Contextual abnormality for teleological explanation

Alexandra Varga (Alexandra.Varga@psychol.uni-giessen.de)
Experimental Psychology and Cognitive Science, Justus-Liebig Universität
Giessen, Germany

Abstract

How can we make sense of observed instrumental actions that are on a first glance bizarre, i.e., different from what "I myself would have done"? In an attempt to answer this question, the paper sets forth a two-staged reasoning procedure for teleological action explanation: goal assignment, and backward planning. Closed-world assumptions about abnormalities frame reasoning to a manageable format under limited processing capacities. Non-default instrumental actions may be explained with respect to a goal hypothesis by encountering an abnormality in the action context. The proposed procedure can be modelled in logic programming, and thereby subserve empirical research on the more generic topic of of defeasible reasoning.

Keywords: teleological reasoning; action explanation; closed-world assumption; abnormality.

Introduction and road map

I propose a reasoning procedure that fosters the explanation of intentional instrumental actions, i.e., actions meant by an agent to achieve a particular goal in the context of performance. My grounding assumption is that the two mirroring phenomena of planning one's own performance of instrumental actions, and understanding those performed by other agents, are underpinned by inferential processes. Planning and explaining actions are instances of high-level cognition. The assumption is supported by empirical evidence, e.g., in the developmental literature (Gergely, Bekkering, & Király, 2002; Király, Csibra, & Gergely, 2013), as well as in adult studies (Brass, Schmitt, Spengler, & Gergely, 2007; Hickok, 2009). My locus of concern is the use of teleological reasoning for interpreting goal-directed actions. The distinctiveness of the proposed mechanism resides in the fact that it is useful for real agents in real time to make sense of atypical instrumental actions, i.e., of alternatives to the default ways of achieving goals.

It has been shown recently (Varga, 2013) that the use of teleological reasoning for imitative learning from observations by human agents as young as 14-month-old, i.e., the results of Gergely et al. (2002), can be modeled in the non-monotonic formal system of constraint logic programming (Lambalgen & Hamm, 2005). The account of teleological reasoning set forth in this paper lays the ground for computational models of human explanatory practices.

I start with introducing some distinctive features of goal-centered reasoning in the service of teleological explanation of actions, which reveal its considerable computational complexity. I go on to present the proposal for a realistic procedure by which human agents with limited cognitive capacities may succeed, i.e., reasoning with closed-world assumptions. In the next Section I describe the reasoning steps. I then briefly present constraint logic programming, focusing

on the features that recommend it for modelling teleological reasoning for action explanation. Upon wrapping up, I emphasize the potential of closed-world reasoning about abnormalities to provide a conducive conceptual framework in cognitive science, and end with a related methodological upshot.

Teleological reasoning about actions

Goals are a peculiar kind of action effects that motivate agents to plan their actions. Because reasons for action are grounded in agents' prior motivations, goals fulfill an explanatory role. Inasmuch as goals are motivational factors for action performance, they also focus the explanatory processing of other agents' actions. If a goal *g* gives agent *X* reasons for doing *a*, then *g* explains *X*'s *a*-ing in context *c*. This means that an observer agent *Y* may use the goal in order to make sense of *X*'s action performance. The human propensity for teleological explanation is empirically well-documented in the psychological literature (Csibra & Gergely, 2007; Lombrozo & Carey, 2006).

Teleological reasoning is intrinsically linked with the features of the situation where it is applied. The goal status is hypothetical because goal inference is an inverse problem (Csibra & Gergely, 2007): contextual information does not deductively recommend a single solution. The choice of means fit for goal achievement is also guided by contextual features. For example, depending on whether the road is frosty or not, I may choose to cycle or to walk to university on a Monday morning. Relatedly, action explanation with respect to the agent's reason for action is context dependent too. I might interpret my colleague, a convinced cyclist, walking to university in a different manner on a warm day of spring than on a freezing cold day.

