Journal of Problem Solving

SPECIAL ISSUE

Deep Analogical Inference as the Origin of Hypotheses

Mark Blokpoel,[1] Todd Wareham,[2] Pim Haselager,[1] Ivan Toni,[1] and Iris van Rooij[1]

[1]Radboud University, Donders Institute for Brain, Cognition and Behaviour, The Netherlands, [2]Department of Computer Science, Memorial University of Newfoundland, Canada

Correspondence:
Correspondence concerning this article should be addressed to Mark Blokpoel, via email to m.blokpoel@donders.ru.nl.

Keywords:
abductive inference, hypothesis generation, analogy, computational-level model

Acknowledgments:
The authors thank the editors, Sarah Carruthers and Ulrike Stege, and the two anonymous reviewers for their comments that improved the presentation of the paper. M.B. was funded by a DCC PhD grant awarded to I.v.R., I.T., and P.H. and the Netherlands Organization for Scientific Research (Gravitation Grant 024.001.006 of the Language in Interaction consortium). T.W. was funded by NSERC Discovery Grants 228104-2010 and 228104-2015. I.T. was supported by the Netherlands Organization for Scientific Research (VICI Grant 453-08-002 and Gravitation Grant 024.001.006 of the Language in Interaction consortium).

The ability to generate novel hypotheses is an important problem-solving capacity of humans. This ability is vital for making sense of the complex and unfamiliar world we live in. Often, this capacity is characterized as an inference to the best explanation—selecting the "best" explanation from a given set of candidate hypotheses. However, it remains unclear where these candidate hypotheses originate from. In this paper we contribute to computationally explaining these origins by providing the contours of the computational problem solved when humans generate hypotheses. The origin of hypotheses, otherwise known as abduction proper, is hallmarked by seven properties: (1) isotropy, (2) open-endedness, (3) novelty, (4) groundedness, (5) sensibility, (6) psychological realism, and (7) computational tractability. In this paper we provide a computational-level theory of abduction proper that unifies the first six of these properties and lays the groundwork for the seventh property of computational tractability. We conjecture that abduction proper is best seen as a process of deep analogical inference.

1. INTRODUCTION

In order to interact with their (social) environment, human beings are continuously faced with the problem of making sense of the world they live in. The capacity to formulate an explanation for a given observation is called abductive inference (Peirce, 1974). This type of inference is inherently uncertain and fallible, which is contrasted by deductive inference where the inferences are truths derived from the observation using deduction rules. Abductive inference is considered to be a central part of human cognition (Chater & Oaksford, 2000; Fodor, 1983; Haselager, 1997; Peirce, 1974). Often, this capacity is characterized as an inference to the best explanation (IBE)—selecting the "best" explanation from a set of candidate hypotheses (Chater, 1999; Chater & Manning, 2006; Glass, 2007; Holland, Holyoak, Nisbett, & Thagard, 1986; Lipton, 1991; Thagard, 1988, 2000; van der Helm, 2000). However, accounts of IBE assume that the set of candidate hypotheses is given, and therefore they do not explain the origin of the set of candidate hypotheses, also known as *abduction proper* (Fodor, 2000; Perfors, 2012). Recently, there has been an increased interest among cognitive scientists in developing accounts that do explain the origin of candidate hypotheses (Gentner & Colhoun, 2010; Goodman, Mansinghka, Roy, Bonawitz, & Tenenbaum, 2008; Goodman, Tenenbaum, & Gerstenberg, 2015; Lake, Salakhutdinov, & Tenenbaum, 2015; Tenenbaum, Griffiths, & Kemp, 2006). In this paper we will contribute to these efforts by unifying seven necessary properties of abduction proper in one theory.

The question "where do candidate hypotheses come from?" can be illustrated with problems solved during human communication. For example, imagine two friends in a loud and crowded pub. From across the crowd one friend sees the other making a gesture: she puts the fingertips of her hands together to form a triangular shape. The observer realizes that his friend is going home. The capacity for generating and understanding communicative signals showcases several key properties of the origin of hypotheses. We highlight seven necessary properties of sets of candidate hypotheses: isotropy, open-endedness, novelty, groundedness, sensibility, psychological realism, and computational tractability.

http://dx.doi.org/10.7771/1932-6246.1187

http://dx.doi.org/10.7771/1932-6246.XXXX

Isotropy: Any knowledge that a person has can potentially be relevant for making the abductive inference (Fodor, 1983). In the context of the example above, even seemingly unrelated knowledge about role-playing games might be relevant. Knowing that role-playing games include wizards wearing pointed hats, and the friend likes role-playing games, then one may hypothesize that she is inviting us to come play a game with her.

Open-endedness: The set of candidate hypotheses can contain each hypothesis that a person can in principle generate (Goodman, 1983). That is, it contains all hypotheses that can be generated (in all possible ways) based on all knowledge a person possesses, which in principle can be infinitely many (Fodor & Pylyshyn, 1988). For example, your friend's gesture may mean home or wizard, but it could also mean roof, house, logical and, diving, ship, wedging, beak, space shuttle, diving, hat, etc.

Novelty: The set of candidate hypotheses can contain hypotheses that a person has never generated before (Fodor, 1983; Goodman, 1983). For example, this may be the first time one has encountered a "going home" gesture; hence the hypothesis about its meaning has to be generated *de novo*. This implies that the meaning of a gesture cannot always be inferred by a simple look-up table or priming mechanism (Pickering & Garrod, 2004, 2013). If it cannot be inferred as such, it has to be generated *de novo*.

Groundedness: A candidate hypothesis must have a well-defined relation to a representation of the observation that is to be explained (Lakoff & Johnson, 2003). There are two types of observations that abduction proper can explain: perceptual observations such as the gesture from the example, and internal "observations" such as the scientific conclusion that some elementary particle must exist. Assuming that both types of observation are explained by the same abduction proper capacity, that capacity minimally needs to be able to account for hypotheses being grounded into perceptual representations. In the case of perceptual observations, this property can be construed as classic groundedness (Barsalou, 1999). Whereas groundedness requires the existence of some well-defined relationship between observation and hypothesis, sensibility (see next) requires that the nature of that relationship is explanatory.

Sensibility: A common criticism of accounts of IBE is that picking the "best" hypothesis need not return a "good" hypothesis if the set does not contain any "good" hypotheses (what has been called "the best of a bad lot" by van Fraasen, 1985; and also by Kuipers, 2000). This criticism can be addressed by asserting that the set of candidate hypotheses contains only sensible candidate hypotheses. A "sensible" hypothesis is not just grounded in the sense above, but in principle (in some context) each candidate hypothesis must be able to explain the observation (Kuipers, 2000; van Fraasen, 1985).

Psychological realism: The computational processes that support abduction proper must be psychologically realistic. Whether or not a computational characterization of abduction proper has this property is an empirical question. When it does, the set of candidate hypotheses is naturally constrained to those hypotheses that can (in principle) be inferred by humans.

Computational tractability: Although much of human cognition is computationally impressive, ultimately it is bounded by limited computational power. This implies that any computational account of abductive inference must be computationally tractable (Frixione, 2001; van Rooij, 2008). From a theoretical perspective, this property seems antagonistic to isotropy and open-endedness, yet it is necessary if the theory is to explain how resource-bounded humans can perform abduction proper.

In this paper we present a computational-level theory (Bechtel & Shagrir, 2015; Blokpoel, 2017; Marr, 1982) that aims to unify these seven properties. We propose this unification can be achieved by viewing the origin of hypotheses as a process of deep analogical inference. Whereas a single analogical inference finds one structural relation between two representations, deep analogical inference allows many consecutive and branching analogical inferences that lead to sets of candidate hypotheses. In the main paper, we focus on our theoretical contributions. We present a formal characterization of abductive inference and the origin of hypotheses. Throughout the paper, we will highlight how the computational-level theory incorporates these properties. For an illustrative case study on how the theory can explain abduction proper in communication, we refer the reader to the Appendix where we show how deep analogical inference can explain the interpretation of a communicative signal in a director-matcher-type communication game (de Ruiter, Noordzij, Newman-Norlund, Hagoort, & Toni, 2007).

1.1. ABDUCTIVE INFERENCE

To understand the nature of the origin of hypotheses, it is necessary to understand how this origin relates to IBE. Unless one assumes that all candidate hypotheses are predefined, there must be some process that works either before or in tandem with IBE, providing the candidate hypotheses. This generative process is called abduction proper (Haselager, 1997; Lipton, 1991). Both abduction proper and IBE together make up abductive inference. Following notational conventions from computer science (see Ausiello et al., 1999; and also see van Rooij, Wright, & Wareham, 2012), we can characterize abductive inference as an input–output mapping:

Complete Abductive Inference (informal)
Input: Evidence *e* and knowledge *K*.
Output: The hypothesis *h*, where *h* = Inference to the Best Explanation (*e*, Abduction Proper(*e, K*)).

Based on observed evidence *e* and all internal knowledge *K*, the cognizer generates a hypothesis *h* that can explain the evidence. Complete abductive inference is based on two sub-functions: Inference to the Best Explanation(.) and Abduction Proper(.). Note that Inference to the Best Explanation requires Abduction Proper to provide a set of candidate explanations. Furthermore, because the theory is a functional characterization, we can characterize Inference to the Best Explanation and Abduction Proper separately, even if algorithmically the two functions might be intricately intertwined. To continue, Inference to the Best Explanation can be characterized as follows:

Inference to the Best Explanation *(*informal*)*
Input: Evidence e and a set of candidate hypotheses *H*.
Output: The hypothesis *h* ∈ *H* that best explains *e*.

The nature of IBE has been extensively debated (Glass, 2007; Hanson, 1958; Hobbs, 2004; Lipton, 1991; Peirce, 1974; Thagard, 1991) and many characterizations of the notion of "best" have been proposed, such as "most probable" and "most likely" (Lipton, 1991), "most coherent" (Glass, 2007; Thagard, 2000), "simplest" (Chater, 1999; van der Helm, 2000), or mixtures of these (Holland et al., 1986; Thagard, 1988). Regardless of the nature of IBE, its functioning critically depends on the presumed availability of a set of candidate hypotheses *H*. Without a set of candidate hypotheses *H* from which to pick the best, IBE does not do anything. This is theoretically problematic, because we cannot always presuppose that a set of candidate hypotheses *H* is given. Therefore, a complete account of abductive inference should also specify the origin of hypotheses, i.e., abduction proper:

Abduction Proper *(*informal*)*
Input: Evidence e and knowledge *K*.
Output: A set of candidate hypotheses *H* based on *e* and *K*.

Characterized in this way, the input–output mapping of abduction proper is underspecified because it does not specify the relationship between a set of candidate hypotheses and the evidence and knowledge. In this paper, we build on the structure-mapping theory (SMT) of analogy to specify exactly this relationship.

1.2. THE ANALOGICAL ORIGIN OF HYPOTHESES

Analogical reasoning has been conjectured to lie at the core of the human capacity for understanding the world around them, sometimes with a strong emphasis on embodied–embedded cognition (Lakoff & Johnson, 1999, 2003), in particular in domains that require creative leaps such as language learning and understanding (Gentner & Christie, 2010), concept learning (Gentner, 2010), (insight) problem solving (Gick & Holyoak, 1980, 1983), similarity judgment (Gentner & Markman, 1997; Gentner & Medina, 1998), scientific explanation (Gentner et al., 1997), perception (Chalmers, French, & Hofstadter, 1992; Hesse, 1974; Hofstadter & Sander, 2013), and generalization (Christie & Gentner, 2010). Given that these domains all involve the generation of hypotheses, it suggests that analogical reasoning may be at the foundation of abduction proper.

