Formal Argumentation and Human Reasoning: The Case of Reinstatement

Mohammed Iqbal Madakkatel,[1] **Iyad Rahwan,**[1,2] **Jean-François Bonnefon,**[3]
Ruqiyabi N. Awan,[1] **Sherief Abdallah**[1,2]

[1]British University in Dubai, P.O.Box 502216, Dubai, UAE
[2]University of Edinburgh, Edinburgh, EH8 9LE, UK
[3]CNRS and Université de Toulouse, France

Abstract

Argumentation is now a very fertile area of research in Artificial Intelligence. Yet, most approaches to reasoning with arguments in AI are based on a *normative* perspective, relying on intuition as to what constitutes correct reasoning, sometimes aided by purpose-built hypothetical examples. For these models to be useful in agent-human argumentation, they can benefit from an alternative, *positivist* perspective that takes into account the empirical reality of human reasoning. To give a flavour of the kinds of lessons that this methodology can provide, we report on a psychological study exploring simple reinstatement in argumentation semantics. Empirical results show that while reinstatement is cognitively plausible in principle, it does not yield full recovery of the argument status, a notion not captured in Dung's classical model. This result suggests some possible avenues for research relevant to making formal models of argument more useful.

Introduction

Argumentation has become a very fertile area of research in Artificial Intelligence (Rahwan & Simari 2009). A highly influential framework for studying argumentation-based reasoning has been introduced by Dung (Dung 1995). An *argumentation framework* is simply a pair $AF = \langle \mathcal{A}, \rightharpoonup \rangle$ where $\mathcal{A}$ is a set of arguments and $\rightharpoonup \subseteq \mathcal{A} \times \mathcal{A}$ is a defeat relation between arguments. This approach abstracts away from the origin of individual arguments and their internal structures, and focuses instead on the defeat relationship between them.

Figure 1 shows an example textual argument and its corresponding graph structure. This structure is the canonical example for the notion of *reinstatement*. In particular, while argument A is defeated by argument B, the presence of C reinstates A since C undermines A's only defeater.

Given an argument framework (or graph), a semantics assigns a *status* to each argument. Classically, we distinguish between arguments that are *accepted* and those that are not (Dung 1995). Other approaches distinguish *accepted*, *rejected* and *undecided* arguments (Caminada 2006a).

In some cases, all semantics agree on the result. For example, in Figure 1, all classical argumentation semantics agree that we should accept C (for lack of any counter-argument), reject B (because there is a good reason to), and accept A (since every objection to it has been defeated). When there are cycles, different semantics may prescribe different results.

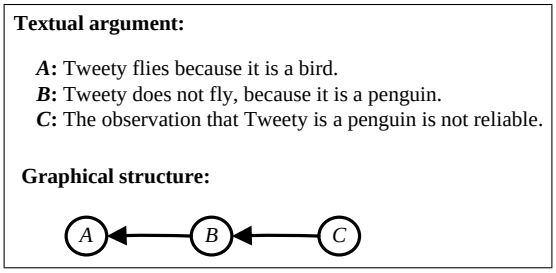

Figure 1: Defeat structure with reinstatement

Most models of argument evaluation in AI are normally based on a *normative* perspective, relying on intuition as to what constitutes correct reasoning, sometimes aided by purpose-built hypothetical examples. As we will argue below, there are limits to relying solely on this approach.

The main aim of this paper is to promote the use of psychological experiment as a methodological tool for informing and validating intuitions about argument-based reasoning. To give a flavour of the kinds of lessons that this methodology can provide, we report on an empirical study exploring simple reinstatement in Dung's argumentation semantics. The study reveals that while simple reinstatement is cognitively plausible, it does not fully restore the status of the reinstated argument. This notion is not captured in Dung's classical model of reinstatement, thus our result motivates work on new probabilistic semantics.