Agents' teleological inferences are flexible. The context-relative hierarchy of means and ends makes teleological reasoning an epitome of hypothetical defeasible reasoning; it is a form of 'explanationist abduction' (Gabbay & Woods, 2005). Its conclusions are open to revision as new information becomes available. For this and other reasons (see Pollock, 1995; Kowalski, 2011; McCarthy & Hayes, 1969), goal-centered reasoning is remarkably complex from a computational point of view.

Closed-world assumptions in teleological reasoning. Taking into account cognitive economy presumptions (Chater & Vitanyi, 2003; Gigerenzer, Todd, & ABC group, 1999), according to which reasoners tend to invest a minimal cognitive effort for a maximally advantageous outcome, the computational complexity may be dealt with in real time by as-

sumptions that help to 'frame' (McCarthy & Hayes, 1969) the inferential scope. One such assumption is that unless positive information is available, no abnormality occurs in the context[1]. This is the closed-world assumption for reasoning about abnormalities CWA_{ab} (Stenning & Lambalgen, 2008). It calls for the formation of the minimal, simplest possible interpretation of the context to be explained.

Attached to the CWA_{ab} is a disclaimer based on the reasoners' commonsense background knowledge, i.e., a set of conditions specifying the kind of positive evidence that may constitute abnormality. At the psychological level, the set of abnormality conditions can be said to be 'at the back of the reasoner's mind'. The set is the result of interaction between working memory and long-term memory, under limitations imposed by the capacity of the former. The whys and wherefores of such relations are depicted in detail in J. R. Anderson's (1983) prominent model of human thinking, ACT-R.

The set of abnormality conditions is $\Sigma = \{p_1, \ldots, p_n\}$, where IF p_1 THEN $ab, \ldots,$ IF p_n THEN ab. The states of the world represented by $p_1, \ldots, p_n$ are explicit evidence of abnormality. $p_1 \ldots p_n$ are hierarchically ordered based on the likelihood of overriding the assumption. n is potentially infinite, because so is the cardinality of the set of logical possibilities; in principle, there could be gremlins, fairies, the laws of gravity may stop holding, etc. Some of these abnormal cases however are more realistic than gremlins. They are exceptional contextual features that may actually prevent the smooth, habitual running of a process. The set of conditions considered depends on various factors, such as the importance of a satisfactory explanation, the degree of certainty that counts as satisfactory, or the amount of time available for computations.

In contexts where there is evidence of any of $p_1 \ldots p_n$, the assumption is *justifiably* overridden. For instance, a broken right elbow could define an abnormal context for a right-handed person. It is crucial for the current purposes to note that her using the left hand to sign a petition would be explained by observation of p_1 = 'broken hand' in a context where the rule 'IF p_1 THEN ab' is active. It goes without saying that abnormality conditions are checked by reasoners only if teleological inferences produce unsatisfactory outcomes under the assumption that nothing abnormal is the case. Otherwise the CWA_{ab} would be self-defeating.

Under this view of teleological closed-world reasoning, human agents' instrumental actions can be represented in action rules that connect a goal to an ordered sequence of actions in a particular context. They read as "In order to g in context c do a[2], unless something abnormal is the case". The CWA_{ab} is captured by the 'unless' proviso. Under the assumption that nothing abnormal is the case, a is a default action. The proviso makes explicit that the rule may not be applicable in a different context where exceptions do occur. Exceptions to

action rules are features of a context, say c_1, which make abnormality conditions obtain, and thereby override the application of the closed-world assumption to c_1. The action rule may be revised for c_1, and prescribe performance of an action a_1, $a_1 \neq a$, in order to g. For the corresponding case of action explanation, positive evidence of an exception p_1 in context c_1 would justify performance of a_1 instead of a in order to g.

The reasoning steps

Let us begin with a simple example of two agents, an actor and an observer, in an action context. It will serve to illustrate the reasoning steps.

Imagine a Dutch university, where the bike is the default means of transport to work. X works there as a researcher. One day her colleague Y sees her walking towards the University not long before the regular arrival time. How can Y make sense of X walking? The action is *prima facie* not understood, and it calls for explanation.

The reasoning involved in teleological explanation of actions is roughly two-staged: formation of a goal conjecture, followed by testing its explanatory capacity in the current context.