The goal of this paper is to characterize abduction proper at Marr's (1982) computation level and to unify six properties (excluding computational tractability). Hence, the theory that we present is a characterization of the *what* of abduction proper, and not yet the *how* (Bechtel & Shagrir, 2015). In the last few decades, many accounts of analogical reasoning have been proposed across different levels of explanation. Examples include Tabletop (French & Hofstadter, 1992), Copycat (Hofstadter, 1996), ACME (Holyoak & Thagard, 1989), LISA (Hummel & Holyoak, 1997), and SMT (Gentner, 1983) and its associated Engine (Falkenhainer, Forbus, & Gentner, 1989). For extensive overviews, see French (2002) and Gentner & Forbus (2011). We will build our theory on SMT, as it is one of the few theories that includes a computational-level characterization. Researchers interested in developing algorithmic-level models of abduction proper—models that explain *how* the inferences are computed—can use the model that we present as a constraining guide on possible algorithmic-level theories (cf. Blokpoel, 2017).

The SMT of analogy is a good choice to characterize abduction proper (Gentner, 1983; Gentner & Colhoun, 2010; Gentner & Smith, 2013), because SMT already has the potential to cover three of the seven necessary properties, i.e., novelty, sensibility, and psychological realism. First, a candidate hypothesis in SMT consists of an analogical match between two (relational) representations and possible (projected) inferences from one to the other. Because analogical matches and inferences can cross domains, a candidate hypothesis can result in representations that are novel. The classical solar system to atom analogy illustrates this nicely, where the explanation for planetary revolution is transferred by analogy to explain electron revolution in an atom (see Gentner, 1983). Second, analogy makes a reasonable operationalization of sensibility. If, in an analogical inference, one of the two representations is that of an observation, then the resulting hypothesis can be used to explain that observation by analogy. In the example from the Introduction, the representation resulting from observing the gesture by a friend may match with the representation of home. One can thus explain her gesture, by analogy, as meaning home. Third, SMT has much empirical evidence supporting it and it has been used to model cognition in various domains (Bowdle & Gentner, 2005;

Forbus, Gentner, Everett, & Wu, 1997; Forbus, Gentner, & Law, 1995; Gentner, 1989, 2003a; Gentner & Christie, 2010; Gentner & Markman, 1997; Kuehne, Forbus, Gentner, & Quinn, 2000; Lovett, Gentner, & Forbus, 2006; Wolff & Gentner, 2011). This contributes to the psychological realism of the processes postulated by SMT. Hence, if the set of candidate hypotheses consists only of hypotheses generated by SMT processes, the set is naturally constrained in the sense that it excludes candidate hypotheses that are outside the scope of the psychologically plausible processes.

This leaves four properties yet unexplained: isotropy, open-endedness, groundedness, and computational tractability. We propose an extension of SMT that uses its key processes of *analogical matching* and *projection* to generate a set of candidate hypotheses through deep analogical inference. We conjecture that sets of candidate hypotheses may be built through recursive analogical matching and projection, which we call *deep analogical inference*. This extension of SMT will impart three additional properties to the theory (isotropy, open-endedness, and groundedness), and set the stage for addressing the final property: computational tractability. It does so in the following ways. First, all candidate hypotheses generated by deep analogical inference are potentially grounded, because deep analogical inference guarantees a relationship between the observation and the candidate hypothesis, and the representation of the observation can be perceptual in nature. Second, the set of candidate hypotheses is also open-ended, as the characterization includes all possible hypotheses a person can in principle generate. Third, isotropy is guaranteed because all knowledge representations available to a person serve as a potential link in the chain of deep analogical inference. Finally, by formalizing the theory at the computational level, we will lay the groundwork for investigating under which conditions it is computationally tractable (cf. van Rooij, 2008; van Rooij, Evans, Müller, Gedge, & Wareham, 2008).

In the next section we review alternative accounts of abduction proper, after which we cover the key processes of SMT that the theory of deep analogical inference extends.

1.3. CURRENT ACCOUNTS OF ABDUCTION PROPER

Accounts that aim to explain abduction proper other than the one presented in this paper exist. Although some accounts may have the potential to unify the seven necessary properties, it is not yet clear if or how they do that. Proponents of these accounts may find it valuable to investigate to what extent these properties are already incorporated, or ensure that they are in future iterations of their accounts.

Church: Church (Goodman et al., 2008) is a modeling framework capable of generating hypotheses by performing probabilistic inference over computational expressions (λ-calculus). Because λ-calculus is Turing-complete, there are no restrictions on the hypotheses that Church can generate. An argument can be made that Church incorporates the isotropy, open-endedness, and novelty properties. However, it is not clear how it can incorporate the groundedness, sensibility, psychological realism, and computational tractability properties as this is left to modelers using the framework.

Hierarchical Bayesian models: Hierarchical Bayesian models (Lake et al., 2015; Tenenbaum, Kemp, Griffiths, & Goodman, 2011) form a modeling framework that can generate hypotheses by virtue of grammar or programming-language structures that are built in. Hierarchical Bayesian models are meant to bridge symbolic representations with probabilistic computations. In principle, like Church, hierarchical Bayesian models as a framework have the capacity to unify all seven necessary properties. However, whether or not they do that depends crucially on the structures being able to generate the right sets of hypotheses. The approach, so far, seems not to have focused on identifying which structures lead to the seven properties.

Structure-mapping theory and engine: As mentioned, SMT (Gentner, 1983) can explain three properties: novelty, sensibility, and psychological realism. To explain the remaining properties, the theory needs to be extended. Although there is room for debate as to what extent algorithmic-level incarnations of SMT incorporate some of the properties, to our best knowledge none of the four remaining properties have been addressed at the computational level.

2. STRUCTURE-MAPPING THEORY

The groundwork for the model of abduction proper given in this paper lies in SMT. According to SMT (Gentner, 1983), analogical reasoning consists of finding *analogical matches* between a base and target and then *projecting inferences* from the base to the target. Analogical matches are determined by finding structural overlap between two *relational representations*. These three concepts are used to characterize the theory Analogical Abduction Proper: relational representations, analogical matches, and projection. We briefly introduce them here.

2.1. RELATIONAL REPRESENTATION

Knowledge in SMT is represented relationally, i.e., knowledge is represented in terms of *objects*, *attributes*, *functions*, and most importantly *relations*. Objects such as Ball, Red and Mary may form the basic elements of a representation such as "A girl named Mary kicked the red ball." Attributes and functions such as isGirl(.) and isSphere(.) are relations that can have only one object as their argument and they return, respectively, true or false or an ordinal value. Finally, relations such as hasColor(.,.) and kicked(.,.) can take two or more arguments which can be other relations, attributes, functions, or objects. Using these building blocks, one can define relational representations. For example:

(1) Kicked (isGirl (Mary), hasColor (isSphere (Ball), Red))

2.2. ANALOGICAL MATCHING

An analogical match is defined as the structural overlap between two relational representations. The overlap consists of correspondences between entities in both representations. Some entities, like objects, attributes, and functions, can correspond to any entity of the same type. Relational entities, however, can only correspond to entities of the same type and with the same label. Furthermore, matches in SMT have to be structurally consistent (Gentner, 1983) in the sense that matches have to satisfy the following two constraints:

1. *1:1 correspondence*: Each entity that is part of the match can only be part of one correspondence.
2. *Parallel connectivity*: If an entity is part of the match, then all its arguments should also be part of the match.

Analogical matching can, for example, explain how another child, James, can believe that kicking a can is good pretend play for playing soccer (i.e., kicking a round ball). The following relational representation matches to Representation (1), because kicked(.,.) corresponds and so do all its arguments.

(2) kicked (isBoy (James), hasColor (isCylindrical (Can), Silver))

Note that this match (and analogical matches in general) only works because structural overlap is guaranteed and because labels of objects, attributes, and functions can be ignored. This is how SMT can explain why analogical inferences can transcend domains, yet remain sound.

A high-quality analogical match is one that has high *systematicity*. Systematicity is assumed to be higher the more the analogical match is interconnected and the more deeply nested substructures it contains (Clement & Gentner, 1991; Forbus & Gentner, 1989; Gentner, 1983, 1989). There is much empirical evidence that SMT accurately captures how humans make analogical inferences (see Gentner, 2010; Gentner & Colhoun, 2010; Gentner & Smith, 2013).

2.3. INFERENCE PROJECTION AND VARIABLE INSTANTIATION

Based on an analogical match, it is possible to transfer information from one representation (the base) to the other (the target). The main constraint on projections is that the projected part of the base has to connect to at least one attribute, function, or relation in the overlapping structure. Additional projection constraints exist, e.g., based on goal relevance (Spellman & Holyoak, 1996), adaptability (Keane, 1996), and support and/or extrapolation (Forbus et al., 1997; Gentner, 2003b; Wareham, Evans, & van Rooij, 2011). An important feature of projections is that they can transfer knowledge from one domain to another, because analogical matches can cross domains. Consider the following extension to Representation 1:

(1a) is (kicked (isGirl (Mary), hasColor (isSphere (Ball), Red)), PlayingSoccer)

Representation 1a also matches Representation 2, because kicked(.,.) corresponds. Based on that match is(.,PlayingSoccer) can be projected onto Representation 2 (indicated in bold), further modeling how James can pretend-play soccer with a tin can.

(2a) **is**(kicked (isBoy (James), hasColor (isCylindrical (Can), Silver)), **PlayingSoccer**)

In addition to projection there is a second way to transfer knowledge from the base to the target representation called variable instantiation (Gentner & Medina, 1998). With variable instantiation, objects in the target representation can be replaced by objects from the base representation if they analogically match. A target object that is replaced in this way can be seen, in a sense, as a variable that is instantiated by the value from the base object.

2.4. CANDIDATE HYPOTHESIS

An analogical match, a projection, and a variable instantiation can be combined to form a candidate hypothesis. Such a hypothesis is a quintuple $\langle b,t,m,\rho,\iota\rangle$, where b is the base representation, t the target representation, m the analogical match between them, ρ a projection function that transfers structure from b to t, and ι an instantiation function that replaces objects in t with objects from b. This structure is a *candidate* hypothesis, because there is no guarantee that the information transformed onto the target is correct. This is not a problem for the purpose of characterizing abduction proper, because abduction proper precisely is about generating hypotheses, whereas IBE is about selecting the best hypothesis.

The processes and concepts from SMT are the basic operators used in the computational-level characterization of Analogical Abduction Proper. Analogical Abduction Proper goes beyond SMT, because it characterizes a set of candidate hypotheses compared to a single candidate hypothesis.

3. ABDUCTION PROPER BY DEEP ANALOGICAL INFERENCE

Using the formal notions of representation, matching, inference projection, and variable instantiation from SMT, we can formally characterize Analogical Abduction Proper. The foundation of Analogical Abduction Proper lies in the recursive application of analogical matching and inference projection. We call this Deep Analogical Inference because it consists of (potentially) many consecutive analogical inferences. Analogical Abduction Proper unifies six out of the seven necessary properties. Three are derived properties from SMT: novelty, sensibility, and psychological realism. Three more properties come by virtue of the extension: isotropy, open-endedness, and groundedness. Finally, in the discussion section we reflect on how the extension lays the groundwork for satisfying computational tractability. We highlight the relevant parts of the theory for each property. We explain the theory in a top-down manner so that it is clear what role each sub-function plays in the function in which it is contained.

We start by defining a candidate hypothesis as a quintuple $\langle b,t,m,\rho,\iota\rangle$. Here, m is an analogical match between two relational representations b and t, ρ is the related inference projection, and ι is the variable instantiation. We start by providing a formal characterization of Analogical Abduction Proper and then continue by formalizing each sub-function.

Analogical Abduction Proper
Input: A relational representation of evidence e and a set of relational representations of knowledge K.
Output: A complete set of candidate hypotheses H, where $H = \bigcup_{k \in K}$ Analogical Candidate Hypotheses(e, K, k).