The study presented in this paper is not meant to constitute a comprehensive psychological assessment of Dung's semantics –this is beyond the scope of any single paper. But the results do show how psychological studies can provide new insights. Such results are relevant not only to the evaluation of existing semantics, but also to the design of new semantics. They are also relevant to the design of software agents that can argue persuasively with humans, or that can provide reliable support to humans in evaluating arguments (e.g. on top of argument diagramming tools). Before presenting our experimental results, we first review current methodologies for validating argumentation semantics.

Origins and Benchmarks of Semantics

The evaluation of semantics in the argumentation literature falls into two main categories: the example-based approach, and the principle-based approach, as discussed below.

The Example-based Approach

Most semantics for argumentation-based reasoning in AI are based on intuition as to what constitutes correct reasoning. This intuition is expressed formally in a precise manner, making it amenable to formal analysis. To test this intuition, a typical article presents a hypothetical scenario (e.g. on whether Tweety can fly) that correspond to a particular argument structure (e.g. argument defeat or reinstatement). Then, authors indicate the kinds of conclusions a system *ought* to draw, based on intuition pertaining to the particular example at hand. Subsequently, the semantics is shown to draw the desired conclusion.

However, often one can construct other examples, exhibiting the same logical structure, but in which the previous semantics leads to counter-intuitive results (e.g. Horty devoted a whole paper to demonstrate counter-intuitive results with floating conclusions in default reasoning (Horty 2002)). This motivates work on new semantic criteria to capture the new example, or class of examples. For example, CF2 semantics (Baroni, Giacomin, & Guida 2005) was conceived to deal with the problematic behaviour of preferred semantics in argument graphs with odd-length cycles. Semi-stable semantics (Caminada 2006b) was introduced mainly to deal with cases in which no stable extension exists.

The above approach to designing argumentation semantics has been recently termed the *example-based approach* (Baroni & Giacomin 2007). More importantly, examples act as the main tool for comparing one semantics with another. Baroni and Giacomin make a compelling case for the limitations of the example-based approach to semantics.

> *"example-based comparisons . . . are affected by the inherent limitation of relying more on intuition than on formally stated principles. In fact, even in relatively simple examples there may not be a general agreement on the "desired" outcome, due to different underlying intuitions"* (Baroni & Giacomin 2007)

This is consistent with Prakken, who argues that due to the wide-spread differences in underlying intuitions:

> *"it is better to use intuitions not as critical tests but as generators for further investigation"* (Prakken 2002)

Thus, it seems that there is a recognised difficulty in relying on intuition alone as the benchmark for designing and evaluating semantics for argumentation.

The Principle-based Approach

Motivated by the observations discussed above, a number of authors have recently advocated a more systematic, *axiomatic* (or *principle-based*) approach (Baroni & Giacomin 2007; Caminada & Amgoud 2007). In this approach, one analyses whether a semantics satisfies certain 'principles' or 'quality postulates.' Baroni and Giacomin present criteria such as the *reinstatement criterion*, according to which an argument must be included in any extension that reinstates it. Another example is the *directionality criterion* which requires that an argument's status should only be affected by the status of its defeaters. Such postulates can be used for systematic comparison between semantics.

Caminada recently provided postulates for the notion of reinstatement in order to characterise the labelling of arguments in an argument graph (in, out, and undecided) (Caminada 2006a). One postulate states that an argument must be 'in' if and only if all of its defeaters are 'out.' Another postulate states that an argument must be 'out' if and only if at least one of its defeaters is 'in.' This enabled characterising different semantics by the kinds of labellings they allow.

The principle-based approach provides a significant improvement, since it enables making statements that transcend individual examples and characterise semantics more generally. Having said that, the source of the general postulates themselves remains an intuition as to what correct reasoning ought to look like. In summary, most existing work on argumentation semantics, be it example-based or principle-based, is mostly *normative*. Research on argumentation semantics is rarely based on *empirical* evidence about how people actually reason with conflicting information.

Towards an Empirical Approach

There remains a largely unexplored source of intuition and validation for argumentation semantics, namely the psychology of human reasoning. We refer to this as the *empirical approach* to argumentation semantics. It aims at *cognitively plausible* (as opposed to normatively optimal) semantics.