(1) Goal hypothesis formation. In the first step, the observer Y conjectures a goal of the instrumental action a performed by X.

Given that goals are conceptualized as a kind of action effects, observations are first causally individuated. Organizing observations along means – ends hierarchies supervenes on setting up a causal model (Lambalgen & Hamm, 2005) of the current context. Empirical research supports this proposal; the causal individuation of events appears to be a quasi-automatic processing mode. From very early infancy events are perceived as forming cause – effect sequences. This is evidenced by empirical findings, e.g., $6 - 7$ month-old infants dishabituate to causal sequences of motion events after having been familiarized with situations in which spatio-temporal contiguity (presumably the crucial cue for causal relations) between events is disrupted (Saxe & Carey, 2006).

In our working example, X's walking is perceived as the cause of gradual minimization of the distance to university, which eventually results in arrival at the office[3] – a = 'walking'.

The causal model of the current situation, relevant bits of background knowledge, and other observable cues for goal attribution, such as:

· the number of effects per cause (potential multifinality) and of causes per effect (potential equifinality),

· the availability and salience of action effects,

· the agent's (emotional) reaction to the effects,

constitute the database for further computations. The computational complexity of goal hypothesis formation depends

[1] 'Positive information of abnormality' may refer to current observation, or consequences derived from it by forward reasoning.

[2] For simplicity I will henceforth use the singular 'action' for 'ordered sequence of actions'.

[3] This is an instance of perceiving continuous causation (Lambalgen & Hamm, 2005).

on which of these elements are available, and on their consistency (or lack of). Given the database, Y establishes a goal hypothesis g that presumably calls for the observed agent's action a in context c. In the current example, Y's knowledge that it is a working day, that the path on which he sees X walking leads to university, and that X works for the university, plausibly leads to assigning X's action a the goal to come to university.

Clearly goal assignment is often much more complicated that this, e.g., in contexts of unsuccessful or unfulfilled actions, multifinality, equifinality (Baker, Tenenbaum, & Saxe, 2009; Csibra & Gergely, 2007; Luo & Baillargeon, 2005; Paulus & Király, 2013), but a comprehensive mechanism for goal inference is beyond the scope of this paper. This example is sufficient to grasp the structure of the reasoning process.

(2) Testing the explanatory potential of the goal hypothesis. Because goal assignment is hypothetical, the explanatory function of the goal requires confirmation. The uncertainty with respect to fulfilling the function is even higher in the case of unusual or atypical actions. And it is precisely such actions that usually trigger explanatory processes.

I propose offline plan simulation as a method for hypothesis testing. The procedure is offline because the result of planning is not overt action, rather an action representation to be compared with observations.

Reasoning amounts to computing a sequence of actions for the goal g, in an attempt to answer the question "what would I have done in order to g?". This gives voice to the widespread human tendency to use own behavior as a standard for understanding the actions of other agents, when the observer can resonate either with the behavior itself, or with its conjectured goal. In this sense, it follows the guiding idea behind simulationist approaches to action understanding (Gallese & Goldman, 1998; Zentgraf, Munzert, Bischoff, & Newman-Norlund, 2011). The crucial difference lies in the fact that the proposed account is inference-driven.

The inferential strategy for planning is backward closed-world reasoning (Lambalgen & Hamm, 2005), from the goal g to actions[4]. The input for reasoning is the goal, and the expected output is a temporally ordered sequence of actions whose performance achieves the goal, unless something abnormal is the case in the context.

The output is represented in the format of action rules introduced above. The action in the rule is a default action (Mueller, 2006), i.e., what the observer would have done in order to g in a context where nothing abnormal is the case. Defaults are typical actions[5]. The action rule computed by the observer in offline planning is "In order to g do b unless

---

[4]The backward direction of reasoning overlaps with the goal-to-action inferences in teleological action interpretation, as described by Csibra and Gergely (2007). In a discussion of deontic conditional reasoning, Beller (2008) applies it to inferences from the action side to the condition side of deontic rules.