Compared to the informal definition from the Introduction we add the assumption that e is a relational representation of evidence and that K is a set of relational representations of all knowledge. All candidate hypotheses in the output are based on the evidence e, which guarantees that all candidate hypotheses are grounded. The output is based on Analogical Candidate Hypotheses, which returns all possible candidate hypotheses for e relative to a core k. The complete set of candidate hypotheses is the unified set of all possible candidate hypotheses for all cores $k \in K$. This is the first part of the theory that contributes to its isotropy, i.e., by considering all cores.

Analogical Candidate Hypotheses
Input: A relational representation of evidence e, a set of relational representations of knowledge K, and a relational representation of a core k.
Output: A set of candidate hypotheses (relative to k) H_k, where

$$H_k = \bigcup_{\substack{e' \in D(e,K) \\ k' \in D(k,K)}} \bigcup_{\substack{m \in \textsc{Match}(e',k') \\ \rho \in \textsc{Proj}(e',k',m) \\ \iota \in \textsc{Inst}(e',k',m)}} \langle e',k',m,\rho,\iota \rangle$$

Analogical Candidate Hypotheses outputs a set of candidate hypotheses for e relative to a core k. These candidate hypotheses $\langle e',k',m,\rho,\iota\rangle$ are based on every analogical match m, projection ρ, and instantiation ι that can be found between *all possible* representations of the evidence e and *all possible* representations of the core k. We characterize all possible representations of a base representation with Deep Analogical Inference (D), which outputs all representations that can be built from the base representation by recursively making analogical inferences using all pieces of knowledge in K. This is the second part of the theory that contributes to its isotropy, i.e., by considering all candidate hypotheses between all possible representations of evidence and knowledge. Note that each hypothesis contains a well-defined relation between the core and the evidence e (see also Figures 1 and 2). Hence, the theory is grounded even in the classical sense since the evidence can (but need not) be perceptual (Forbus, Gentner, Markman, & Ferguson, 1998). Furthermore, Analogical Candidate Hypotheses outputs an open-ended set of candidate hypotheses in the sense that the set contains all hypotheses that can be generated based on all knowledge a person possesses.

As explained in the previous section, SMT allows for any part of the base that connects to the match to be transferred and instantiated onto the target. Exactly how much is projected and instantiated is still debated in the literature (Gentner & Colhoun, 2010; Gentner & Smith, 2013), but various proposals for characterizations have been made based on goal relevance (Spellman & Holyoak, 1996), adaptability (Keane, 1996), and support and/or extrapolation (Forbus et al., 1997; Gentner, 2003b; Wareham et al., 2011). At the time of writing there are two options for the theory. The first option is to choose one of the (debated) proposals and (possibly incorrectly) assume that it produces the relevant projections and instantiations. However, which projections and instantiations are relevant may vary wildly, hence the debate. The second option defines $\textsc{Match}(e',k')$, $\textsc{Proj}(e',k',m)$ and $\textsc{Inst}(e',k',m)$ such that they actually return all possible matches, projections, and instantiations between e' and k' that conform to SMT in general. In this way, the theory does not exclude potentially relevant candidate hypotheses. Here we choose the second option.

The final two pieces of the puzzle are Deep Analogical Inference and Analogical Augmentation.

Deep Analogical Inference (D)
Input: A relational representation a and a set of relational representations K.
Output: The set of all possible representations of a relative to K:

$$D(a,K) = \begin{cases} \bigcup_{\substack{k \in K \\ \varnothing,}} \mathrm{AA}(\mathrm{a},\mathrm{k}) \cup D(\mathrm{AA}(a,k),K), \\ \\ \textit{if} \quad \underset{\substack{k \in K \\ \text{otherwise}}}{\exists} \mathrm{AA}(a,k) \neq \varnothing \end{cases} \tag{1}$$

Analogical Augmentation (AA)
Input: Two relational representations a and k.
Output: Given the analogical match $m = \text{Match}(k, a)$ with the highest systematicity, return $a' = \rho(a)$, where $\rho \in \text{Proj}(k,a,m)$ is the biggest possible projection. If no match is possible, return $\varnothing$.

The second, given representations a and k, returns an augmentation of a by finding the most systematic match with k and projecting the biggest structure possible from k to a. Deep Analogical Inference recursively applies Analogical Augmentation (AA) as often as possible. Thereby, it returns the set of all possible representations of a. The third and final part of the theory that contributes to its isotropy is that deep analogical inference returns sets of representations that are based on all possible sequences of analogical inferences (i.e., match and projection) with all knowledge. This also contributes to the theory being open-ended. Figure 1 provides an illustration of Deep Analogical Inference.

The four formal characterizations (Analogical Abduction Proper, Analogical Candidate Hypotheses, Deep Analogical Inference, and Analogical Augmentation) presented in this section together form a complete theory of Abduction Proper that unifies six of the seven necessary properties. Figure 2 provides an illustration of Analogical Candidate Explanations. It shows how analogical matches and projections between representations from two spaces of reconceptualized representations (one space for the evidence e and one for the core k) make up a set of candidate hypotheses relative to the core k. Analogical Abduction Proper combines each subset of candidate hypotheses for all $k \in K$.

4. DISCUSSION

In this paper we have proposed that the origin of hypotheses (otherwise known as abduction proper) may lie in deep analogical inference. We identified seven necessary properties that any account of abduction proper should have: isotropy (Fodor, 1983), open-endedness (Goodman, 1983), novelty (Fodor, 1983; Goodman, 1983), groundedness (Barsalou, 1999; Lakoff & Johnson, 2003), sensibility (Kuipers, 2000; van Fraassen, 1985), psychological realism, and computational tractability (Frixione, 2001; van Rooij, 2008). We characterized abduction proper based on processes from the SMT of analogy. This characterization, called analogical

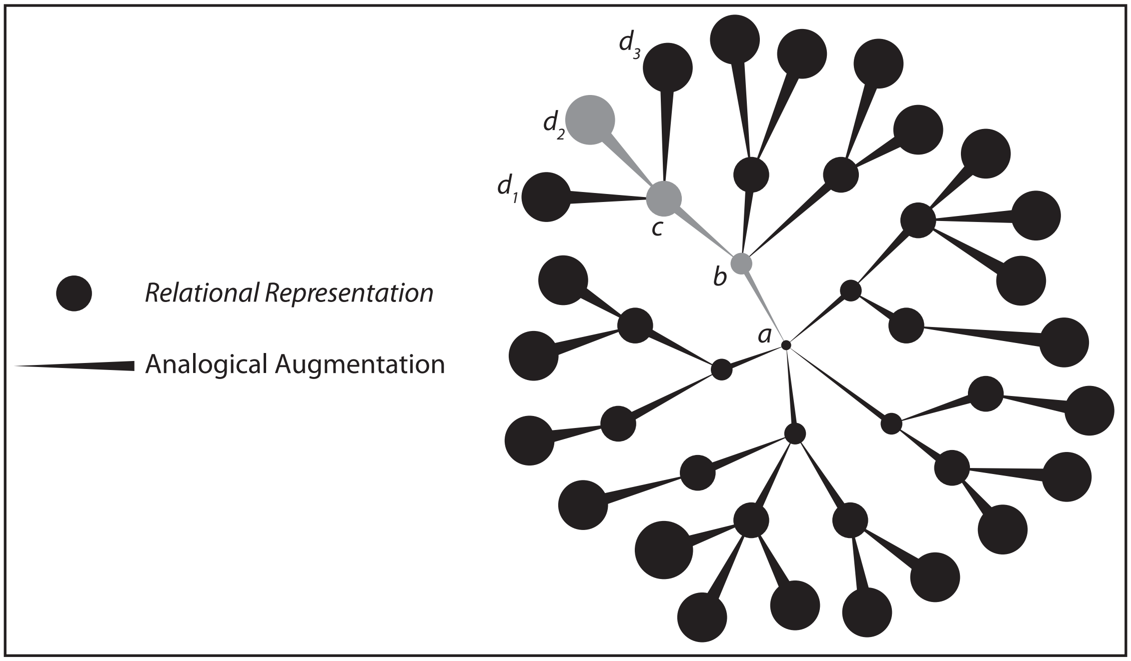


Figure 1.
Deep Analogical Inference *D*. By recursively applying Analogical Augmentation to the relational representation *a*, this function characterizes all possible representations of *a*. The gray sequence highlights one such representation path: *a* analogically matches to some knowledge *k* and can be augmented via analogical projection into *b*. Then *b* is similarly transformed into *c* and *c* into d_1, d_2, and d_3. Finally, the *d*'s do not match to any knowledge $k \in K$ which ends the recursion. Here, we can observe that if *a* is a perceptual representation, then each representation built on top of *a* is grounded in (i.e., has a well-defined relationship with) *a* by virtue of the transitive analogical relation.

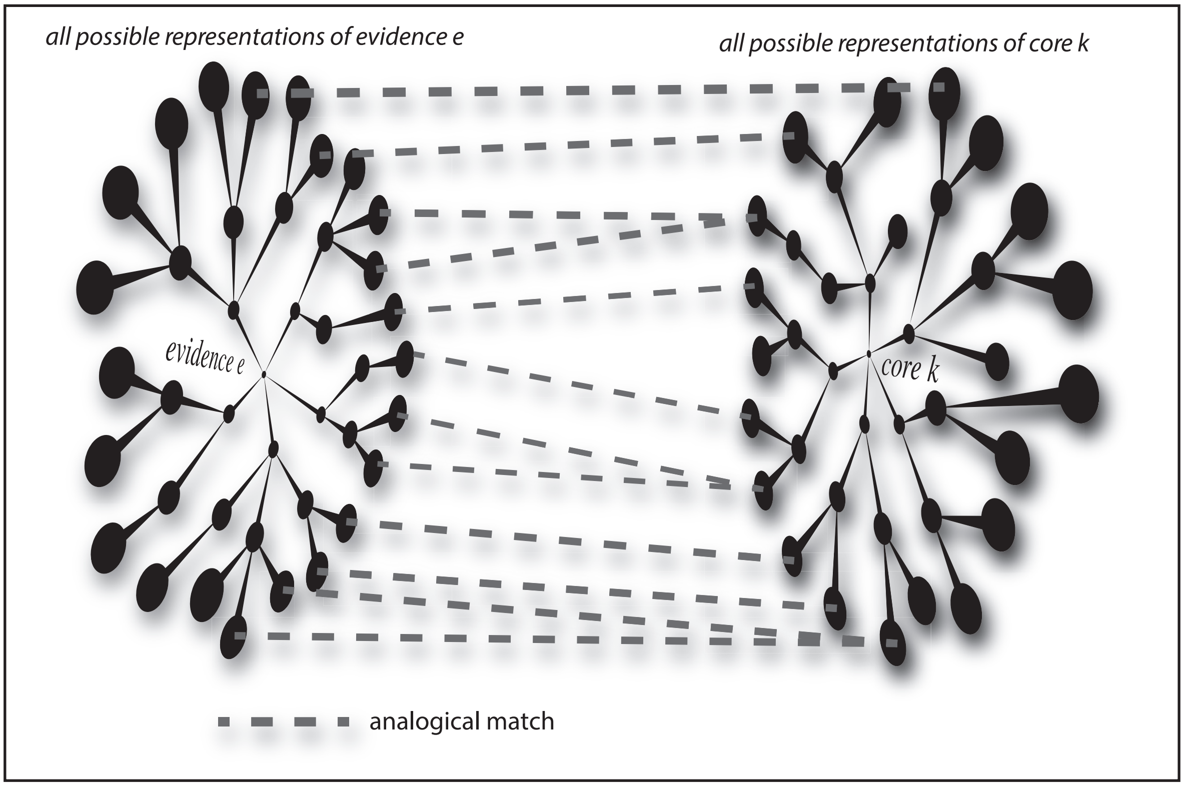

all possible representations of evidence e
all possible representations of core k
evidence e
core k
analogical match

Figure 2.
Analogical Abduction Proper. Analogical matches and projections between representations from two spaces of all possible representations make up a set of candidate hypotheses relative to the core k. On the left is the space of all possible representations for the evidence e and on the right is the space for the core k. The complete figure would include multiple of these analogy "networks," namely one for each core $k \in K$. This figure also illustrates how candidate hypotheses can be grounded in perception.

abduction proper, has six out of seven properties and lays the groundwork for pursuing the computational tractability of abduction proper. This opens up two new research lines: the development of and integration with theoretical accounts of knowledge acquisition and solving the paradox of tractable abductive inference. Before we cover these research lines, we first explain how analogical abduction proper covers six necessary properties of abduction proper.

