, J., Wareham, T., and van Rooij, I. (2011). Bayesian intractability is not an ailment that approximation can cure. *Cogn. Sci.* doi: 10.1111/j.1551-6709.2011.01182.x

Levinson, S. C. (1995). "Interactional biases in human thinking," in *Social Intelligence and Interaction*, ed. E. Goody (Cambridge: Cambridge University Press), 221–260.

Levinson, S. C. (2006). "On the human 'interactional engine'," in *Roots of Human Sociality: Culture Cognition, and Interaction*, eds N. J. Enfield and S. C. Levinson (London: Berg), 39–69.

Lueg, C. (2004). "Looking under the rug: context and context-aware artifacts," in *Cognition and Technology. Co-existence, Convergence and Co-evolution*, eds B. Gorayska and J. L. Mey (Amsterdam: John Benjamins).

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438.

Marr, D. (1982). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. San Franciso, CA: W. H. Freeman and Company.

Marx, D. (2008). Parameterized complexity and approximation algorithms. *Comput. J.* 51, 60–78.

Miall, R. C., and Wolpert, D. (1996). Forward models for physiological motor control. *Neural Netw.* 9, 1265–1279.

Nordh, G., and Zanuttini, B. (2005). "Propositional abduction is almost always hard," in *28 Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-2005)*, Edinburgh, 534–539.

Nordh, G., and Zanuttini, B. (2008). What makes propositional abduction tractable. *Artif. Intell.* 172, 1245–1284.

Oztop, E., Wolpert, D., and Kawato, M. (2005). Mental state inference using visual control parameters. *Cogn. Brain Res.* 22, 129–151.

Paine, R. W., and Tani, J. (2004). Motor primitive and sequence self-organization in a hierarchical recurrent neural network. *Neural Netw.* 17, 1291–1309.

Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–226.

Pylyshyn, Z. (ed.) (1987). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Norwood: Ablex Publishing.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670.

Russell, S., and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*, 3rd Edn. Upper Saddle River, NJ: Prentice Hall.

Sanborn, A. N., Griffiths, T. L., and Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychol. Rev.* 117, 1144–1167.

Shafto, P., and Goodman, N. (2008). "Teaching games: statistical sampling assumptions for learning in pedagogical situations," in *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*, Austin, TX.

Thagard, P. (2000). *Coherence in Thought and Action*. Cambridge, MA: MIT Press.

Thagard, P., and Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cogn. Sci.* 22, 1–24.

Toni, I., de Lange, F. P., Noordzij, M. L., and Hagoort, P. (2008). Language beyond action. *J. Physiol. Paris* 102, 71–79.

Uithol, S., van Rooij, I., Bekkering, H., and Haselager, P. (2011). What do mirror neurons mirror? *Philos. Psychol.* 1–17. doi: 10.1080/09515089.2011.562604

van Ditmarsch, H., van de Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*, Vol. 337. Berlin: Springer. [Synthese Library Series].

van Rooij, I. (2008). The tractable cognition thesis. *Cogn. Sci.* 32, 939–984.

van Rooij, I., Evans, P., Müller, M., Gedge, J., and Wareham, T. (2008). "Identifying sources of intractability in cognitive models: an illustration using analogical structure mapping," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, eds B. C Love, K. McRae, and V. M. Sloutsky (Austin, TX: Cognitive Science Society), 915–920.

van Rooij, I., and Wareham, T. (2008). Parameterized complexity in cognitive modeling: foundations, applications and opportunities. *Comput. J.* 51, 385–404.

van Rooij, I., Wright, C., and Wareham, H. T. (2010). Intractability and the use of heuristics in psychological explanations. *Synthese*. doi: 10.1007/s11229-010-9847–9847

Vul, E., Goodman, N. D., Griffiths, T. L., and Tenenbaum, J. B. (2009). "One and done? Optimal decisions from very few samples," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (Austin, TX: Cognitive Science Society), 148–153.

Wolpert, D. M., and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat. Neurosci.* 3, 1212–1217.

Yoshida, W., Dolan, R. J., and Friston, K. J. (2008). Game theory of mind. *PLoS Comput. Biol.* 4, e1000254. doi: 10.1371/journal.pcbi.1000254

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 July 2010; accepted: 16 May 2011; published online: 30 June 2011.

Citation: van Rooij I, Kwisthout J, Blokpoel M, Szymanik J, Wareham T and Toni I (2011) Intentional communication: Computationally easy or difficult? Front. Hum. Neurosci. ***5****:52. doi: 10.3389/fnhum.2011.00052*

Copyright © 2011 van Rooij, Kwisthout, Blokpoel, Szymanik, Wareham and Toni. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

APPENDIX

In this Appendix we explain general concepts, notation and terminology that are used in the main text. In particular, we introduce the mathematical notions of dynamic Bayesian Networks and a special case of such networks, called Bayesian Inverse Planning (BIP-)Bayesian Networks (Baker et al., 2007, 2009; Blokpoel et al., 2010). Further, we explain the conceptual bases of computational complexity analysis, including definitions and proof techniques. Last, we present formal proofs for all results in the main text.

PRELIMINARIES FROM BAYESIAN MODELING

Bayesian or probabilistic networks are tools for modeling uncertain knowledge (Jensen and Nielsen, 2007). A Bayesian network B is a graphical structure that models a set of stochastic variables, the (in-)dependencies among these variables, and a joint probability distribution over these variables. B includes a directed acyclic graph N, modeling the variables and (in-)dependencies in the network, and a set of parameter probabilities Γ in the form of conditional probability tables (CPTs), capturing the strengths of the relationships between the variables. A CPT is associated with each variable in the network and contains the probability distribution of that variable X for each joint value assignment to its parents $Y_1 \ldots Y_n$, denoted by $\Pr(X|Y_1,\ldots,Y_n)$. The network models a joint probability distribution $\Pr(V)$ over its variables V, where the arcs denote dependencies between variables and lack of arcs between variables denote independencies in the probability distribution. **Figure A1** illustrates an example network that models the dependencies between the variables WetGrass (G), SprinklerOn (S), Rain (R) and PowerFailure (P). The grass can be wet both due to rain and to the sprinkler being on. Whether the sprinkler is on depends on whether there is a power failure. To model abduction there are two types of Bayesian inference we use in our proofs. For completeness we introduce their formal input-output mappings here.

Most probable explanation

Input: A probabilistic network $B = (\mathbf{N}, \Gamma)$, where $\mathbf{N} = (\mathbf{V}, \mathbf{A})$ and $\mathbf{V}$ is partitioned into a set of evidence nodes $\mathbf{E}$ with a joint value assignment $\mathbf{e}$ and an explanation set $\mathbf{M}$, such that $\mathbf{E} \cup \mathbf{M} = \mathbf{V}$.
Output: What is the most probable joint value assignment $\mathbf{m}$ to the nodes in $\mathbf{M}$ given evidence $\mathbf{e}$?