Presumably, people's capacity to perform reliable common sense reasoning in the context of conflicting information is the ultimate inspiration to researchers who want to build intelligent machines. It is quite striking that very little work in the 'argumentation in AI' community has investigated how people reason with argument structures.

One might ask: *why should we care about how humans reason?* One answer to this question is that, while deductive reasoning is inherently normative (e.g. helps us discover mathematical theorems), nonmonotonic reasoning and other common sense reasoning tasks are inherently 'psychologistic' (Pelletier & Elio 2005). That is, they are *defined* by what people do in real common sense situations, since there is not necessarily one obviously correct normative answer.

A second answer to the above question is that cognitively realistic reasoning is more successful when interacting with humans. For example, in the context of human-agent negotiation, it was shown that software agents that account for human reciprocity outperform agents that play the normative equilibrium strategies prescribed by game-theory (Gal & Pfeffer 2007). In the context of agent-human argumentation, it has also been argued that emotional arguments are more successful than purely rational arguments at persuading people (Mazzotta, de Rosis, & Carofiglio 2007). More generally, identifying which argument evaluation criteria are cognitively plausible can potentially enable building agents that are able to better argue with humans. It remains an open question, however, whether any of the existing models of argument evaluation in AI are cognitively plausible.

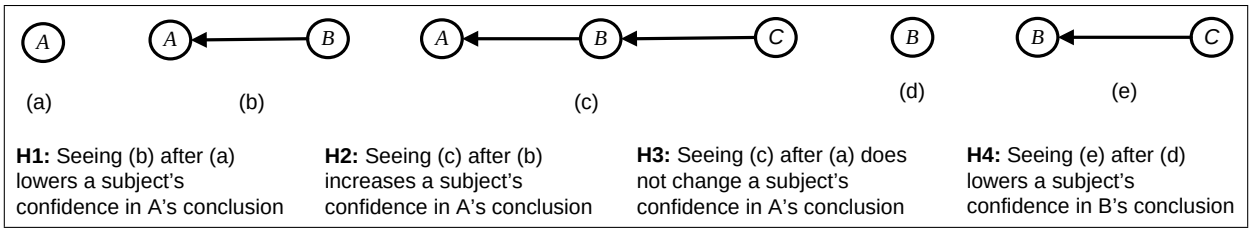

Figure 2: Summary of Hypotheses

Plausibility of Simple Reinstatement

We now report on a study of the cognitive plausibility of argument evaluation criteria for the basic notion of argument reinstatement (recall Figure 1). Abstractly, this structure is defined in the following argumentation framework: $AF = \langle\{A, B, C\}, \{(B, A), (C, B)\}\rangle$ in which argument A is attacked by argument B but reinstated by argument C.

Hypotheses

We are interested in testing three hypotheses on how a subject's confidence changes when attacking and reinstating arguments are presented.

> **H1**: An attack on an argument lowers the confidence that a subject has in the conclusion of that argument. That is, the confidence that a subject has in the conclusion of argument A is higher than the confidence the subject has in the conclusion of argument A when it is attacked by argument B.

Hypothesis $H1$ simply tests the cognitive plausibility of the basic notion of defeat, which has been subject to much research (Byrne 1999; Ford 2005).

> **H2**: Reinstatement of an argument raises the confidence that a subject has in the conclusion of the argument from the level of confidence that the subject has when the argument is attacked. That is, the confidence that a subject has in the conclusion of argument A is higher when it is attacked by argument B but reinstated by argument C compared to the confidence that the subject has in the conclusion of argument A when it is attacked by argument B but not reinstated.