4.1. NECESSARY PROPERTIES OF ABDUCTION PROPER

The computational-level theory of analogical abduction proper unifies six out of seven necessary properties of abduction proper under one theory. We briefly summarize these properties and explain how analogical abduction proper satisfies them.

Isotropy: Abduction proper is isotropic in the sense that any knowledge that a person has is potentially relevant for some candidate hypothesis. Analogical abduction proper is isotropic because the set of candidate hypotheses it characterizes contains all hypotheses that can be generated through deep analogical inference using every possible knowledge representation available. This means that if a piece of knowledge is possibly relevant, it will be part of at least one deep analogical inference path leading to a candidate hypothesis.

Open-endedness: A set of candidate hypotheses is open-ended if it contains all hypotheses a person can *in principle* infer. Analogical abduction proper generates the set of all possible candidate hypotheses based on all possible deep analogical reconceptualizations and is therefore open-ended.

Novelty: A set of candidate hypotheses is novel if it can contain hypotheses that an individual has never generated before. Because analogical abduction proper is based on analogical inference, it can transfer knowledge from one domain to another, thereby reconceptualizing representations which can lead to candidate hypotheses that the individual has never generated before.

Grounded: For hypotheses to be grounded, any hypothesis needs to contain some well-defined relationship between the representation of the observation and its explanation. Each candidate hypothesis in analogical abduction proper consists of an analogical match and inference between (deeply) reconceptualized knowledge and (deeply) reconceptualized observation. Therefore, analogical abduction proper is grounded, even in the classical sense given that the observation might be perceptual in nature.

Sensible: A sensible candidate hypothesis is one that can be used to explain an observation. In analogical abduction proper, all candidate hypotheses are sensible, as they can relate the observation via (deep) analogical inference to a concept. Because analogical inference is only possible when structural overlap between representations exists, the model avoids candidate hypotheses where anything goes.

Psychological realism: Analogical abduction proper is constrained by the processes that underlie SMT. This means that the set of candidate hypotheses is constrained to those candidate hypotheses that can be generated through analogical matching and projection. The model is empirically supported to the extent that its component processes from SMT have considerable empirical support (Bowdle & Gentner, 2005; Forbus et al., 1995, 1997; Gentner, 1989, 2003a; Gentner & Christie, 2010; Gentner & Markman, 1997; Kuehne et al., 2000; Lovett et al., 2006; Wolff & Gentner, 2011). In addition, we illustrated with an empirical case study how one might explain observations of abduction proper as it occurs in a communicative game (see the Appendix).

Computational tractability: Despite analogical abduction proper being constrained by psychological realism and sensibility, the sets of candidate hypotheses it generates are extremely large, potentially even infinite, due to isotropy and open-endedness. Although at first sight one may reject the theory for this computational intractability, we believe rejection to be too strong a response. The fact that many theories of abductive inference (including analogical abduction proper) are computationally intractable (Bylander, Allemang, Tanner, & Josephson, 1991; Nordh & Zanuttini, 2005) can be seen as a sign that cognitive science is currently unable to solve Fodor's frame problem: How can abductive inference be isotropic, yet computationally explained (Fodor, 2000)? Where Fodor was pessimistic about the chances of computational cognitive science fully solving this problem, we are not and propose a way forward in the section below.

4.2. COMPUTATIONAL TRACTABILITY OF ABDUCTIVE INFERENCE

It is well known that IBE can be computationally intractable (e.g., NP-hard or worse) even for hypothesis spaces that are closed and predefined (Abdelbar & Hedetniemi, 1998; Kwisthout, 2011; Thagard, 2000; Thagard & Verbeurgt, 1998). Inference to the best explanation over open-ended hypothesis spaces, such as those generated by analogical abduction proper, can potentially make IBE more difficult to compute. However, it is not an option to exclude abduction proper from our theories as it is inherently part of abductive inference. This leads to a paradox. People can make abductive inferences quickly, but our best theories of complete abductive inference cannot explain how people can be so quick.

We think that the approach is far from defeated and that the apparent intractability is no reason to reject analogical abduction proper, including its unification of six necessary properties. The reason for our optimism is based on the fact that computational intractability is not a property of the size of the search space (even if it is infinite), but of the ability to search that space efficiently. In fact, it is known that certain functions can become tractable when their search space is appropriately constrained by adding structure to it that can be exploited for efficient search (Downey & Fellows, 1999). This is the basis of a methodology called parameterized complexity analysis. It can be used to analyze under which constraints a computational-level characterization can be tractable (Blokpoel, Kwisthout, van der Weide, Wareham, & van Rooij, 2013; van Rooij, 2008; van Rooij et al., 2008). It has already been successfully applied to analyze models of analogy (van Rooij et al., 2008; Wareham et al., 2011) and communication (Blokpoel et al., 2012; van Rooij et al., 2011). This type of analysis can only be applied to well-defined formal computational-level models, such as the one we presented in this paper. Hence, although our theory is not (yet) computationally tractable, it opens up the possibility for future exploration of ways in which its search-space can be constrained to render it tractable. Such an exploration might, for example, lead to understanding how structure in the set of candidate hypotheses (by virtue of sensibility) may constrain IBE and render it tractable.

4.3. KNOWLEDGE ACQUISITION

Our model crucially depends on the availability of the set of all knowledge K to guarantee isotropy. One might argue that we have shifted the burden of explaining where candidate hypotheses come from to explaining where knowledge comes from. Knowledge acquisition is, however, a different explanatory target. Even if a fully satisfactory and agreed upon account of knowledge acquisition existed, that account would only explain where knowledge comes from, not how knowledge is used to form candidate hypotheses that explain the observations.

The set of knowledge does constrain which candidate hypotheses can be generated. This means that knowledge acquisition may play a constraining role in the theory of analogical abduction proper. Hence, it is important in future research on abduction proper to understand the nature of knowledge acquisition.

5. CONCLUSIONS

The human capacity for generating hypotheses is a phenomenon that is difficult to characterize, mainly because any such characterization will have to be isotropic, open-ended, novel, grounded, sensible, psychologically realistic, and computationally tractable. We have provided a computational-level characterization based on deep analogical inference that unifies six of the seven necessary properties and lays the groundwork for pursuing the seventh, i.e., computational tractability. We believe that this contribution is fundamental to taking the next step towards fully explaining abduction proper, as it establishes firm ground to address future challenges: integrating IBE and abduction proper, developing algorithmic-level explanations of abductive inference, integrating theories of knowledge acquisition, and solving the paradox of tractable abductive inference.

REFERENCES

Abdelbar, A. M., & Hedetniemi, S. M. (1998). Approximating MAPS for belief networks is NP-hard and other theorems. *Artificial Intelligence*, *102*(1), 21–38.

Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A., & Protasi, M. (1999).*Complexity and approximation: Combinatorial optimization problems and their approximability properties*. Berlin: Springer.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660.

Bechtel, W., & Shagrir, O. (2015). The non-redundant contributions of Marr's three levels of analysis for explaining information processing mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322.

Blokpoel, M. (2017). Sculpting computational-level models. *Topics in Cognitive Science*, *10*(3), 641–648.

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

Blokpoel, M., van Kesteren, M., Stolk, A., Haselager, P., Toni, I., & van Rooij, I. (2012). Recipient design in human communication: Simple heuristics or perspective taking? *Frontiers in Human Neuroscience*, *6*, 253.

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193–216.

Bylander, T., Allemang, D., Tanner, M. C., & Josephson, J. R. (1991). The computational complexity of abduction. *Artificial Intelligence*, *49*(1–3), 25–60.

Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental & Theoretical Artificial Intelligence*, *4*(3), 185–211.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology Section A*, *52*(2), 273–302.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335–344.

Chater, N., & Oaksford, M. (2000). The rational analysis of mind and behavior. *Synthese*, *122*(1–2), 93–131.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373.

Clement, C. A. & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, *15*(1), 89–132.

de Ruiter, J. P., Noordzij, M., Newman-Norlund, S., Hagoort, P., & Toni, I. (2007). On the origin of intentions. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Attention and performance. XXII. Sensorimotor foundation of higher cognition* (pp. 593–610). Oxford: Oxford University Press.

de Ruiter, J. P., Noordzij, M. L., Newman-Norlund, S., Newman-Norlund, R., Hagoort, P., Levinson, S. C., & Toni, I. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies*, *11*(1), 51–77.

Downey, R., & Fellows, M. (1999). *Parameterized complexity*. Berlin: Springer.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*(1), 1–63.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Fodor, J., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture. *Cognition*, *28*(1–2), 3–71.

Forbus, K. D., & Gentner, D. (1989). Structural evaluation of analogies: What counts. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 341–348).

Forbus, K. D., Gentner, D., Everett, J. O., & Wu, M. (1997). Towards a computational model of evaluating and using analogical inferences. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 229–234).

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*(2), 141–205.

Forbus, K. D., Gentner, D., Markman, A. B., & Ferguson, R. W. (1998). Analogy just looks like high level perception: Why a domain-general approach to analogical mapping is right. *Journal of Experimental Theoretical Artificial Intelligence*, *10*(2), 231–257.

French, R. M. (2002). The computational modeling of analogy-making. *Trends in Cognitive Sciences*, *6*(5), 200–205.

French, R. M., & Hofstadter, D. R. (1992). Tabletop: An emergent stochastic model of analogymaking. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 175–182).

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

Galantucci, B. (2009). Experimental semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science*, *1*(2), 393–410.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: A review. *Frontiers in Human Neuroscience*, *5*, 11.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*(3), 181–215.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170.

Gentner, D. (1989). Mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). London: Cambridge University Press.

Gentner, D. (2003a). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.

Gentner, D. (2003b). Psychology of analogical reasoning. In *Encyclopedia of Cognitive Science* (pp. 106–112). London: Nature Publishing Group.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), 752–775.

Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B. B., Wolff, P., & Forbus, K. D. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of the Learning Sciences*, *6*(1), 3–40.

Gentner, D., & Christie, S. (2010). Mutual bootstrapping between language and analogical processing. *Language and Cognition*, *2*(2), 261–283.

Gentner, D., & Colhoun, J. (2010). Analogical processes in human thinking and learning. In B. Glatzeder, V. Goel, & A. Müller (Eds.), *Towards a theory of thinking* (pp. 35–48). Berlin: Springer.

Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 266–276.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45–56.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2–3), 263–297.

Gentner, D., & Smith, L. A. (2013). Analogical learning and reasoning. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 668–681). New York, NY: Oxford University Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38.

Glass, D. H. (2007). Coherence measures and inference to the best explanation. *Synthese*, *157*(3), 275–296.

Goodman, N. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.

Goodman, N. D., Mansinghka, V. K., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2008). Church: a language for generative models. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, UAI* (pp. 220–229).

Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts*. Cambridge, MA: MIT Press.

Hanson, N. R. (1958). *Patterns of discovery*. Cambridge: Cambridge University Press.