Partial maximum a-posteriori probability

Input: A probabilistic network $B = (\mathbf{G},\Gamma)$, where $\mathbf{V}$ is partitioned into a set of evidence nodes $\mathbf{E}$ with a joint value assignment $\mathbf{e}$, a set of intermediate nodes $\mathbf{I} \neq \varnothing$ and an explanation set $\mathbf{M}$, such that $\mathbf{E} \cup \mathbf{I} \cup \mathbf{M} = \mathbf{V}$.
Output: What is the most probable joint value assignment $\mathbf{m}$ to the nodes in $\mathbf{M}$ given evidence $\mathbf{e}$?

While Bayesian networks denote static knowledge, they can be made dynamic by incorporating a discrete notion of time. In *dynamic* Bayesian networks (Grahamani, 1998), sequences of variables are used that are indexed by a time stamp; variables with the same time stamp t form a *time slice*. In addition to the static dependencies between variables in the same time slice, dynamic dependencies may then be modeled by arcs between variables in different time slices. For example, if nothing happens, wet grass will dry up eventually; a power failure will probably be solved somewhere in the future; if it's raining now, chances are high that it will be raining as well in the near future. Dynamic dependencies can be between similar or different variables in each time slice; for example the temperature in time slice $t + 1$ will depend on both the temperature and the state of the thermostat in slice t, and vice versa. In **Figure A2** the example network is enhanced with dynamic dependencies.

A BIP-Bayesian network (BIPBN) is a dynamic Bayesian network with specific connectivity, that models goal inference in the form of inverse planning (Baker et al., 2007, 2009). In a BIPBN $\mathbf{D}$ each slice consists of a state variable $S_t \in \mathbf{S}$ and action variable

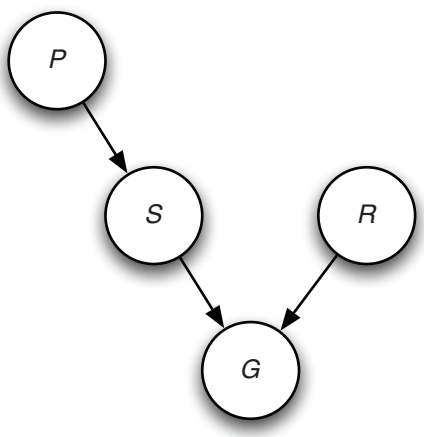

FIGURE A1 | A Bayesian network denoting static knowledge. This BN contains four variables: *P, S, R,* and *G*. These variables are dependent on each other as denoted by the arcs: *S* depends on *P, G* depends on *S* and *R*.

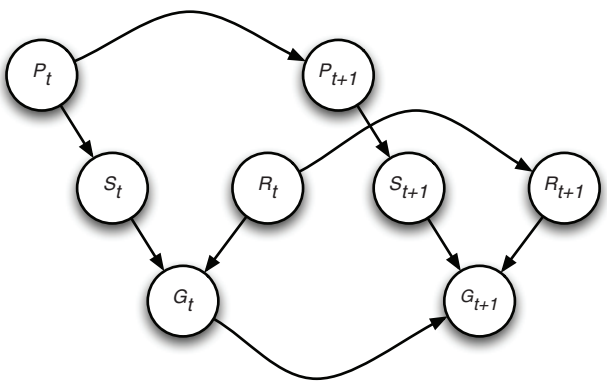

FIGURE A2 | Bayesian network denoting dynamic knowledge. This BN contains four variables: *P, S, R,* and *G*. These variables exist multiple times, once each time slice. As in **Figure 1** their dependencies are denoted by arcs, but in addition there are also dependencies between various variables in different time slices: P_{t+1} depends on P_t, etcetera. However, S_{t+1} is independent of S_t in this example.

$A_t \in \mathbf{A}$. Additionally there is a set of goal variables **G** that contains an arbitrary number of variables that encode the goal(s). In this framework, at a particular time t, the action A_t depends on the current state S_t and on (at least one) goal variable in **G**. State variables S_{t+1} depend on the previous state S_t and action variable A_t. See **Figure A3** for a graphical illustration.

A BIPBN is based on the assumption that agents choose actions that maximize the probability that their goals are achieved. In particular, a BIPBN assumes that agents solve a Markov Decision Problem (Bellman, 1957) to achieve their goals, i.e., they pick exactly these set of actions $A_1,\ldots,A_t$ for which $\Pr(A_1,\ldots,A_t \mid G_1,\ldots,G_K)$ is maximal. Given this assumption, a prior probability distribution on the goals, and Baye's rule, we can infer the probability of the agent's goals, given an observation of its actions, since $\Pr(\mathbf{G} \mid \mathbf{A}) \propto \Pr(\mathbf{A} \mid \mathbf{G}) \Pr(\mathbf{G})$ according to Bayes rule.

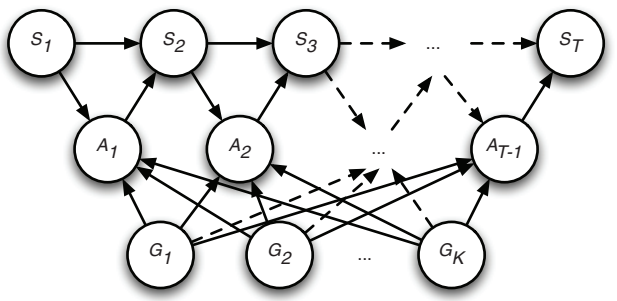

FIGURE A3 | [Same as Figure 2 in main text] The BIP-Bayesian network with T time slices: S_t where $t \in [1, T]$, is a state variable at time t; A_t where $t \in [1, T-1]$, is an action variable at time t. The goal variables, G_1, G_2, …, G_K are fixed over the different time slices. Arrows indicate probabilistic dependencies between variables in the network. Note that no direct dependencies exist between states and goals, but that any indirect dependencies between states and goals are mediated by action variables.

Table A1 | Complexity results for RECEIVER.