Hypothesis $H2$ explicitly investigates the effect of reinstatement. It simply states that reinstatement *helps* the reinstated argument by improving our confidence in it. It says nothing about how the status of a reinstated argument compares to an argument that has never been attacked. This is a non-trivial question. Consider the example in Figure 1. Suppose a person only hears the statement "*Tweety flies because it is a bird*." Here, the person may presume, with some confidence, that Tweety indeed flies. Suppose a second person hears all three arguments, namely that "*Tweety flies because it is a bird*," that "*Tweety does not fly, because it is a penguin*," and that "*the observation that Tweety is a penguin is not reliable*." Would this person have more or less confidence in Tweety's flight? Classical Dung semantics all agree that the reinstated argument should be accepted, and have the same status as the undefeated argument.

However, there are two other sensible intuitions about this problem. On one hand, we might say that the first person should have higher confidence in Tweety's flight, since she was not exposed to any potential counter-arguments that may raise her doubt. On the other hand, we might say that the second person, who saw all three arguments, should have higher confidence in Tweety's ability to fly, since he saw that a possible objection/exception has been ruled out.

Against this background, hypothesis $H3$ tests whether the classical assumption, that reinstatement perfectly restores the argument status, is cognitively plausible.

> **H3**: The confidence that a subject has in the conclusion of an argument does not change when it is attacked by an argument but reinstated by another argument. That is, the confidence that the subject has in the conclusion of argument A is the same as the confidence in the conclusion of argument A when it is attacked by argument B but reinstated by argument C.

As a manipulation check, we introduce hypothesis $H4$ to ensure that in a reinstatement scenario, the attacking argument is indeed defeated by the reinstating argument. Hypothesis $H3$ is relevant only if hypothesis $H4$ holds.

> **H4**: In a reinstatement framework $A \leftarrow B \leftarrow C$, an attack on argument B by argument C lowers the confidence that a subject has in the conclusion of argument B. Argument A is not revealed to the subject.

The hypotheses above are summarised in Figure 2.

Dependent And Independent Variables

The independent variable (IV) in the experiment is the argumentation framework presented to the participant. Different values to this independent variable are:[1] A; $A \leftarrow B$; $A \leftarrow B \leftarrow C$; B; $B \leftarrow C$. Natural language arguments are used to instantiate the independent variable (see below).

The dependent variable (DV) is the confidence marked by the participant in the conclusion of arguments, argument A and argument B on a given 7-point interval scale. We used a 7-point scale (following (Politzer & Bonnefon 2006)) to mark the confidence, with 1 denoting 'certainly false' and 7 denoting 'certainly true.' A visual representation of the scale

[1] Here, we are using $A \leftarrow B$ to denote an argument graph in which argument A is defeated by argument B, and $A \leftarrow B \leftarrow C$ to denote the reinstatement graph, etc.

with black representing false and white representing true is given in Figure 3.

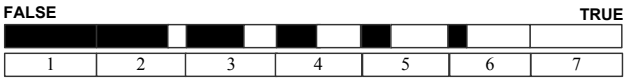

Figure 3: Visualization of the Scale

Participants

The survey was conducted using questionnaires in paper form. Random sample of people from offices, shopping malls and open spaces in Dubai, UAE were surveyed. There were two participant groups in this survey. These groups were labelled $G1$ and $G2$. Group $G1$ had 20 participants and group $G2$ had 18 participants totalling 38 participants in the survey. Group $G1$ tested hypothesis $H1$, $H2$ and $H3$ whereas group $G2$ tested hypothesis $H4$. Following are the IV value(s) assigned to each group, in sequence. For group $G1$, there were three assignments and for group $G2$, there were two assignments.

$G1 : A;\ \ A \leftarrow B;\ \ A \leftarrow B \leftarrow C$

$G2 : B;\ B \leftarrow C$

Questionnaire Design

There was an introduction in the beginning of the questionnaire informing the participants that the experiment was just to collect information on how people think. They were informed how much time it would take on an average to solve the problems. They were also informed that there were no trick questions and all that the participant had to do was to mark the answer that they felt correct, on the given scale.