Haselager, W. F. (1997). *Cognitive science and folk psychology: The right frame of mind*. London: Sage.

Hesse, M. B. (1974). *The structure of scientific inference*. Berkeley, CA: University of California Press.

Hobbs, J. R. (2004). Abduction in natural language understanding. In *Handbook of pragmatics* (pp. 724–741). Oxford: Blackwell Publishing.

Hofstadter, D. R. (1996). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. New York, NY: Basic Books.

Hofstadter, D. R., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. New York, NY: Basic Books.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*(3), 295–355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427–466.

Keane, M. T. (1996). On adaptation in analogy: Tests of pragmatic importance and adaptability in analogical problem solving. *Quarterly Journal of Experimental Psychology Section A*, *49*(4), 1062–1085.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language.

Proceedings of the National Academy of Sciences of the United States of America, *105*(31), 10681–10686.

Kuehne, S., Forbus, K. D., Gentner, D., & Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 770–775).

Kuipers, T. A. F. (2000). *From instrumentalism to constructive realism*. Dordrecht: Kluwer Academic Publishers.

Kwisthout, J. (2011). Most probable explanations in Bayesian networks: Complexity and tractability. *International Journal of Approximate Reasoning*, *52*(9), 1452–1469.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind & its challenge to Western thought*. New York, NY: Basic Books.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lipton, P. (1991). *Inference to the best explanation*. New York, NY: Routledge.

Lovett, A., Gentner, D., & Forbus, K. (2006). Simulating time-course phenomena in perceptual similarity via incremental encoding. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A. C., de Ruiter, J. P., Hagoort, P., & Toni, I. (2009). Recipient design in tacit communication. *Cognition*, *111*(1), 46–54.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2010). Neural correlates of intentional communication. *Frontiers in Neuroscience*, *4*, 188.

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

Peirce, C. S. (1974). *The collected papers of Charles S. Peirce*. C. Hartshorne, P. Weiss, & A. W. Burks (Eds.). Cambridge, MA: Harvard University Press.

Perfors, A. (2012). Bayesian models of cognition: What's built in after all? *Philosophy Compass*, *7*(2), 127–138.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347.

Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology*, *31*(3), 307–346.

Stolk, A., Hunnius, S., Bekkering, H., & Toni, I. (2013a). Early social experience predicts referential communicative adjustments in five-year-old children. *PLoS One*, *8*(8), e72667.

Stolk, A., Noordzij, M. L., Verhagen, L., Volman, I., Schoffelen, J., Oostenveld, R.,… Toni, I. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), 18183–18188.

Stolk, A., Verhagen, L., Schoffelen, J., Oostenveld, R., Blokpoel, M., Hagoort, P.,… Toni, I. (2013b). Neural mechanisms of communicative innovation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14574–14579.

Stolk, A., Verhagen, L., & Toni, I. (2016). Conceptual alignment: How brains achieve mutual understanding. *Trends in Cognitive Sciences*, *20*(3), 180–191.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.

Thagard, P. (1991). The dinosaur debate: Explanatory coherence and the problem of competing hypotheses. In R. Cummins & J. Pollock (Eds.), *Philosophy and AI: Essays at the interface* (pp. 279–300). Cambridge, MA: MIT Press/Bradford Books.

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

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

van der Helm, P. A. (2000). Simplicity versus likelihood in visual perception: from surprisals to precisals. *Psychological Bulletin*, *126*(5), 770–800.

van Fraassen, B. C. (1985). Empiricism in the philosophy of science. In P. Churchland & C. Hooker (Eds.), *Images of science: Essays on realism and empiricism* (p. 245). Chicago, IL: University of Chicago Press.

van Rooij, I. (2008). The tractable cognition thesis. *Cognitive Science*, *32*(6), 939–984.

van Rooij, I., Evans, P., Müller, M., Gedge, J., & 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* (pp. 915–920). Austin, TX: Cognitive Science Society.

van Rooij, I., Kwisthout, J., Blokpoel, M., Szymanik, J., Wareham, T., & Toni, I. (2011). Intentional communication: Computationally easy or difficult? *Frontiers in Human Neuroscience*, *5*(52), 1–18.

van Rooij, I., Wright, C. D., & Wareham, T. (2012). Intractability and the use of heuristics in psychological explanations. *Synthese*, *187*(2), 471–487.

Volman, I. A. C., Noordzij, M. L., & Toni, I. (2012). Sources of variability in human communicative skills. *Frontiers in Human Neuroscience*, *6*, 310.

Wareham, T., Evans, P., & van Rooij, I. (2011). What does (and doesn't) make analogical problem solving easy? A complexity-theoretic perspective. *Journal of Problem Solving*, *3*(2), 30–71.

Wolff, P., & Gentner, D. (2011). Structure-mapping in metaphor comprehension. *Cognitive Science*, *35*(8), 1456–1488.

APPENDIX: A CASE STUDY OF DEEP ANALOGICAL INFERENCE

These supplementary materials illustrate a case study for the computational theory presented in the main paper. We first explain the target phenomenon: the interpretation of an innovative communicative signal in a communication game. We then show how a candidate hypothesis can be generated by using deep analogical inference. We assume that readers are familiar with the theory as presented in Sections 2 and 3 in the main paper.

A.1. A WINDOW INTO ABDUCTION PROPER: THE TACIT COMMUNICATION GAME

The ability to generate novel hypotheses is difficult to isolate and study empirically; however, the phenomenon of communicative innovations provides a window into abduction proper. Communicative innovations are novel signals that have novel meanings (Stolk et al., 2013b); hence they require communicators and listeners to generate (novel) hypotheses about their meaning. They may occur when interlocutors do not have conventionalized signals available. Unfortunately, communicative innovations are often interspersed with conventional signals in daily communication, making it difficult to cleanly observe hypothesis generation. Interest in studying the capacity to generate and understand novel signals has led to the emergence of a research field called *experimental semiotics*. Experimental semioticians have developed many experimental paradigms to isolate and study phenomena related to the emergence of communicative innovations (de Ruiter et al., 2010; Galantucci, 2009; Galantucci & Garrod, 2011; Garrod & Doherty, 1994; Kirby, Cornish, & Smith, 2008). These phenomena range from pair interactions (de Ruiter et al., 2010; Galantucci, 2009) to communities and the evolution of communication systems (Kirby et al., 2008) and from developmental capacities (Stolk, Hunnius, Bekkering, & Toni, 2013a) to neural mechanisms (Noordzij et al., 2010; Stolk et al., 2013b). We focus on observations from the Tacit Communication Game (TCG) for two reasons. First, the TCG was developed to study the emergence of novel signals and recipient design in pair interactions. It therefore provides a clear view on abduction proper as it underlies communication by communicative innovations without adding influences of (cultural) evolution and development. Second, it is one of the most well-studied semiotic paradigms, offering a solid empirical platform for isolating instances of abduction proper in human communication (Blokpoel et al., 2012; de Ruiter et al., 2010; Noordzij et al., 2010; Stolk et al., 2013a, 2013b, 2014; Stolk, Verhagen, & Toni, 2016; Volman, Noordzij, & Toni, 2012).

The TCG is a collaborative task between two participants. To solve the task both participants need certain information, but information is unevenly distributed. This means that one of the participants (the sender) has to confer information to the other participant (the receiver) such that he can solve his part of the joint task. To prevent participants having direct access to conventionalized signals, the TCG has communicators design signals in an unconventional medium. This means that senders are required to generate communicative innovations and receivers are required to understand those innovations. Both communicating and understanding require the ability to generate genuinely novel candidate hypotheses. In this section we give details of both the TCG paradigm and the observations that will form the basis of the case study.

Paradigm: In the TCG, two players cooperate to solve a joint task: placing two tokens, each controlled by one player, correctly on a game board. The 3 × 3 game board has nine locations and each player's token (identified by a color) can vary in shape (see Figure A1). Furthermore, one of the players—and only one—receives privileged information about the correct placement of the two tokens in each trial. This player is the sender and she has to share this privileged information with the other player, the receiver, in order to successfully play the game. Movement is done in turns: first the sender moves her token, then the receiver may move his token, after which a trial ends. Player tokens start at the center of the board and players can move only orthogonally and rotate clockwise 90 degrees[1]. This means that one movement sequence of the sender contains both communicative (i.e., the signal) and instrumental movements (i.e., moving to her goal placement). The TCG is an experimental semiotic paradigm precisely because senders have to generate a signal in an unconventional medium, i.e., by moving and rotating their token on a game board. As explained earlier, this requires senders to generate communicative innovations, because they cannot simply use conventionalized signals that they have learned for other mediums. Consequently, it requires the receiver to be able to understand communicative innovations. Given that generating and understanding of communicative innovations require the ability to generate novel candidate hypotheses, the novel signals and meanings observed in TCG experiments are good subjects for our case study of analogical abduction proper. We detail some of these signals and their meanings later in this section.

A trial of the TCG breaks down into the following sequence (see also Figure A2):

1. *Preparation:* Both players are shown their respective tokens for this trial.

2. *Informing/planning:* The sender is presented with the target positions of both herself and the receiver and she is given time to plan her actions.
3. *Communicating:* The sender's token is placed in the center location of the board and she is given time to execute her planned movement using orthogonal movement and rotation (note that the circle cannot rotate). During this step the receiver observes the sender's token movement.
4. *Task completion:* The receiver's token is placed in the center location of the board and he is given time to move and rotate his token to what he believes is the correct location

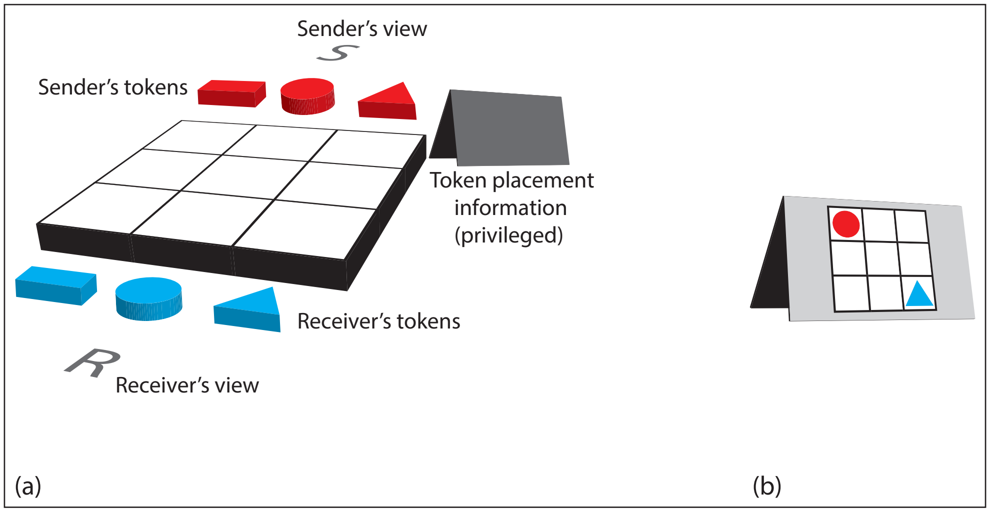


Figure A1.
The Tacit Communication Game. (a) The game is played with the following components: three different shaped tokens for each player (circle, rectangle, and equilateral triangle) and a 3 × 3 game board. The starting position for both players is in the center. (b) The sender has access to privileged information (unavailable to the receiver) about the correct placement of both tokens.