RECEIVER	—	$\lvert\mathbf{G}_i\rvert$	$\lvert\mathbf{G}_c\rvert$	$\lvert\mathbf{G}_i\rvert$, $\lvert\mathbf{G}_c\rvert$
–	*NP-hard* Theorem A	*fp-intractable*	*fp-intractable*	*fp-tractable* Theorem C
$\lvert\mathbf{A}\rvert$	*fp-intractable*	*fp-intractable* Corollary A	*fp-intractable*	*fp-tractable*
$1-p$	*fp-intractable*	*fp-tractable* Corollary B	*fp-intractable*	*fp-tractable*
$\lvert\mathbf{A}\rvert$, $1-p$	*fp-intractable*	*fp-tractable*	*fp-intractable* Theorem D	*fp-tractable*

Cells without theorem or corollary are implied by Lemma 2.

Table A2 | Complexity results for SENDER.

SENDER	—	$\lvert\mathbf{G}_i\rvert$	$\lvert\mathbf{G}_c\rvert$	$\lvert\mathbf{G}_i\rvert$, $\lvert\mathbf{G}_c\rvert$
–	*NP-hard* Theorem B	*fp-intractable*	*fp-intractable*	*fp-tractable*
$\lvert\mathbf{A}\rvert$	*fp-intractable*	*fp-intractable* Corollary C	*fp-intractable*	*fp-tractable* Theorem E
$1-p$	*fp-intractable*	*fp-intractable*	*fp-intractable*	*fp-intractable* Theorem F
$\lvert\mathbf{A}\rvert$, $1-p$	*fp-intractable*	*?*	*fp-intractable*	*fp-tractable*

Cells without theorem or corollary are implied by Lemma 2.

Some proofs require the degradation of dependencies, i.e., when the dependencies are per definition required in the instance but should not have an effect on the Bayesian inference. The following lemma states how to achieve this:

Lemma 1. *Let X be a variable and let* **P** *denote the parents of X, i.e. the set of variables on which X depends. To degrade the dependencies between X and* $\mathbf{O} \subseteq \mathbf{P}$ *we set the conditional probabilities as follows, where* $\mathbf{P}' = \mathbf{P} \backslash \mathbf{O}$:

$$\forall_{o \in \Omega(O)} \; [\Pr(X \mid O = o, P) = \Pr(X \mid P)]$$

PRELIMINARIES FROM COMPLEXITY THEORY

In computational complexity analyses one studies the amount of computational resources required to compute (or solve) a problem $P{:}I \rightarrow R$, mapping inputs in the range I to the domain of outputs R. Our focus here is on the resource time. We express the time complexity of a problem as a function of the size of the input, using the Big-Oh notation $O(.)$. A function $f(x)$ is said to be $O(g(x))$, if there are constants $c > 0$ and a red number x_0 such that $f(x) \leq cg(x)$ for all $x \geq x_0$. Note how the function $O(.)$ describes an asymptotic upper-bound, as it ignores constants and low-order function terms. For this reason $O(.)$ it is also called the order of magnitude.

We are interested in the time complexity of models in terms of the size of the input. The input i of a problem has size $n = |i|$ which is the number of symbols used in a typical encoding (usually in binary). A problem P is said to be solvable in time $O(g(n))$ if there exists at least one algorithm that solves P in time $O(g(n))$. The time complexity of a problem P is measured by the fastest algorithm that solves P. We will say that a problem P is computationally intractable (for all but small inputs) if the time required to compute it grows excessively fast as a function of input size. To make precise what we mean by "excessively fast", we adopt a definition that is widely used in both computer science and cognitive science:

Definition 1. *Classical (in)tractability. A problem P is said to be tractable if it can be computed in polynomial-time, i.e., time* $O(|i|^{\alpha})$*, where* α *is a constant. If an problem requires super-polynomial time, e.g., exponential-time* $O(\alpha^{|i|})$*, where* α *is a constant, then P is intractable.*

To see why this definition has merit, compare the speed with which polynomial functions (say, $2n$, n^2, or n^3) and an exponential functions (say, 2^n) grow as a function of input size (n). **Table A3** shows how for small n, the numbers n^2 and 2^n do not differ much, and 2^n is even smaller than n^3, but as n grows, 2^n rockets up so fast that is is no longer plausible that a resource limited mind can perform that number of computations in a reasonable time. As reference points, consider that the number of neurons in a human brain is estimated to be 10^{12}, and 10^{27} is about the number of seconds that have past since the birth of the universe. It seems highly unlikely that a human mind could perform this number of operations in only a couple of minutes (which is a generous upper bound on the time scale of most cognitive processes of interest). Even if a human mind could perform

Table A3 | Illustration of how polynomial growth rates (n, $2n$, n^2, n^3) compare with an exponential growth rate (2^n).

n	$2n$	n^2	n^3	2^n
2	4	4	8	4
5	10	25	125	32
10	20	100	1000	1024
20	40	400	8000	1048576
50	100	2500	125000	$>10^{15}$
100	200	10000	1000000	$>10^{30}$
200	400	40000	8000000	$>10^{60}$

as many parallel computations per second as there are neurons in the brain, it would take days for it to complete 10^{18} operations and as much as centuries for it to complete 10^{27} operations. In other words, knowing that an optimization problem is of superpolynomial time-complexity is good reason to consider that optimization problem computationally intractable for all but small input sizes.

NP-hard problems are problems that cannot[1] be solved in polynomial time, and hence are computationally intractable according to above definition. A problem P_1 can be proven NP-hard, by taking a known NP-hard problem P_2 and polynomial reducing it to the problem P_1. A polynomial time reduction from P_2 to P_1 involves the construction of a transformation from P_2 to P_1 such that any solution for the latter also implies a solution for the former. Further, the transformation must be performable by an algorithm that runs in polynomial time. Given such reduction we can conclude that P_1 is solvable in polynomial time only if P_2 is as well. However, because P_2 is not solvable in polynomial time neither can P_1 be.

The reductions in the proofs in Sections "Computational Models and Computational Complexity Analyses" and "Parameterized Computational Complexity Analyses" use the following NP-hard computational problems.

Clique

Input: An undirected graph $G = (V,E)$ where V is ordered and $k \in \mathbb{N} > 0$.
Output: Does there exist a subset $V' \subseteq V$ such that $|V'| = k$ and $\forall_{u,v \in V'}[(u,v) \in E]$?

3-Satisfiability (3SAT)

Input: A tuple (U,C), where C is a set of clauses on Boolean variables in U. Each clause is a disjunction of at most three variables.
Output: Does there exist a truth assignment to the variables in U that satisfies the conjunction of all clauses in C?

Our analyses not only consider (classical) tractability as in Definition 1, but also fixed-parameter tractability. The latter type of complexity assessment is done using the tools and proof techniques from parameterized complexity theory. This mathematical theory is motivated by the observation that many NP-hard problems can be computed by algorithms whose running time is polynomial in the overall input size $|i|$ and non-polynomial only in one or more small aspects of the input. These aspects are called *parameters*. As the main part of the input contributes to the overall complexity in a "good" way, and only the parameters contribute to the overall complexity in a "bad" way, the problem is well-solved even for large inputs provided only that the parameters remain small. This intuitive characterization is captured by the formal notion of fixed-parameter tractability (see also Downey and Fellows, 1999).