The questionnaire was scrutinized to avoid double-barrelled questions (questions with more than one question embedded within them) and leading questions (questions that suggest the answer or an answer that is intended). All questions were short and concise to reduce the fatigue effect. The assignment of participants to questionnaires was random. All questions used simple common words and the words were chosen in such a way that they would not be misinterpreted by the participants.

There were six sets of natural language arguments (see appendix). Each set of arguments contained three arguments. $G1$ tested three conditions per argument set. So each participant in group $G1$ responded to 3 (conditions) x 6 (argument sets) = 18 problems. The order of conditions in each argument set was fixed and corresponded to the order of the IV assignments: A, $A \leftarrow B$, $A \leftarrow B \leftarrow C$. To reduce the order effect, half of the questionnaires contained the argument sets presented in reverse order but preserving the order of problems within each argument set.

Group $G2$ used the same argument sets that were used for group $G1$, but tested only two conditions per argument set. So each participant in group $G2$ responded to 2 (conditions) x 6 (argument sets) = 12 problems. Again the order of problem within an argument set was fixed and corresponded to the order of IV assignment B, $B \leftarrow C$. In order to reduce the order effect, half of the questionnaires contained the argument sets presented in reverse order but preserving the order of problems within an argument set.

The participants were allowed to answer problems only in the order of the problems as per the questionnaire. It was mandatory to answer all problems.

Hypothesis Analysis

Hypotheses $H1$, $H2$, and $H3$ were tested through an analysis of variance featuring conclusion acceptance as the dependent variable, problem as a 3-level predictor, and argument set as a 6-level measure. Standard contrast analyses were performed to compare the effects of different levels of the predictor. $H4$ was tested using a multivariate test featuring conclusion acceptance as the dependent variable, problem as 2-level predictor and argument set as 6-level measure.

Results

The base (when argument A is presented alone) acceptance rating of the conclusion when we average the scores across the 6 contents was 5.9 (SD = 0.8) whereas acceptance rating of the defeated conclusion (when argument A is attacked by argument B) as 4 (SD = 1.4). The acceptance rating of reinstated (when argument A is attacked by argument B but reinstated by argument C) was 5.2 (SD = 1.0).

Acceptance ratings were analyzed with a repeated-measure analysis of variance, with pattern as a 3-level predictor (*base*, *defeated* and *reinstated*) and 6 measures corresponding to the 6 contents. The repeated-measure analysis of variance detected a significant effect of pattern, $F(2, 18) = 14.1, p < 0.001, n_p^2 = 0.61$. We found that this overall effect is due to both the defeat and the reinstatement. As shown by a contrast analysis, ratings in the *base* condition were significantly higher than ratings in the *defeated* condition, $F(1, 19) = 26.8, p < 0.001, n_p^2 = 0.59$, and ratings in the *defeated* condition were themselves significantly lower than ratings in the *reinstated* condition, $F(1, 19) = 9.9, p < 0.005, n_p^2 = 0.34$. Although reinstatement increased the acceptability of a conclusion, the recovery was not perfect. Indeed, the ratings in the *reinstated* condition were still significantly lower than the ratings in the *base* condition, $F(1, 19) = 9.1, p = 0.007, n_p^2 = 0.32$.

The reliable effect of reinstatement must be related to the success of the reinstating manipulation, as shown by the results of the manipulation check. Averaging across 6 contents, the base acceptance ratings of defeaters was 5.1 (SD = 0.8) and the acceptance ratings for the attacked defeaters was 4.1 (SD = 0.7). The acceptance ratings for the manipulation check were analyzed with a repeated-measure analysis of variance, with pattern as 2-level predictor(*base defeater*, *attacked defeater*), and 6 measures corresponding to 6 contents. The test detected a significant effect of pattern $F(6, 12) = 3.8, p = 0.02, n_p^2 = 0.66$.