[5]It is worth noting that although there may of course be individually specific defaults (e.g., my own special way of typically making coffee), most are shared across communities.

something abnormal is the case". The 'unless' proviso allows that additional contextual information modify the action b prescribed by backward reasoning for goal achievement. A non-default action may be prescribed for the same g in an abnormal context.

Planning with CWA_{ab} is a rather automatic reasoning procedure, e.g., it can be implemented in a spreading activation network (Stenning & Lambalgen, 2008). A passive, unsupervised process like spreading activation is essential granted the size of potentially relevant long-term memory. In the case of habitual goals that pertain to the reasoner's procedural knowledge (e.g., get to work, write a paper, make coffee), offline planning details activation of an action response most strongly associated with the goal (e.g., mount on the bike, turn on the computer, turn on the stove). The use of the CWA_{ab} for planning is the crucial element that justifies calling this process inferential. Low-level action – effect association processes do not allow the kind of flexibility, and thus contextual adjustment of actions to goals, that closed-world planning does.

The observer's own action rule is compared with observations; the output is either match or mismatch.

The case of match, i.e., $a = b$, is rather trivial. The fact that the observed agent did what the observer would have done to attain the hypothesized goal confirms the hypothesis and its explanatory function. In fact, the word 'explanation' may appear as a misnomer here, since the teleological structure of observations is self-evident. Nothing calls for what we normally mean by explanation – a deliberative, consciously engaged process. However inferential and automatic processing need not be seen as contrasting modes; as mentioned above, closed-world reasoning instantiates both (Stenning & Lambalgen, 2008).

The more interesting and explanatorily substantial case is when observations conflict with the output of offline planning, i.e., $a \neq b$. At this point, the assigned goal does not fulfill the expected explanatory function. Computations proceed stepwise. The goal hypothesis is not canceled immediately. Cognitive economy, or the least effort principle, recommends a computationally less expensive conflict resolution procedure first, i.e., retry explanation in light of the same goal.

The working example instantiates this case. Let us spell out the two mismatching action rules.

Default: In order to come to university use the bike in context c unless something abnormal is the case.

Observation: In order to come to university walk in context c_1 unless something abnormal is the case.

In Y's simulation of X's plan for the goal "come to university", the CWA_{ab} prescribes the default use of the bike in a normal context c. Further contextual analysis of the observed c_1 may provide evidence for abnormality. This calls for checking whether CWA_{ab} does indeed hold in c_1. Sup-

pose that X's default rule comes with the following hierarchy of abnormality conditions:

1. IF wind is too strong THEN *ab*,

2. IF there is snow or frost on the way THEN *ab*,

3. IF bike has flat tire THEN *ab*,

4. ...

They are scanned from the most likely in descending order. Suppose that the whole story takes place in a warm sunny day. It is thus easy to reject (1) and (2) – the conditions' truth value is 0. Then X encounters (3). He has no positive information about the state of his colleague's tires. However, he does not know that it is false either. The truth state of condition (3) is u (uncertain), but it may evolve towards either 0 or 1[6]. At this point the CWA_{ab} *may be* justifiably overridden, and thereby the unusual instrumental action may be explained by the assigned goal. Suppose that Y finally enters the office and starts complaining about the poor quality of bike tires nowadays. This indicates that the antecedent of condition (3) holds (its truth value is 1), thus something abnormal is the case, thus c_1 where the bizarre action takes place is an abnormal context with respect to the assigned goal. X's walking to university on a warm sunny day is then contextually explained by g in light of the abnormality for which positive information about the flat tire provides evidence. The minimal teleological interpretation is extended to include the abnormality.

If the scan of abnormality conditions provides no evidence of relevant exceptions to the default action rule for the goal g, the goal conjecture is dropped. The teleological structure of the context of observation is recomputed from step (1). A different goal g_1 is assigned to the observed action, and its explanatory function is verified along the lines of (2). The process is thus recursive.

Potential for formal implementation

The logical notion that corresponds to psychological flexibility is non-monotonicity (Mueller, 2006; Stenning & Lambalgen, 2008). It means that validly derived conclusions may become false when new premises are added to the database. A non-monotonic formal system is needed as a computational format for the flexible, context adaptive reasoning that I proposed to subserve action explanation. Throughout the paper I described these processes as a form of closed-world teleological reasoning. The computational logic system of constraint logic programming (CLP) provides a suitable framework to capture goal-centered reasoning with closed-world assumptions. The technical background for the description of CLP below is taken mainly from Lambalgen and Hamm (2005).