Definition 2. *Fixed-parameter (in)tractability. A problem P is said to be fixed-parameter (fp-)tractable for parameter set* $K = k_1, k_2, \ldots, k_m$ if there exists at least one algorithm that computes P for any input of size n in time $f(k_1, k_2, \ldots, k_m)\, n^{\alpha}$ *where f(.) is an arbitrary computable function and* α *is a constant. If no such algorithm exists then P is said to be fixed-parameter (fp-)intractable for parameter set K.*

Proving fixed-parameter tractability is conceptually straightforward: It suffices to produce just one algorithm that computes the problem in fixed-parameter tractable time (see, e.g., Sloper and Telle, 2008, for a review of generic techniques for building such algorithms).

W[1]-hard problems are problems, including a set of parameters, that cannot[2] be solved in fixed parameter time even when all parameters in their set are small, and hence are fixed-parameter intractable according to above definition. A problem K_1-P_1 can be proven W[1]-hard, by taking a known W[1]-hard problem K_2-P_2 and parameterized reducing it to the problem K_1-P_1. A parameterized reduction from K_2-P_2 to K_1-P_1 involves the construction of a transformation from K_2-P_2 to K_1-P_1 such that any solution for the latter also implies a solution for the former. Further, the transformation must be performable by an algorithm that runs in fixed parameter tractable time, and all parameters in K_2 must be a function of a parameter in K_1. Given such reduction we can conclude that K_1-P_1 is solvable in fixed parameter tractable time only if K_2-P_2 is as well. However, because K_2-P_2 is not solvable in fixed parameter tractable time neither can K_1-P_1 be.

The proofs in the Sections "Computational Models and Computational Complexity Analyses" and "Parameterized Computational Complexity Analyses" use the following W[1]-hard parameterized problem.

k-Clique

Input: A undirected graph $G = (V,E)$ where V is ordered and $k \in \mathbb{N} > 0$.
Parameters: k, the size of the desired clique.
Output: Does there exist a subset $V' \subseteq V$ such that $|V'| = k$ and $\forall_{u,v \in V'}[(u,v) \in E]$?

Finally, the following Lemma is used to propagate fp-(in)tractability results derived for one parameter set K to another K'.

Lemma 2. *Let P be a problem that is fp-intractable for the parameter set* $K = \{k_1 \ldots k_n\}$, *then P is also fp-intractable for any subset* $K' \subseteq K$. *Conversely, let P be a problem that is fp-tractable for the parameter set* $K = \{k_1 \ldots k_n\}$, *then P is also fp-tractable for any super-set* $K' \supseteq K$.

[1]This is true, assuming that the class of problems solvable in polynomial time (P) is not equal to the class of problems whose solutions can be verified in polynomial time (NP). This "P ≠ NP" conjecture is believed by most living mathematicians, both on theoretical and empirical grounds (for more details see Garey and Johnson, 1979; Aaronson, 2005; Fortnow, 2009).

[2]This is true, assuming that FPT ≠ W[1] (Downey and Fellows, 1999).

COMPUTATIONAL MODELS AND COMPUTATIONAL COMPLEXITY ANALYSES

Sender

Input: A Bayesian network $B = (N, \Gamma)$, where $\mathbf{S}, \mathbf{A}, \mathbf{G_I}, \mathbf{G_C} \in N$ and Γ is an arbitrary conditional probability distribution over N, a value assignment $\mathbf{g_I}$ for $\mathbf{G_I}$, and a value assignment $\mathbf{g_C}$ for $\mathbf{G_C}$ encoding the communicator's goals. The probabilistic dependencies in N are illustrated in **Figure 4**.

Output: A value assignment a to A, such that $\mathbf{a} = \text{argmax}_{\mathbf{a}} \Pr(\mathbf{A} = \mathbf{a} \mid \mathbf{G_I} = \mathbf{g_I})$ and Receiver $(B, \mathbf{a}, \mathbf{s}) = \mathbf{g_C}$, or $\varnothing$ if no sequence of actions a is possible. Here $\mathbf{s} = \text{argmax}_{\mathbf{s}} \Pr(\mathbf{S} = \mathbf{s} \mid \mathbf{A} = \mathbf{a})$, i.e., the most likely states **s** to follow from the actions.

Receiver

Input: A Bayesian network $B = (N, \Gamma)$, similar as in the Sender network, a value assignment **a** for **A** and a value assignment **s** for **S** encoding the observed actions and states.

Output: The most probable value assignment $\mathbf{g_C}$ to the communicative goals $\mathbf{G_C}$, i.e., $\text{argmax}_{\mathbf{g}} \Pr(\mathbf{G_C} = \mathbf{g_C} \mid = \mathbf{a}, \mathbf{S} = \mathbf{s})$, or $\varnothing$ if $\Pr(\mathbf{G_C} = \mathbf{g_C} \mid \mathbf{A} = \mathbf{a}, \mathbf{S} = \mathbf{s}) = 0$ for all possible values for $\mathbf{G_C}$.

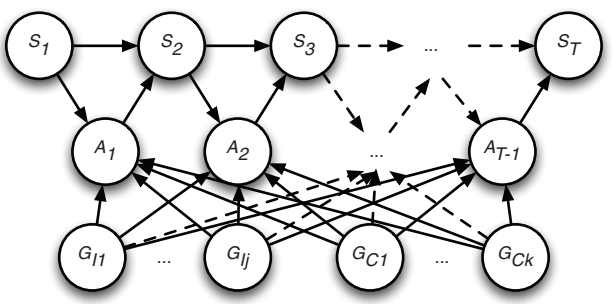

FIGURE A4 | The Bayesian network showing the dependencies between the variables in the Sender and Receiver models. Arrows denote dependencies, and all actions ($A \in \mathbf{A}$) are arbitrarily dependent on one or more (instrumental or communicative) goal variables ($\mathbf{G_I'} \in \mathbf{G_I}$ and/or $\mathbf{G_C'} \in \mathbf{G_C}$).

Theorem A. Receiver is NP-hard.