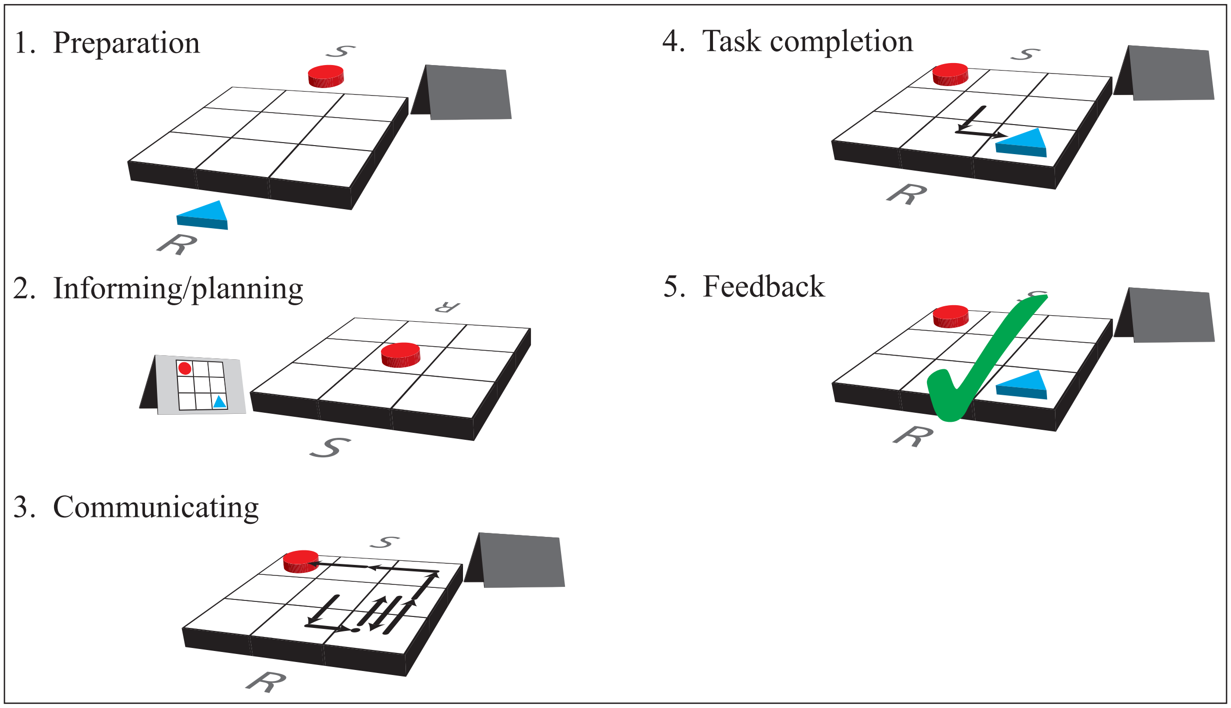


Figure A2.
Turn order. Each turn consists of five steps. Tokens can be moved freely in orthogonal directions, and onto other tokens should they be on the board. They can, if the shape shows it, also be rotated by increments of 90°.

and orientation based on the communicator's observed movement.

5. *Feedback:* Both players receive confirmation on whether or not they solved the joint task. The task is solved when both players' tokens are in their correct location and orientation as shown in Step 2. If both tokens are correctly placed, both players are notified by a green check mark. If at least one token is misplaced, then the trial is lost and both players are notified by a red cross. Note that the players do not receive feedback on what would have been the correct location and orientation.

We next review key observations of the TCG, including specific communicative innovations, which offer a window into abductive inference including abduction proper.

Key observations: We have already explained that due to the unconventional nature of the communication medium in the TCG, senders have to use communicative innovations. Players have been observed to generate a *wide variety* of communicative innovations (see Table A1 for an overview of TCG signals). This variety is reflected most obviously in the signal itself, but more importantly in the signal's meaning, of which any one signal can have many. To appreciate this second observation we have to define the concept of "meaning" in the context of the TCG.

Even though there are only 3 × 3 × 4 (board width × board height × maximum number of different orientations) possible configurations for a token, there are many different ways to represent a token in a position (e.g., "token at (3,2)," "circle on a board," "blue 1 cm by 2 cm object on white 3 cm by 3 cm square with 1 mm black border," etc.). This means that, if the meaning of a signal is a hypothesis about the sender's intended meaning, then all of the different representations of the signal and concept lead to uniquely different meanings. This is reflected in the TCG, where signals that look identical can have different behavioral replies on the receiver's side (Stolk et al., 2013b). It also, counterintuitively, suggests that the same behavioral response on the receiver's part can be the result of different meanings. If the latter should be the case, then the observed diversity is actually an underestimate of the true diversity of communicative innovations.

Table A1.
The wide variety of signals and their meanings in the Tacit Communication Game. This list is compiled from observations made by de Ruiter et al. (2010) and Blokpoel et al. (2012). Here S-token and R-token stand for sender's and receiver's token shape, respectively. These columns indicate with which token shapes (circle, rectangle, or triangle) the signal has been observed.

Signal	Variant	S-token	R-token	Description
Wiggle	Apex	C	R, T	Repetitive motion along an axis from A to B means the apex should point to B.
	Opposite	C	R, T	Repetitive motion along an axis from A to B, where one repetition means the apex should point to B and two repetitions mean the opposite orientation.
	Rotate	C	R, T	The number of repetitive motions is the number of times the receiver should rotate his token.
Exit to point	From target location	C, R, T	R, T	The direction in which the sender leaves the target location is the receiver's orientation.
	From start location	C, R, T	R, T	The direction in which the sender leaves the start location is the receiver's orientation.
Mirror	Exact match	C, R, T	C, R, T	Using the same shaped token, a pause in the receiver's target location and orientation signals the receiver's target.
	Non-match	C, R, T	C, R, T	Using a different shaped token, a pause in the receiver's target location and (as closely matched) orientation signals the receiver's target.
Motion to point		C	T	A fast motion from one side of the board to the receiver's target location signals orientation.
Rotate to rotate		C	R, T	The number of times the sender rotates signals the number of times the receiver should rotate his token.

In many studies, TCG players begin the game with easy trials, i.e., trials where both players have identically shaped tokens. This allows communicators to successfully use "mirror" signals. They can move their own token to the receiver's correct location and orientation, then pause, and then continue to their own position. While one can argue whether or not this signal and its meaning is a communicative innovation, the more interesting communicative behaviors emerge when the shape of the sender's token has less rotational options than that of the receiver's token shape. For example, a circle cannot show rotations. A circle thus has fewer rotational options than a triangle, which can be oriented in four different configurations. The trials where senders play with circle tokens and receivers play with triangle tokens result clearly in communicative innovations.

One such communicative innovation is called the "wiggle" and this communicative behavior has been observed in many different studies (de Ruiter et al., 2010; Newman-Norlund et al., 2009; Stolk et al., 2013b). The wiggle is a communicative innovation generated by senders to indicate location and orientation of a receiver's token, when the sender's token has less rotational freedom than that of the receiver. For example, in trials where the sender has to use a circle (which cannot show rotations) to communicate the orientation of a triangle (which has four orientations), often—but not always—senders adopt a wiggle signal. Figure A1b displays an example of the information given to a sender on such a trial. The wiggle signal consists of the communicator pausing her token at the receiver's target location to convey that his token should be positioned there. Then, unable to orient her own circular token, the sender uses repetitive movements along an axis to signal the orientation of the receiver's triangle (see Figure A3). This signal, however, can mean various things within the context of the game. For instance, it can mean the pointing direction of a triangle ("wiggle apex"), or the number of times a receiver needs to perform a "rotate" action ("wiggle rotate"), or even the opposite of the pointing direction depending on the number of repetitions ("wiggle opposite"; see Table A1).

In order for these communicative innovations to be generated or understood candidate meanings have to be generated *de novo*. The possible meanings of these innovations are not predefined, and they are also open-ended. These properties are best observed with the wiggle. Therefore, we will use the wiggle as a case study for Analogical Abduction Proper in the next section.

A.2. AN EMPIRICAL TEST CASE: GENERATING THE MEANING OF A "WIGGLE"

As seen in the previous section, the wiggle signal can be interpreted in different ways. We first sketch informally how the "wiggle apex," "wiggle opposite," and "wiggle rotate" meanings can be hypothesized by analogical abduction proper. Then we present a more detailed and formal analysis of the wiggle apex meaning. It is important to note that we will limit our case study to a single augmentation path (the gray annotation in Figure 1 in the main text) and single candidate hypothesis (the gray dashed line in Figure 2 in the main text). The key point here is not to show an entire hypothesis space (as this would require too many pages), but to show that Analogical Abduction Proper can in principle generate novel candidate hypotheses. It will become clear that, given a large knowledge base, it can generate a set of candidate hypotheses.

Informal wiggle analysis: Tables A2, A3 and A4 illustrate possible augmentations that can be performed on representations of the signal and cores such that a candidate meaning can be hypothesized by finding a match (and projecting structure) between the augmented representations. Generating one possible meaning hypothesis starts with two representations, one of the signal and one of a possible core. Each consecutive row is an augmentation of the previous representation with some knowledge via Analogical Augmentation (depicted by ⇝). On the final row (match level) an analogical match between the reconceptualized evidence and reconceptualized core is found. This final match, including a potential projection and instantion, is a candidate hypothesis about the meaning of the signal.

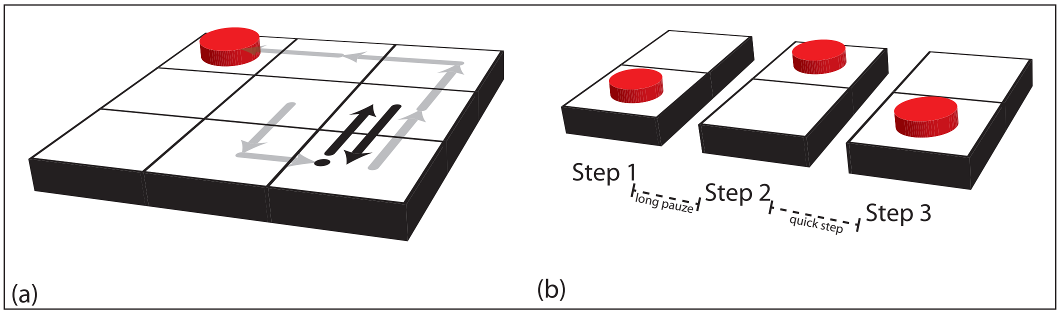


Figure A3.
Zooming in on the wiggle signal. The two locations in the time steps in (b) are parts of the (a) bigger 3 × 3 board.

Table A2.
Wiggle apex: augmentation and analogical match sketch.

		Representations of evidence *e*			Representations of core *k*
Lowest level	(1)	timed sequence of circle locations		(1)	equilateral triangle
					augment with symmetry ⇝
				(2)	equilateral triangle with axis of symmetry
					augment with apex⇝
				(3)	triangle with axis and apex
		augment with lines			*augment with* alignment ⇝
	(2)	path		(4)	triangle with apex, aligned to frame of reference by axis
		augment with alignment			*augment with* base ⇝
	(3)	path aligned to frame of reference		(5)	triangle aligned to frame of reference by axis, with apex and base
		augment with start location			*augment with* location ⇝
	(4)	path aligned with start		(6)	triangle aligned to frame of reference by axis, with apex and location
		augment with direction			*augment with* direction ⇝
	(5)	vector with starting location		(7)	pointing triangle with location *augment with* orientation ⇝
		augment with orientation			
Match level	(6)	vector with start and orientation	⇔	(8)	triangle with location and orientation

From these sketches we can already make some interesting observations. The first observation is that the augmentations that are required to reconceptualize representations to find a candidate hypothesis are not trivial. The second observation is that the same core representation can lead to different candidate hypotheses, e.g., the "wiggle apex" and "wiggle opposite" candidate meanings start with the same core representation. The third observation is that multiple different core representations can form the basis of different candidate meanings, e.g., the "wiggle apex" and "wiggle rotate" are based on two different core representations. Finally, each candidate hypothesis is grounded in perception, because it is an analogy between a reconceptualized representation of observed evidence and knowledge.