Negation as failure (NAF) is the basic formal manifestation of closed-world reasoning; it can be encountered at the levels of the CLP's semantics, syntax, and consequence relation. An exhaustive description of CLP exceeds the present purpose.

In what follows I focus on the appropriateness of CLP semantics for modelling teleological reasoning. NAF is a weaker form of negation than the one in classical logic; the negation of a sentence is true whenever there is no evidence for the truth of the (positive) sentence. In Section *Closed-world assumptions in teleological reasoning* I introduced the CWA_{ab} by saying that it is applicable as a constraint on reasoning in the absence of *positive* evidence of exceptional cases that constitute abnormalities. Therefore formalization in terms of NAF is appropriate.

Furthermore it is worth noting that the notion of model construction[7] in formal semantics is tantamount to the psychological notion of interpretation. Action explanation proceeds via the construction of models that fit observations and relevant bits of background knowledge, taking seriously the constraints of cognitive economy. Formal model construction had thus better be uniformly and efficiently computable in real time. The weak notion of negation, NAF, fares well in this respect (Etzioni, Golden, & Weld, 1997). It allows computations to be performed in minimal models. Such minimal models are 'closed worlds'. Minimal model semantics is a useful modelling device because of its approximation of principles of least effort, as expressed in the CWA_{ab}. A minimal model may be (minimally) extended to cover abnormalities in the face of positive evidence.

Lambalgen and Hamm (2005) have argued for the use of a three-valued Kleene semantics for CLP as a modelling instrument for cognitive phenomena. Kleene semantics has three possible truth values: *1* for 'true', *0* for 'false', and *u* for 'uncertain'. What is special about the Kleene's *u* is that it is not a degree of truth intermediary between 0 and 1. Rather *u* is undecided and can evolve toward *0* or *1* as a computational upshot. This fits nicely with the potential indeterminacy of truth value of the conditions for abnormality with respect to action rules (such as was the case in the working example).

At the level of syntactic operations, CLP has the capacity to capture the use of abnormality conditions in teleological reasoning by representing them as integrity constraints (Kowalski & Sadri, 2009; Reiter, 1988). Integrity constraints impose local norms on computations, in the form of obligations or prohibitions. They are expressed in conditional form; they call upon, or prohibit, certain computational movements, ensuring that the database satisfies the conditions expressed in the consequent of the conditional. For the case of abnormality conditions, when evidence of exceptions becomes available, the database must be updated with abnormalities. This affects further computations, to which the closed-world assumption no longer applies.

The semantics of the conditional in integrity constraints is a matter of ongoing debate in the computer science and AI literature (Godfrey, Grant, Gryz, & Minker, 1998; Kowalski, 2011). For the current purpose of modelling teleological reasoning for explanation, I propose that the 'IF exception

---

[6]This is Kleene's three-valued semantics. More on this in the next Section.

[7]The term 'model' is used here as 'semantic model', not to be confounded with its homonymic meaning of 'formal theory'.

...THEN *ab*' expression of abnormality conditions should be monotonic. An abnormality that is validly inferred from positive evidence of an exceptional condition cannot be afterwards withdrawn; 'exceptions to exceptions' are not accepted by the formalism thus construed[8]. This possibility would be too permissive with respect to the flexibility of reasoning about actions, to the point that it could be detrimental to the desired efficiency of reasoning under real time constraints.

Finally, the cognitive relevance of CLP has been shown in a variety of domains. It has been used to construct a formal cognitive semantics of tensed speech (Lambalgen & Hamm, 2005) and thereby applied to discourse interpretation (R. Baggio, Lambalgen, & Hagoort, 2008), or to formalize conditional reasoning tasks which facilitated the derivation of predictions with respect to autistic subjects' performance on those tasks (Pijnacker et al., 2009; Pijnacker, Geurts, Lambalgen, Buitelaar, & Hagoort, 2010). For instance Pijnacker et al. (2009) have shown that different reasoning patterns between people with autism and normal controls are to be expected upon abstracting the logical form of the task using such logical formalization methods. The hypotheses generated by CLP formal models have been confirmed in behavioral and neural studies. Furthermore CLP has been conducive to appealing implementations in neural networks (Stenning & Lambalgen, 2008).