Proof. Given an instance $\langle G = (V,E) \rangle$ of Clique, construct an instance $\langle B, \mathbf{a}, \mathbf{s} \rangle$ of Receiver as follows (see **Figure A5** for an example):

1. Assume an arbitrary order on the vertices in V such that $V = V_1 < V_2 < \ldots < V_{|V|}$.
2. Assume the basic structure of B as in the definition of Receiver and create $1 + (k-1) + (k(k-1)/2)$ Boolean state variables $S_0, \ldots, S_{(k-1)+k(k-1)/2}$, $(k-1) + (k(k-1)/2)$ Boolean action variables $A_1, \ldots, A_{(k-1)+k(k-1)/2}$, and $k\lceil \log_2 |V| \rceil$ Boolean goal variables $\mathbf{G} = G_1, \ldots, G_{k\lceil \log_2 |V| \rceil}$ such that $\mathbf{G_I} = \varnothing$ and $\mathbf{G_C} = G_1, \ldots, G_{k\lceil \log_2 |V| \rceil}$. Divide $\mathbf{G}$ into k blocks of $\lceil \log_2 |V| \rceil$ Boolean goal variables, i.e., $\mathbf{B} = \mathbf{B_1}, \ldots, \mathbf{B_k}$ where $\mathbf{B_i} = G_{((i-1)\times\lceil \log_2 |V| \rceil)+1}, \ldots, G_{i\times\lceil \log_2 |V| \rceil}$ for $1 \leq i \leq k$. Now define $\nu: \mathbf{B} \rightarrow V$ such that $\nu(\mathbf{B_i})$ returns V_j, where j is the number between 1 and $|V|$ encoded in binary in the (Boolean) values of the $\lceil \log_2 |V| \rceil$ goal-variables in $\mathbf{B_i}$.
3. Set $S_0 = true$ and for $1 \leq i \leq (k-1) + (k(k-1)/2)$, let S_i depend on $S_i - 1$ and $A_i - 1$ and have the following conditional probability:

$$\Pr(S_i = true \mid S_{i-1}, A_{i-1}) = \begin{cases} 1 & \text{if } S_{i-1} = true \text{ and } A_{i-1} = true \\ 0 & \text{otherwise} \end{cases}$$

These state variables effectively function as conjunctions which ensure that there is some assignment $\mathbf{g}$ to $\mathbf{G}$ such that $\Pr(\mathbf{g}) > 0$ iff. all action variables are set to *true*.

4. All dependencies between S_i and A_i are degraded as defined in Lemma 1.
5. For $1 \leq i < k$ let A_i depend on $\mathbf{B_i}$, $\mathbf{B_{i+1}}$, and S_i and have the following conditional probability:

$$\Pr(A_i = true \mid \mathbf{B_i}, \mathbf{B_{i+1}}, S_i) = \begin{cases} 1 & \text{if } \nu(\mathbf{B_i}) < \nu(\mathbf{B_{i+1}}) \text{ and } S_i = true \\ 0 & \text{otherwise} \end{cases}$$

These action variables ensure that a joint value assignment $\mathbf{g}$ to $\mathbf{G}$ has $\Pr(\mathbf{g}) > 0$ iff. the values encoded in the Boolean blocks are distinct.

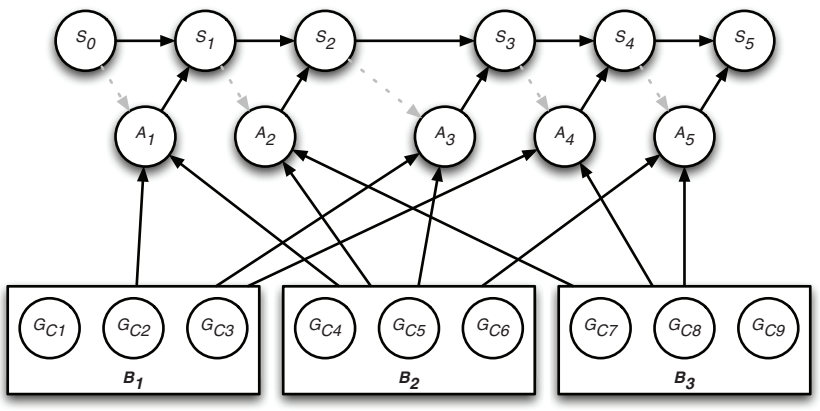

FIGURE A5 | An example of a Clique instance with $k = 3$ and $|V| = 6$ reduced to a Receiver instance. Here Boolean goal variables in the blocks encode vertices from the Clique instance, actions encode "Clique-rules" and states conjoin the rules. Degraded dependencies – i.e., those that have their conditional probabilities set such that they do not influence the Bayesian inference – are depicted by dotted arrows.

6. For $k \leq i \leq (k(k-1)/2)$, let A_i depend on a distinct pair of goal-variable blocks ($\mathbf{B_p}$, $\mathbf{B_q}$) and S_i and have the following conditional probability:

$$\Pr\left(\mathbf{A_i} = true \mid \mathbf{B_p}, \mathbf{B_q}, S_i\right) = \begin{cases} 1 & \text{if } (v(\mathbf{B_p}), v(\mathbf{B_q})) \in E \text{ and } S_i = true \\ 0 & \text{otherwise} \end{cases}$$

These action variables ensure that a joint value assignment **g** to **G** has $\Pr(\mathbf{g}) > 0$ iff. all pair of values encoded in the Boolean blocks is an edge in E from the clique instance.

7. Make the prior probability distribution for each goal variable uniform.
8. Set **a** and **s** such that all action and state variables are assigned *true*.

As the number of conditional probability tables that are constructed by the reduction is proportional to the total number of variables (which is $|\mathbf{S}| + |\mathbf{A}| + |\mathbf{G}| = (1 + (k-1) + \frac{k(k-1)}{2}) + ((k-1) + \frac{k(k-1)}{2}) + k\lceil \log_2 |V| \rceil$) and each table involves at most $1 + 2\lceil \log_2 |V| \rceil$ Boolean variables (resulting in a maximum of $2^{3\lceil \log_2 |V| \rceil} = (2^{\lceil \log_2 |V| \rceil})^3 \leq (2^{1+\log_2 |V|})^3 = (2|V|)^3 = 8|V|^3$ entries per table), the instance of Receiver can be constructed in time polynomial in the size of the given instance of Clique. Moreover, note that in this instance of Receiver $|\mathbf{A}| = (k-1) + [k(k-1)/2]$ and $|\mathbf{G_I}| = 0$.