The results from the study thus support hypothesis $H1$ and $H2$. That is, when an argument is attacked by another argument, then the confidence in the conclusion of the argument being attacked significantly falls. The average acceptance ratings across 6 contents in the *base* condition is

5.9 (SD = 0.8) whereas in the *defeated* condition it is 4 (SD = 1.4). We also found that the reinstatement significantly increases the confidence in the conclusion of the argument being reinstated. We found that the average score across 6 contents in the *reinstated* condition as 5.2 (SD = 1.0). However, the results do not support hypothesis $H3$ which states that the confidence level in the *reinstated* condition is the same as in the *base* condition.

Discussion

Results show that the notion of reinstatement is cognitively plausible by supporting hypotheses $H1$ and $H2$. This is reinforced by the fact that the results support the manipulation check (hypothesis $H4$) showing that the reinstating argument effectively defeats the defeater of the reinstated argument. That is, reinstatement is achieved by defeating the defeater rather than merely supporting the main argument.

From the perspective of abstract argument evaluation criteria, the reinstatement framework produces the same extensions for all classical semantics. Hence, the empirical results do not indicate the preference of one semantic over another (this is an avenue of future research).

Having said that, the results do not support hypothesis $H3$, meaning that the recovery from a defeat, by a reinstatement, is not perfect. On one hand, this phenomenon is not captured by Dung's semantics, and a probabilistic approach to argument evaluation may be more accurate. On the other hand, the partial recovery is somewhat surprising, since it is reasonable to expect that people should have higher confidence in argument A when they see a possible objection/exception being ruled out. More investigation is needed to ascertain the precise nature of the cognitive processes that lead to this partial recovery. For example, revealing a defeater might trigger people to think of other possible objections, thus making full recovery less likely. This observation is relevant to agent-human argumentation. For example, an agent arguing with a human may be better off avoiding discussion that may reveal a defeater, even if the agent has a counter-argument to that defeater.

Related Work

Johan van Benthem has recently become a strong proponent of taking empirical findings from cognitive science seriously when working in logic at large. He terms this movement the '*new psychologism*' and states:

> '*Logicians analyzing natural language, or computer scientists modeling common sense, tend to go by their own intuitions, anecdotal evidence from colleagues, email surveys of sometimes surprising naiveness, and other easy procedures that avoid the laboratories and statistical packages of the world of careful experimental design. But even so, experimental evidence is relevant, in that these theories can be, and sometimes are, modified under pressure of evidence from actual usage, even when it comes through these home-grown sources.*' (van Benthem 2008)

In the context of epistemic logic, Pietarinen (Pietarinen 2003) argues for the important role of empirical findings from cognitive science in revising our logical conceptions of knowledge and belief. He argues that "*the interplay between logic and cognition is likely to reach increasingly wider and become increasingly prominent, encouraging fresh perspectives both from logical and semantic fields and from cognitive and neuroscientific communities.*"

Pelletier and Elio also argued for the importance of empirical studies of human reasoning in the formalisation of AI theories of default and inheritance reasoning (Pelletier & Elio 2005). They present empirical results on how people deal with benchmark problems in nonmonotonic reasoning, and identified a number of cases in which people deviate from the "AI answer."

There is also a long history of relevant literature from the psychology camp. Many psychologists defended the proposition that formal logic is an important part of human reasoning (Braine & O'Brien 1998; Rips 1994). These researchers argued for the existence of a *mental logic* that characterises 'core schemas' in human abstract reasoning. Another approach is the so-called *mental models* theory (Johnson-Laird 1983), which postulates that deductive reasoning is best explained by abstract models, as opposed to formal logic or schemas. For example, when people interpret a conditional "*If a card has an A on one side, then it has a 4 on the other*," they construct a mental representation such as '*[A] 4*.' A process of matching then determines the kinds of inferences people make, and also accounts for difficulties in logical reasoning. Other approaches to modelling human logical reasoning include probabilistic (Chater & Oaksford 2008) and neural-based approaches (Stenning & van Lambalgen 2008).