Conclusions: wrapping up and further on

I proposed that, upon establishing a causal model of actions in an observed context, agents explain actions via goal-centered inferences. First, a goal hypothesis is formed; in so doing, the agent constructs a minimal teleological model of observations. The expectation is that the goal explains the observed sequence of actions. Second, an attempt is made to corroborate the explanatory role of the hypothesis. This is done by computing a plan that answers the question 'what would I have done in order to achieve this goal?'. The offline planning is constrained by closed-world assumptions. When observations mismatch what the observer would have done, actions are not explained by the assigned goal. The flexible use of closed-world assumptions can foster explanatory computations without having to re-compute the teleological structure 'from scratch' (i.e., engage in recursion starting from goal attribution anew, and attempting to validate the secondary goal conjecture). Although different from what the observer would have normally done, an observed action may be explained by the initial goal assigned to it if the context of performance turns out to be abnormal. Consequently, the observer engages in further contextual analysis, by going through the abnormality conditions and checking for positive evidence that at least one of those obtains. The bottom line is that finding contextual abnormality supports efficient teleological explanation of non-default instrumental actions.

The proposed reasoning strategies can be formalized in CLP. Given that at all points I took into consideration cog-

nitive limitations, and that the chosen formalism is well able to capture the kind of processing required by these limitations, the proposal is likely to subserve the construction of a realistic process model for action explanation[9]. Furthermore such a model is likely to inform the intricacies of abductive reasoning at a descriptive level (Gabbay & Woods, 2005).

Non-monotonic reasoning has a larger scope of applicability in human cognition than teleological reasoning for explanation. The intrinsic role of modelling to provide theoretical generalizations does not need further arguments. Therefore a computational model of closed-world reasoning about abnormalities may also prove useful for empirical investigations of related cognitive phenomena in the psychology of reasoning, e.g., reasoning with counterexamples. In the subfield of legal reasoning, for example, research is currently underway (Gazzo-Castañeda & Knauff, n.d.) regarding the conditions in which exculpatory circumstances are accepted as counterexamples to legal rules. Conceptualizing counterexamples as abnormalities with respect to typical cases where rules apply, supplemented by a principled manner of establishing hierarchies of abnormality conditions, is likely to be beneficial by yielding finer-grained empirical predictions. The factors presumed to influence the cardinality of the set of abnormalities conditions (e.g., the importance of a satisfactory explanation, or the amount of time available for computations), for instance, may be manipulated in the experimental design.

Apart from its intrinsic interest, this proposal also has methodological implications. It is commonly assumed that logical modelling has been superseded by probabilistic techniques; more specifically, Bayesian models give the most prominent accounts of action understanding (Baker, Tenenbaum, & Saxe, 2006; Baker et al., 2009). There are close structural connections between logic programming and Bayesian models. However the latter are high-level normative accounts; their profile makes them unsuitable for process models of mental processes that involve fast-changing conceptual vocabularies (G. Baggio, Stenning, & Lambalgen, in press). The rejection of logics as modelling tools in psychology and cognitive science, in favor of probability, is premature.

References

Anderson, J. R. (1983). *The architecture of cognition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Baggio, G., Stenning, K., & Lambalgen, M. van. (in press). The Cognitive Interface. In M. Aloni & P. Dekker (Eds.),

---

[8]This is a form of regress argument.

[9]A further argument along these lines that is worth developing is that the functioning of production systems such as ACT-R closely resembles the procedural mechanisms of logic programming, which are most congenial to its practical use (Kowalski & Sadri, 2009).

Cambridge Handbook of Formal Semantics. Cambridge: Cambridge University Press.

Baggio, R., Lambalgen, M. van, & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language*, *59*, 36–53.