To prove that the construction above is a valid reduction, we must show that the answer to the given instance of Clique is "Yes" iff. there exists a solution to the constructed Receiver instance unequal to ∅. We will show this we prove both directions of this implication separately:

- If the answer to the given instance of Clique is "Yes", there exists a subset $V' \subseteq V$ such that $|V'| = k$ and $\forall_{u, v \in V'}[(u, v) \in E]$. Let **g** be the assignment to **G** corresponding to the vertices in V' under the assumed order on V. As the vertices in V' are distinct, action variables $A_1, \ldots, A_{k-1}$ will all be *true* with probability 1 relative to **g**. Moreover, as V' is a k-clique and there is thus an edge between each pair of distinct vertices in V', action variables $A_1, \ldots, A_{(k-1)+k(k-1)/2}$ will also be *true* with probability 1 relative to **g**. Given the structure of B, this implies that $\Pr(\mathbf{g}) > 0$, which means that the answer to the constructed instance of Receiver is not empty.
- If the answer to the constructed instance of Receiver is not empty, then there is an assignment **g** to **G** such that $\Pr(\mathbf{g}) > 0$. Given the dependencies in and the conditional probabilities in B, this can only happen if all action variables have value *true* with probability 1. Hence, the values of the goal variables in **g** are not only distinct (by the conditional probabilities in (5) above), but the values encoded in the Boolean blocks also correspond to a set of vertices such that every distinct pair of these vertices is connected by an edge in G (by the conditional probabilities in (6) above), which means that the answer to the given instance of Clique is "Yes." □

Theorem B. Sender is NP-hard.
Proof. Observe that in the Receiver instance constructed in the proof of Theorem A, $\Pr(\mathbf{G_C} = \mathbf{g_C} \mid \mathbf{A} = \mathbf{a}, \mathbf{s})$ for any joint value assignment to **a** to **A** and **s** to **S** other than the assignment that sets all variables in **A** and **S** to *true*. If the output of Sender is not ∅, then there exists a joint value assignment **a** and **s** such that Receiver $(B, \mathbf{a}, \mathbf{s}) \neq \emptyset$ for $\mathbf{g_C}$. By definition of the reduction $\mathbf{g_C}$ then corresponds to a clique in G. This reduces Clique to Sender and thus Sender is NP-hard. □

PARAMETERIZED COMPUTATIONAL COMPLEXITY ANALYSES

Theorem C. Receiver is fp-tractable for parameter set $\{|\mathbf{G_I}|, |\mathbf{G_C}|\}$.
Proof. To calculate the output for for any instance of Receiver one can try out all possible joint value assignments to $\mathbf{G_I} \cup \mathbf{G_C}$ in a time that is only exponential in $|\mathbf{G_I}|$ and $|\mathbf{G_C}|$ (viz. time $O(v^{|\mathbf{G_I} \cup \mathbf{G_C}|})$, where v is the maximum number of possible values per goal variable). As the values of **A** and **S** are given, the conditional probability for each goal value assignment can be computed in a time polynomial in the total number of variables, because all variables are observed. The computational complexity of Receiver is exponential only in the number of instrumental $\mathbf{G_I}$ and communicative goals $\mathbf{G_C}$, thus Receiver is fp-tractable for the parameter set $\{|\mathbf{G_I}|, |\mathbf{G_C}|\}$. □

Corollary A. Receiver is fp-intractable for parameter set $\{|\mathbf{A}|, |\mathbf{G_I}|\}$.
Proof. To prove that Receiver is fp-intractable for parameter set $\{|\mathbf{A}|, |\mathbf{G_I}|\}$, it suffices to provide a parameterized reduction from a W[1]-hard problem, namely κ-Clique, to $\{|\mathbf{A}|, |\mathbf{G_I}|\}$-Receiver. The reduction is exactly the same as the reduction in the proof of Theorem A. That reduction runs in polynomial time and thus also in fixed-parameter tractable time. Furthermore, $|\mathbf{A}|$ is a function of the size k of the requested clique and $|\mathbf{G_I}| = 0$. As κ-Clique is W[1]-hard, $\{|\mathbf{A}|, |\mathbf{G_I}|\}$-Receiver is also W[1]-hard and thus Receiver is fp-intractable for parameter set $\{|\mathbf{A}|, |\mathbf{G_I}|\}$. □

Theorem D. Receiver is fp-intractable for parameter set $\{|\mathbf{A}|, |\mathbf{G_C}|, 1-p\}$.
Proof. In this proof we show that Receiver remains NP-hard even if we allow the probability of the most probable explanation to be as high as $1 - \varepsilon$ for arbitrarily small values of ε. In this proof $|\mathbf{A}|$ is a function of of the size k of the requested clique and $|\mathbf{G_C}| = 1$. This proves Receiver is fp-intractable for the parameter set $\{|\mathbf{A}|, |\mathbf{G_C}|, 1-p\}$, where $1-p$ is the probability of the most probable explanation for **a** and **s** a communicator considers.

Consider a variant of the reduction in the proof of Theorem A. In addition to $G_1, \ldots, G_{k\lceil \log_2 |V| \rceil}$ we create an extra Boolean goal variable G_X. In this variant $\mathbf{G_I} = G_1, \ldots, G_{k\lceil \log_2 |V| \rceil}$ and $\mathbf{G_C} = \{G_X\}$. As before, $\mathbf{G_I}$ is divided into k blocks of $\lceil \log_2 |V| \rceil$ Boolean goal variables $\mathbf{B_1}, \ldots, \mathbf{B_k}$. The conditional probabilities of the action variables are as follows, where $\alpha = (1/|\mathbf{A}|)\varepsilon$:

- For $1 \leq i < k$ let A_i depend on $\mathbf{B_i}$, $\mathbf{B_{i+1}}$ and S_i and have the following conditional probability:

$$\Pr(A_i = true \mid \mathbf{B_i}, \mathbf{B_{i+1}}, S_i, G_X) = \begin{cases} 1 & \text{if } v(\mathbf{B_i}) < v(\mathbf{B_{i+1}}), \quad S_i = true \text{ and } G_X = true \\ 0 & \text{if } v(\mathbf{B_i}) < v(\mathbf{B_{i+1}}), \quad S_i = true \text{ and } G_X = false \\ \alpha & \text{if } v(\mathbf{B_i}) \geq v(\mathbf{B_{i+1}}), \quad S_i = true \text{ and } G_X = true \\ \alpha & \text{if } v(\mathbf{B_i}) \geq v(\mathbf{B_{i+1}}), \quad S_i = true \text{ and } G_X = false \\ 0 & \text{otherwise} \end{cases}$$

These action variables ensure that a joint value assignment **g** to **G** has $\Pr(\mathbf{g}) > 0$ iff. the values encoded in the Boolean blocks are distinct.