Formal wiggle apex analysis: We now present a formal analysis of the "wiggle apex" strategy to illustrate Analogical Abduction Proper. For each candidate hypothesis, Analogical Abduction Proper consists of two parts: perform deep analogical inference on the evidence and core representations, then find possible analogical inferences between the resulting representations. In this case study we first show a representation that results from deep analogical inference of the observed evidence, i.e., the actual movements of the sender token on the board. Second, we assume that a similar process has been done for the core concept "triangle" and illustrate an analogical inference between the two.

Note that we use specific representations in the formal analysis. This, however, does not mean that we commit to these representations being "true." In fact, we would argue that many different representations are psychologically plausible. This is accommodated for in Analogical Abduction proper, because it is agnostic about the content of the representations. Furthermore, the main point here is to show that knowledge that is inherently not about the TCG can be part of the deep analogical inference to generate a candidate hypothesis that can explain the observed wiggle apex signal.

Representations of evidence and core: We start by introducing the representation of the signal. This representation (see Figure A4) involves a number of objects, attributes, functions, and relations. We list these and their interpretation below. Because these representations are quite large, we will use a graphical notation for readability. A relation is depicted by its label and two or more arrows pointing to its arguments. An attribute or function only has one argument, and objects are leaves.

Objects

1. `Board`: represents the TCG board.
2. `(3,3)` and `(3,2)`: represent two locations on the TCG board.
3. `FoR`: "Frame of Reference" represents an abstract spatial/geometrical frame.

Table A3.
Wiggle opposite: augmentation and analogical match sketch.

	Representations of evidence *e*		Representations of core *k*
Lowest level	(1) timed sequence of circle locations *augment with* lines ⇝		(1) equilateral triangle
	(2) path *augment with* alignment		*augment with* symmetry ⇝
	(3) path aligned to frame of reference *augment with* start location		(2) equilateral triangle with axis of symmetry *augment with* apex ⇝
	(4) path aligned with start *augment with* direction ⇝		(3) triangle with axis and apex *augment with* alignment ⇝
	(5) vector with starting location *augment with* orientation		(4) triangle with apex, aligned to frame of reference by axis *augment with* base ⇝
	(6) vector with start and orientation *augment with* symmetry		(5) triangle aligned to frame of reference by axis, with apex and base *augment with* location ⇝
	(7) vector with start and orientation and a symmetrical path *augment with* counting		(6) triangle aligned to frame of reference by axis, with apex and location *augment with* direction ⇝
	(8) vector with start and orientation and a number *augment with* odd/even reverse ⇝		(7) pointing triangle with location *augment with* orientation ⇝
Match level	(9) vector with start and orientation possibly reversed	⇔	(8) triangle with location and orientation

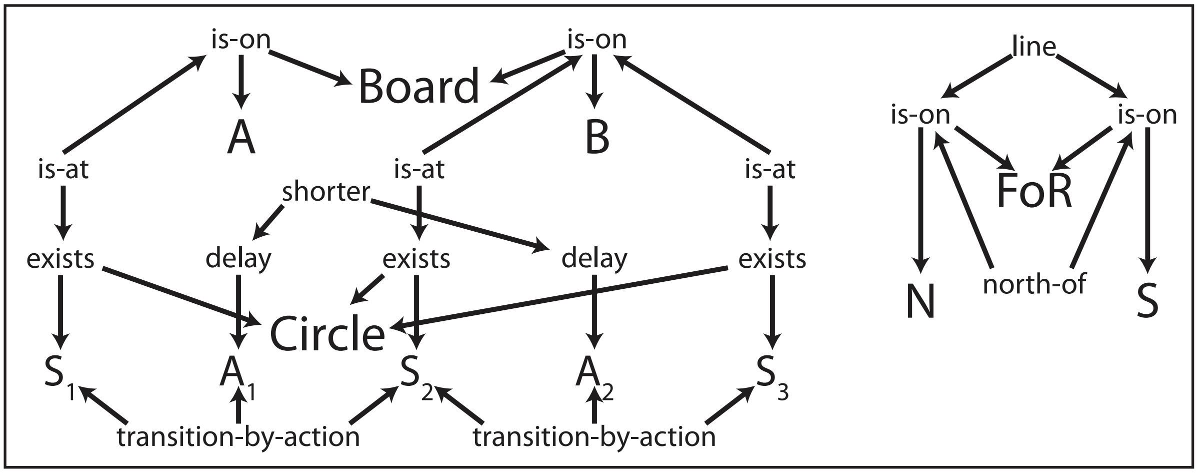


Figure A4.
Relational representation of the evidence. Graphical representation of the representation of the observed sender signal.

4. `N` and `S`: "North" and "South" part of the frame.
5. *S*1,… and *A*1,…: States and actions.
6. `Circle`: The circle token.

Attributes

1. `delay(a)`: there is a delay of `a` milliseconds caused by the action.

Table A4.
Wiggle rotate: augmentation and analogical match sketch.

		Representations of evidence *e*			**Representations of core *k***
Lowest level	(1)	timed sequence of circle locations		(1)	object controlled with directional pad and rotate object button *augment with* 2D space ⇝
		augment with lines ⇝			
	(2)	path *augment with* location ⇝			*augment with* lines ⇝
	(3)	path to location		(2)	controller for moving object in 2D space and rotate object button
		augment with symmetry ⇝			*augment with* lines ⇝
	(4)	path to location then symmetrical path		(3)	controller for moving object along path and rotate object button
		augment with counting ⇝			*augment with* counting
Match level	(5)	path to location and number	⇔	(4)	controller for moving object along path and a button to rotate object a number of times

Relations

1. `is-on (a,b)`: location `a` is on the board `b`.
2. `north-of (a,b)`: location `a` is spatially north of location `b`.
3. `shorter (a,b)`: time `a` is shorter than time `b`.
4. `exists (a,b)`: object `a` exists in state `b`.
5. `is-at (a,b)`: the object that exists in this state `a` is at location `b` on the board.
6. `transition-by-action (a,b,c)`: action `b` transitions state `a` into state `c`.

In Figure A4 we can see that the communicator's behavior is represented with two board locations `(3,2)` and `(3,3)` that lie on a board. These locations are both communicative and there is a line between them. This line has an orientation and is directed. Its starting point is `(3,3)` and the orientation is represented by the fact that it aligns with a north–south frame of reference, directed north.

Augmenting the evidence representation: In this section we show how a process of deep analogical inference can create a representation that can lead to a candidate hypothesis. We illustrate how this process works for the perceptual representation of the communicator's signal. The same process can build a novel representation of the core.

The perceptual representation of the signal *e* that forms the target of the example augmentation only includes the communicator's movement over the board (using simple spatial and temporal relations). This representation is based on a discretized concept of time, i.e., there is a sequence of states and actions. In each state there exists a circle at a location on the board. Between these states there is a certain time-delay: if there is a delay of 0.5 seconds between *S*1 and *S*2, then the world is in state *S*1 for 0.5 seconds and then transitions to state *S*2. Additionally, there is a representation of the board and of a frame of reference. For readability we limit the representation to the parts involved in the analogy. Figure A5 again shows the evidence representation in black. It also shows (in different colors) each individual augmentation that is needed to generate the structure that will be part of the candidate hypothesis.

Each augmentation is performed by first matching to a representation of knowledge $k \in K$ and then projecting the biggest possible structure over from that knowledge representation *k*, increasing the richness of the signal representation *e*. Figure A6 contains all the basic knowledge representations used in the deep analogical inference with the following sequence of augmentations: (1) Communicative pause, (2) Communicative, (3) Lines, (4) Align line, (5) Location, (6) Starting point, (7) Direct line, and (8) Orient. The basic knowledge representations do not contain any knowledge specifically about the TCG. They do make use of several new objects and relations. We list these below and afterwards give an intuitive interpretation of the knowledge representations.

Objects

1. `Communicative`: represents an abstract conceptualization of communicative aspects of the behavior (e.g., this can represent segmentation information).

Relations

1 `is-comm (a,b)`: concept `a` is communicative.
2 `Line (a,b)`: there exists a line between location `a` and `b`. Note that this relation is not ordered, i.e., `line (a,b)` ≡ `line(b,a)`.