Baker, C. L., Tenenbaum, J., & Saxe, R. (2006). Bayesian models of human action understanding. In *Advances in neural information processing systems* (Vol. 18). MIT Press.

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

Beller, S. (2008). Deontic norms, deontic reasoning, and deontic conditionals. *Thinking & Reasoning*, *14*(4), 305–341.

Brass, M., Schmitt, R., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology*, *17*(24), 2117–2121.

Chater, N., & Vitanyi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78.

Etzioni, O., Golden, K., & Weld, D. (1997). Sound and efficient closed-world reasoning for planning. *Artificial Intelligence*, *89*(1–2), 113–148.

Gabbay, D. M., & Woods, J. (2005). *A practical logic of cognitive systems* (Vol. 2). Oxford, UK: Elsevier Ltd.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501.

Gazzo-Castañeda, E. L., & Knauff, M. (n.d.). Defeasibility and the role of counterexamples in reasoning with legal conditionals. *Submitted for publication*.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*, 755.

Gigerenzer, G., Todd, P. M., & ABC group the. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Godfrey, P., Grant, J., Gryz, J., & Minker, J. (1998). Integrity constraints: Semantics and applications. In J. Chomicki & G. Saake (Eds.), *Logics for Databases and Information Systems* (Vol. 436, pp. 265–306). Springer US.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229–1243.

Király, I., Csibra, G., & Gergely, G. (2013). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology*, *116*(2), 471–486.

Kowalski, R. (2011). *Computational logic and human thinking: How to be artificially intelligent*. New York: Cambridge University Press.

Kowalski, R., & Sadri, F. (2009). Integrating logic programming and production systems in abductive logic programming agents. In A. Polleres & T. Swift (Eds.), *Web reasoning and rule systems* (Vol. 5837, pp. 1–23). Springer Berlin Heidelberg.

Lambalgen, M. van, & Hamm, F. (2005). *The Proper Treatment of Events*. Malden: Blackwell Publishing.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*(2), 167–204.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal?: Psychological reasoning in 5-month-old infants. *Psychological Science*, *16*(8), 601–608.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence* (Vol. 4, pp. 463–502). Edinburgh University Press.

Mueller, E. (2006). *Commonsense Reasoning*. San Francisco, CA: Morgan Kaufmann Publishers.

Paulus, M., & Király, I. (2013). Early rationality in action perception and production? a theoretical exposition. *Journal of Experimental Child Psychology*, *116*(2), 407–414.

Pijnacker, J., Geurts, B., Lambalgen, M. van, Buitelaar, J., & Hagoort, P. (2010). Reasoning with exceptions: An event-related brain potentials study. *Journal of Cognitive Neuroscience*, *23*(2), 471–480.

Pijnacker, J., Geurts, B., Lambalgen, M. van, Kan, C., Buitelaar, J., & Hagoort, P. (2009). Defeasible reasoning in high-functioning adults with autism: Evidence for impaired exception handling. *Neuropsychologia*, *47*(644–651).

Pollock, J. (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person*. MA, USA: MIT Press Cambridge.

Reiter, R. (1988). On integrity constraints. In *Proceedings of the 2nd conference on Theoretical Aspects of Reasoning about Knowledge, TARK'88* (pp. 97–111). San Francisco, CA, USA: Morgan Kaufmann Publishers.

Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica*, *123*(1-2), 144–165.

Schueler, G. (2003). *Reasons and purposes: Human rationality and the teleological explanation of action*. New York, NY: Oxford University Press Inc.

Stenning, K., & Lambalgen, M. van. (2008). *Human Reasoning and Cognitive Science*. Bradford Book, The MIT Press, Cambridge, Massachussets.

Varga, A. (2013). *A formal model of infants' acquisition of practical knowledge from observation*. Doctoral dissertation, Department of Philosophy, Central European University, Budapest.

Zentgraf, K., Munzert, J., Bischoff, M., & Newman-Norlund, R. (2011). Simulation during observation of human actions – theories, empirical studies, applications. *Vision Research*, *51*, 827–835.