- For $k \leq i \leq [k(k-1)/2]$, let A_i depend on a distinct pair of goal-variable blocks ($\mathbf{B_p}$, $\mathbf{B_q}$) and S_i and have the following conditional probability:

$$\Pr(\mathbf{A_i} = true \mid \mathbf{B_p}, \mathbf{B_q}, S_i, G_X) = \begin{cases} 1 & \text{if } (v(\mathbf{B_p}), v(\mathbf{B_q})) \in E, \quad S_i = true \text{ and } G_X = true \\ 0 & \text{if } (v(\mathbf{B_p}), v(\mathbf{B_q})) \in E, \quad S_i = true \text{ and } G_X = false \\ \alpha & \text{if } (v(\mathbf{B_p}), v(\mathbf{B_q})) \notin E, \quad S_i = true \text{ and } G_X = true \\ \alpha & \text{if } (v(\mathbf{B_p}), v(\mathbf{B_q})) \notin E, \quad S_i = true \text{ and } G_X = false \\ 0 & \text{otherwise} \end{cases}$$

These action variables ensure that a joint value assignment **g** to **G** has $\Pr(\mathbf{g}) > 0$ iff. all pair of values encoded in the Boolean blocks is an edge in E from the clique instance.

If G has a k-clique and all action and state variables are observed to be true, then $\Pr(G_X = true) = 1 - |\mathbf{A}|\alpha = 1 - |\mathbf{A}|(1/|\mathbf{A}|\varepsilon) = 1 - \varepsilon$. If G does not have a k-clique, then $\Pr(G_X = false) = 1$. Hence, even if the probability of the most probable value of G_X is at least $1 - \varepsilon$, it is still NP-hard to decide on the most probable value. □

Corollary B. Receiver is fp-tractable for the parameter set $\{|\mathbf{G_I}|, 1 - p\}$.

Proof. The Receiver problem is in fact a special case of the more general Partial MAP problem, where the input is an arbitrary Bayesian network B, in which the set of variables is partitioned into a set of observed variables (called the evidence set **E**), a set of variables for which the most probable explanation is sought (called the explanation set **M**), and a set of variables that are neither in the evidence nor explanation set (called the intermediate set **I**). In the Bayesian network in Receiver, the action and state variables **A** and **S** are observed and form the evidence set **E**, the explanation set **M** consists of the communicative goal variables $\mathbf{G_C}$, and the instrumental goal variables form the intermediate set **I**. Partial Map is NP-hard in general and remains so under severe constraints on the structure of the network; however, Partial MAP is tractable when both the probability of the most probable explanation is high, and the number of variables in the intermediate set is low (Park and Darwiche, 2004).

Because of the inheritance from the Partial MAP problem, the Receiver problem is tractable when both the probability p of the most probable joint value assignment to the goal variables is high and when $|\mathbf{G_I}|$, the number of instrumental goal variables, is bounded, i.e., Receiver is fixed parameter tractable for $\{|\mathbf{G_I}|, 1 - p\}$. □

Observation 1. If Receiver is intractable, then Sender is also intractable.

Corollary C. Sender is fp-intractable for parameter set $\{|\mathbf{A}|, |\mathbf{G_I}|\}$.

Proof. Follows from Corollary A and Observation 1. □

Theorem E. Sender is fp-tractable for parameter set $\{|\mathbf{A}|, |\mathbf{G_I}|, |\mathbf{G_C}|\}$.

Proof. To calculate the output of Sender we can try out all possible joint value assignments **a** to **A**. For all **a** we have to calculate the Receiver(B,**a**,**s**) output $\mathbf{g}'_{\mathrm{C}}$. The correct output is that **a** for which $\Pr(\mathbf{A} = \mathbf{a} \mid \mathbf{G_I} = \mathbf{g_I})$ is maximal. This takes time $O(v^{|\mathbf{A}|})$, where v is the maximum number of possible values per goal variable, because Receiver is fp-tractable for parameter set $\{|\mathbf{G_I}|, |\mathbf{G_C}|\}$ (Theorem C). The computational complexity of Sender is only exponential in the number of actions $|\mathbf{A}|$ when $|\mathbf{G_I}|$ and $|\mathbf{G_C}|$ are small, thus Sender is fp-tractable for the parameter set $\{|\mathbf{A}|, |\mathbf{G_I}|, |\mathbf{G_C}|\}$. □

Corollary D. Sender is fp-intractable for parameter set $\{|\mathbf{A}|, |\mathbf{G_C}|, 1 - p\}$.

Proof. Follows from Theorem D and Observation 1. □

Theorem F. Sender is fp-intractable for parameter set $\{|\mathbf{G_I}|, |\mathbf{G_C}|, 1 - p\}$.

Proof. The proof of this theorem uses a variant of the proof construction by Kwisthout (2009), using a reduction from 3SAT to prove NP-hardness of Parameter Tuning restricted to polytrees[3]; which on its turn was inspired by a similar construction by Park and Darwiche (2004), who proved NP-hardness of Partial Map restricted to polytrees. The proof uses a network construction as in **Figure A6**, in which all A_i model the variables of the 3SAT formula, all $S_1 \ldots S_n$ model the clauses and S_0 acts as a clause selector. The conditional probabilities are constructed such that Bayesian inference on the network solves 3SAT.

Let $m = |C|$ denote the number of clauses of a 3SAT formula and let $n = |U|$ denote the number of variables. The conditional probabilities of the variables $S_1 \ldots S_n$ are such that $\Pr(S_{n+1} = true) = m/n$ if and only if there is a joint value assignment to the variables $A_1 \ldots A_n$ that corresponds to a satisfying truth assignment to the 3SAT formula, or $\Pr(S_{n+1} = true) \leq (m-1)/n$

[3]A polytree is a directed acyclic graph for which there are no undirected cycles when the arc direction is dropped.

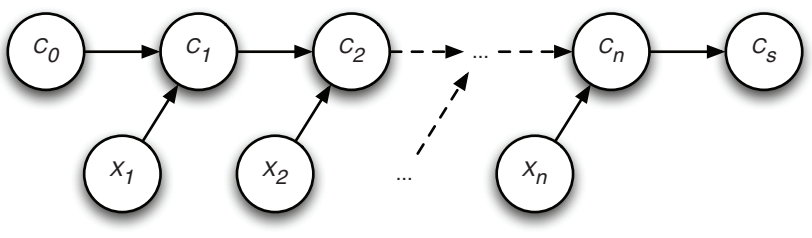

FIGURE A6 | The construction used to prove NP-hardness of Partial MAP, restricted to polytrees, by Park and Darwiche (2004).