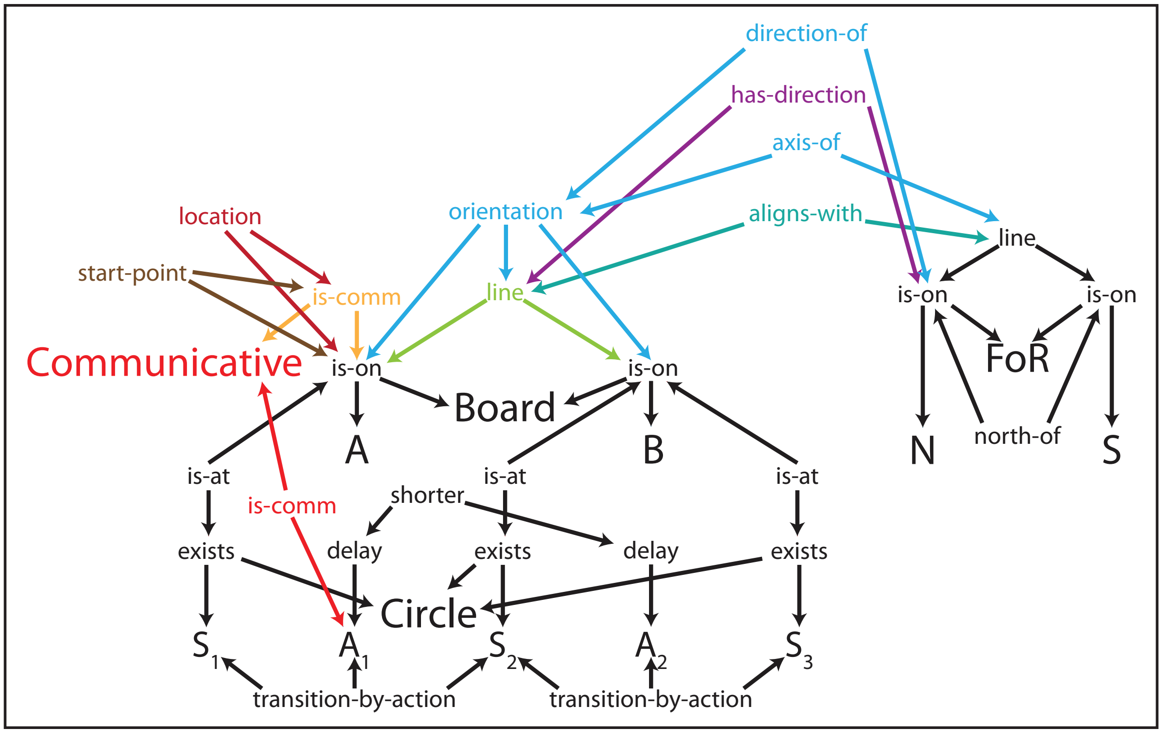


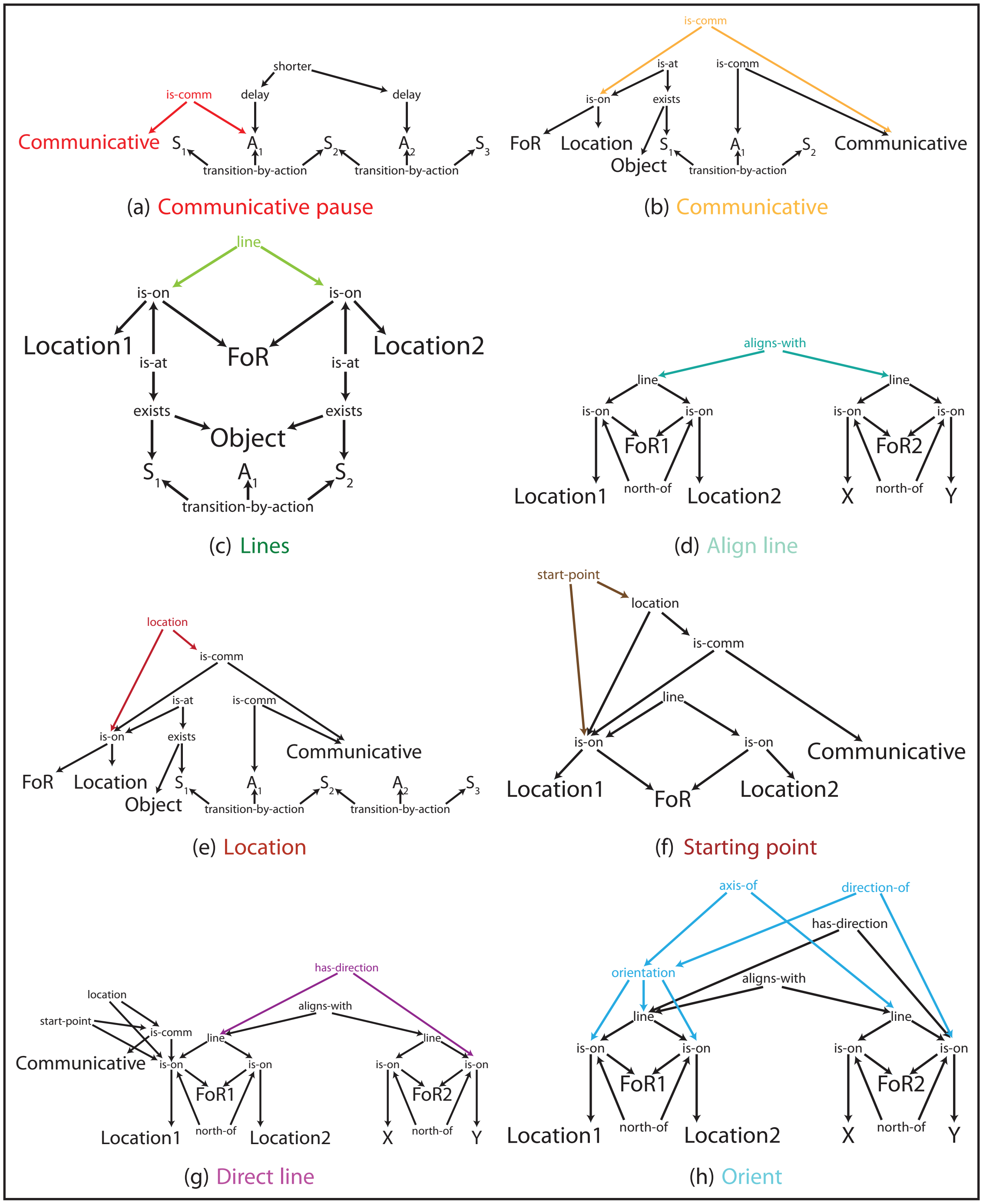
Figure A5.
The final augmented representation of a wiggle. Each color shows a different analogical augmentation.

3 `start-point(a,b)`: location `a` is a starting point with property `b`.
4 `location(a,b)`: object `a` is a location with property `b`.
5 `has-direction(a,b)`: line `a` is directed to part of the frame `b`.
6 `aligns-with (a,b)`: line `a` aligns with axis of the frame `b`.
7 `orientation(a,b,c)`: two locations `a` and `c` that are connected by line `b` have an orientation.
8 `axis-of (a,b)`: axis `b` of the frame applies to orientation `a`.
9 `direction-of (a,b)`: orientation `a` is directed to part of the frame `b`.

Intuitive interpretations

1 Communicative pause: "If a state/action takes more time compared to the state/action that comes after it, this state/action is communicative."
2 Communicative: "If a state/action is communicative, then the location the object is in at that state is communicative."
3 Lines: "If an object is first in location 1 and then in location 2, then one can think of a line being between those locations."
4 Align line: "If there are two lines between two pairs of locations that have a similar relationship (e.g., north-of) then these two lines align."
5 Location: "If an object exists in a communicative location and stays there longer than in its next location, then that location is hypothesized to be *the* location."
6 Starting point: "If two locations on a line are communicative and one of them communicates the location, then that location is the start point of the movement along that line."
7 Direct line: "If a line with a start point aligns with a line in a different frame of reference, then that line has a direction towards the second location of that other frame of reference."
8 Orient: "If a line has a particular direction, then it can be thought of as having an orientation. Here, orientation is a direction along a particular axis with respect to a different frame of reference."

To start deep analogical inference we first augment the evidence representation (black representation in Figure A5) with Communicative pause. This involves, first, finding an analogical match between Communicative pause and the perceptual representation, and then projecting over the biggest possible structure. In this case, the relation is-comm and object Communicative are projected over. The other knowledge representations keep augmenting the representation in a similar fashion, enriching the representation of the communicator's signal. Eventually, the sequence of augmentations leads to the representation that was used previously in this section to generate a candidate hypothesis about the meaning of the communicator's signal.



(a) Communicative pause

(b) Communicative

(c) Lines

(d) Align line

(e) Location

(f) Starting point

(g) Direct line

(h) Orient

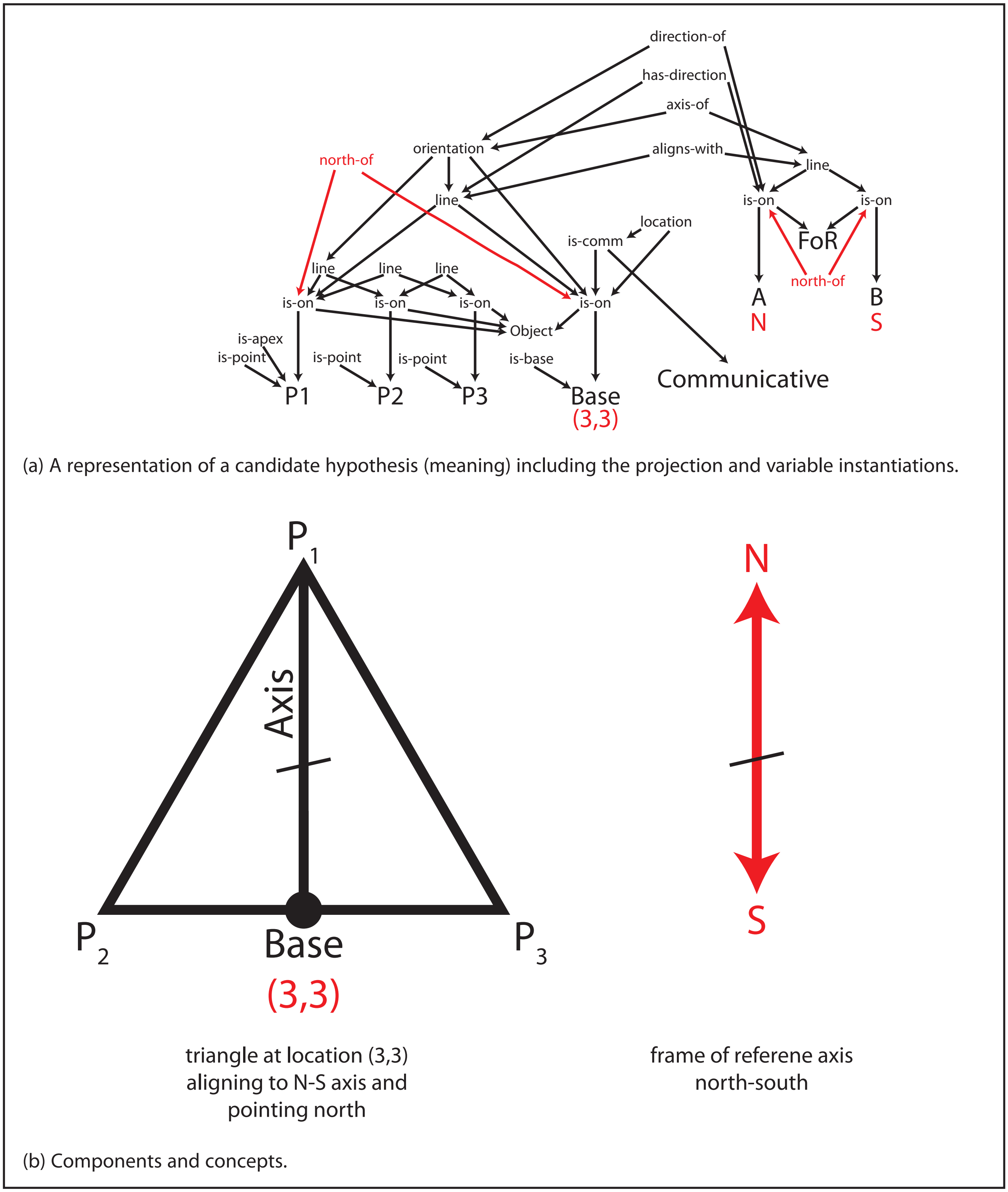
Figure A6.
The knowledge representations that are used for augmenting the wiggle representation.

Candidate hypothesis for "wiggle apex." A candidate hypothesis is based on an analogical match between an augmented representation of evidence and an augmented representation of a core. Next we introduce the representation of the core which represents the concept of a triangle that points (see black representation in Figure A7). It is presupposed that this representation is the result of deep analogical inference and it includes new attributes:

Attributes

1. `is-point(a)`: object `a` is a point.
2. `is-apex(a)`: object `a` is an apex.
3. `is-base(a)`: object `a` is a base.

The triangle representation consists of three points (P1 (the apex), P2, and P3) and a base (Base). These objects are all located on an object, i.e., the abstract triangle. There are



(a) A representation of a candidate hypothesis (meaning) including the projection and variable instantiations.

(b) Components and concepts.

Figure A7.
Candidate hypothesis. Parts annotated in red are structures that have been projected or variable-instantiated. They represent the location of the triangle (3,3) and its orientation ("pointing north").

three lines, the sides of the triangle, and an additional line from the base to the apex representing the axis of symmetry. The base and the apex are communicative, because additional information about these triangle properties is what gives the triangle its location (conceptualized as location of the base), orientation, and direction (conceptualized as an alignment of the axis of symmetry with an axis of the frame of reference).

The evidence and core representations presented here analogically match. This match corresponds to the cognizer understanding that the behavior and candidate meaning are analogous. This, however, is not enough to explain how the signal is hypothesized to have a specific meaning, i.e., we know that the "the repetitive movement along an axis is analogous to the pointing of the apex" but we do not know the specific location and orientation of the triangle. To hypothesize these specifics we need to project relational structures onto and instantiate variables in the core representation from the representation of the signal, based on the analogical match.

In our example we only show one possible projection and variable instantiation, i.e., the red structures in Figure A7a. We project the spatial relation between the base and apex from the base representation of sender behavior to the target representation of triangle. In addition, we instantiate variables: `A→N`, `B→S`, and `Base→(3,3)`. These inferences make the candidate meaning more specific, because they contain information that the triangle apex should be north of the base and that the axis should align to the north–south axis of the frame of reference.

The candidate hypothesis we illustrated in this section is only possible by virtue of a process of deep analogical inference and the knowledge used. Given different representations, analogical abduction proper would generate completely different candidate hypotheses. This illustrates how analogical abduction proper can in principle generate sets of candidate hypotheses. As argued in the main text, this process is isotropic, open-ended, novel, grounded, sensible, and psychologically realistic.

NOTES

[1] Rotation is immediate (not animated), therefore a circle does not appear to rotate and rectangles appear to be in the same orientation after two rotations. Correctness of orientation is evaluated relative to the appearance of the orientation, not the number of times the player has rotated.