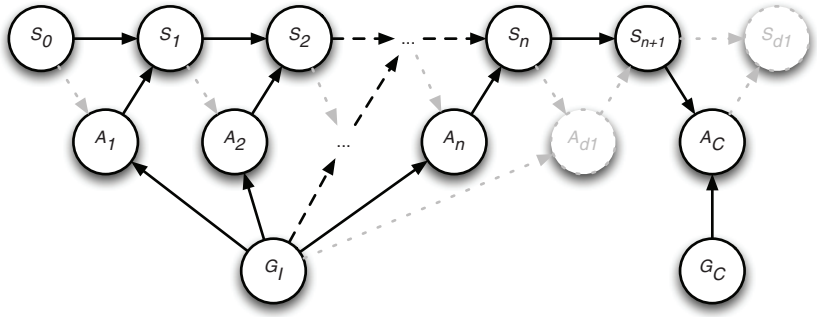

FIGURE A7 | The variant construction used to prove SENDER NP-hard. Degraded dependencies and dummy variables are depicted by dotted lines.

otherwise. We add binary variables S_{n+1}, G_I, G_C and A_C, dummy variables A_{d1} and S_{d1} with the following probability distributions (see **Figure A7**):

- $\Pr(S_{n+1} = true \mid S_n = true) = 1 - \varepsilon + (1 - \frac{m - 1/2}{n}) \times \alpha$

  $\Pr(S_{n+1} = true \mid S_n = false) = 1 - \varepsilon + \frac{m - 1/2}{n} \times \alpha$

  where α is a sufficiently small number to guarantee that all probabilities are in [0,1]. It plays no further role in the proof, so we fix $\alpha = \varepsilon^2$.
- $\Pr(G_I = true) = 1$ and $\Pr(A_i | G_I)$ is uniformly distributed;
- $\Pr(G_C = true) = \Pr(G_C = false) = 1/2$;
- $\Pr(A_C = true \mid S_{n+1} = true, G_C = false) = 1$, $\Pr(A_C = true \mid S_{n+1} = false, G_C = false) = \varepsilon$, and $\Pr(A_C = true \mid S_{n+1} = true, G_C = false) = \Pr(A_C = true \mid S_{n+1} = false, G_C = true) = 1/2\varepsilon$;
- Dependencies between G_I, S_n, S_{n+1}, and A_{d1} are degraded;
- Dependencies between S_{n+1}, A_C, and S_{d1} are degraded;
- All dependencies between S_i and A_i are degraded as defined in Lemma 1.

Using this reduction from 3SAT instances to SENDER we will prove that any joint value assignment **a** to **A** that maximizes the probability of an arbitrary joint value assignment **g** to $\mathbf{G_I} \cup \mathbf{G_C}$ (i.e. a solution to SENDER) also is a solution for the 3SAT instance, even if the probability of that joint value assignment is at least $1 - \varepsilon$.

The following now holds:

$$\begin{aligned}
&\Pr(S_{n+1} = true) = \Pr(S_{n+1} = true \mid S_n = true) \times \Pr(S_n = true) \\
&+ \Pr(S_{n+1} = true \mid S_n = false) \times \Pr(S_n = false) \\
&= \left(1 - \varepsilon + \left(1 - \frac{m - 1/2}{n} \times \alpha\right)\right) \times \Pr(S_n = true) \\
&+ \left(1 - \varepsilon - 1 - \frac{m - 1/2}{n} \times \alpha\right) \times (1 - \Pr(S_n = true)) \\
&= 1 - \varepsilon + \Pr(S_n = true) \times \alpha - \frac{m - 1/2}{n} \times \alpha
\end{aligned}$$

Recall that $\Pr(S_n = true) = m/n$ if the 3SAT formula is satisfiable, and at most $(m-1)/n$ otherwise. Hence, $\Pr(S_{n+1} = true) > 1 - \varepsilon$ if the 3SAT formula is satisfiable, and $\Pr(S_{n+1} = true) < 1 - \varepsilon$ if it is not satisfiable. As we fixed $\alpha = \varepsilon^2$ and given that $m \leq n$ by definition, we have in particular that:

$$1 - 2\varepsilon < \Pr_{\text{UNSAT}}(S_{n+1} = true) < 1 - \varepsilon < \Pr_{\text{SAT}}(S_{n+1} = true) < 1$$

Where $\Pr_{\text{SAT}}(S_{n+1} = true)$ and $\Pr_{\text{UNSAT}}(S_{n+1} = true)$ denote the probability that $S_{n+1} = true$ given that the 3SAT formula is respectively satisfiable or not. Observe that the posterior probability of G_I is independent of the value assignment of the action variables $A_1 \ldots A_n$. Given this independence, if $A_C = true$ then $\Pr(G_C = true) > 1 - \varepsilon$ independent of $\Pr(S_{n+1} = true)$. Also if $A_C = false$ and $\Pr(S_{n+1} = true) > 1 - \varepsilon$ then $\Pr(G_C = false) > 1 - \varepsilon$; and if $A_C = false$ and $\Pr(S_{n+1} = true) < 1 - \varepsilon$ then $\Pr(G_C = false) < 1 - \varepsilon$. Observe that in each of these cases the most probable value assignment to G_C has a probability which is larger than $1 - \varepsilon$, however, the most probable assignment to G_C flips from *true* to *false* depending on the satisfiability of the 3SAT instance.

Thus, if there exists a value assignment **a** to $\mathbf{A} = \{A_1, \ldots, A_n\} \cup A_C$ such that $g = \{G_I = true, G_C = false\}$ is the most probable explanation to $\{G_I, G_C\}$, then the 3SAT instance is satisfiable (viz. those variables in U corresponding to the $A_1 \ldots A_n$ are true iff. $A_i = true$ and false otherwise). Likewise, if the 3SAT instance is satisfiable then there is a value assignment to **A**, with $A_C = false$, such that $\mathbf{g} = \{G_I = true, G_C = false\}$ is the most probable explanation to $\{G_I, G_C\}$.

This proves SENDER fp-intractable for the parameter set $\{|\mathbf{G_I}|, |\mathbf{G_C}|, 1 - p\}$, where $1 - p$ is the minimum probability of all most probable explanations for all possible **a** and **s** a communicator considers. □

REFERENCES

Bellman, R. (1957). A Markovian decision process. *J. Math. Mech.* 6.

Jensen, F. V., and Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs*, 2nd Edn. New York: Springer Verlag.

Kwisthout, J. H. P. (2009). *The Computational Complexity of Probabilistic Networks.* Ph.D. thesis, Faculty of Science, Utrecht University, Utrecht.

Park, J. D., and Darwiche, A. (2004). Complexity results and approximation settings for MAP explanations. *J. Artif. Intell. Res.* 21, 101–133.

Sloper, C., and Telle, J. A. (2008). An overview of techniques for designing parameterized algorithms. *Comput. J.* 51, 122–136.