Closer to the present paper, there is a significant amount of literature on the *suppression of conditional reasoning*. This literature typically examines the effect of new information on previously made conclusions (Byrne 1999; Ford 2005). The degree to which the defeated argument is suppressed can vary depending on the type of argument and defeater involved (Politzer & Bonnefon 2006). In the argumentation jargon, this literature focuses on argument defeat. However, we are not aware of any studies that explore more complex argument structures, even for simple reinstatement.

Conclusion

By focusing on a very specific case study, namely simple reinstatement, we showed how empirical studies can inform the study of formal argumentation semantics. This positivist methodology offers a complementary source of inspiration and validation to the normative, example- and principle-based approaches typically found in the literature.

Results showed that while reinstatement is cognitively plausible in principle, it does not yield full recovery of the argument status, a notion not captured in Dung's classical model. Partial recovery is somewhat surprising, since it is reasonable to expect that people should have higher confidence in argument A when they see a possible objection/exception being ruled out. This observation is relevant to agent-human argumentation: an agent may be better off avoiding discussion that may reveal a defeater, even if the agent has a counter-argument to that defeater.

The reader should note that the aim of this paper is not to provide conclusive or comprehensive answers. Indeed, a different natural language instantiation arguments may lead to different results in subsequent studies. But the important lesson remains: psychological methodology can at least give us a new perspective on the types of problems we face in formalising argumentation, and understanding these problems can be very relevant to making argumentation useful.

References

Baroni, P., and Giacomin, M. 2007. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence* 171(10–15):675–700.

Baroni, P.; Giacomin, M.; and Guida, G. 2005. SCC-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence* 168(1–2):162–210.

Braine, M. D., and O'Brien, D. P. 1998. *Mental Logic*. Mahwah NJ, USA: Erlbaum.

Byrne, R. 1999. Counterexamples and the suppression of inferences. *Journal of Memory and Language* 40:347–373.

Caminada, M., and Amgoud, L. 2007. On the evaluation of argumentation formalisms. *Artificial Intelligence* 171:286–310.

Caminada, M. W. A. 2006a. On the issue of reinstatement in argumentation. In *Logics in Artificial Intelligence, 10th European Conference*, volume 4160 of *Lecture Notes in Computer Science*. Springer. 111–123.

Caminada, M. W. A. 2006b. Semi-stable semantics. In Dunne, P., and Bench-Capon, T., eds., *Proceedings of the 1st International Conference on Computational Models of Argument*, 121–130. Amsterdam, Nethrelands: IOS Press.

Chater, N., and Oaksford, M. 2008. *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. USA: Oxford University Press.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321–358.

Ford, M. 2005. Human nonmonotonic reasoning: the importance of seeing the logical strength of arguments. *Synthese* 146:71–92.

Gal, Y., and Pfeffer, A. 2007. Modeling reciprocity in human bilateral negotiation. In *National Conference on Artificial Intelligence (AAAI)*.

Horty, J. F. 2002. Skepticism and floating conclusions. *Artificial Intelligence* 135(1-2):55–72.

Johnson-Laird, P. N. 1983. *Mental Models*. USA: Harvard University Press.

Mazzotta, I.; de Rosis, F.; and Carofiglio, V. 2007. Portia: A user-adapted persuasion system in the healthy-eating domain. *IEEE Intelligent Systems* 22(6):42–51.

Pelletier, F. J., and Elio, R. 2005. The case for psychologism in default and inheritance reasoning. *Synthese* 146(1-2):7–35.

Pietarinen, A.-V. 2003. What do epistemic logic and cognitive science have to do with each other? *Cognitive Systems Research* 4(3):169–190.

Politzer, G., and Bonnefon, J.-F. 2006. Two varieties of conditionals and two kinds of defeaters help reveal two fundamental types of reasoning. *Mind & Language* 21(4):484–503.

Prakken, H. 2002. Intuitions and the modelling of defeasible reasoning: some case studies. In Benferhat, S., and Giunchiglia, E., eds., *Proceedings of the 9th International Workshop on Non-Monotonic Reasoning*, 91102.

Rahwan, I., and Simari, G. R., eds. 2009. *Argumentation in Artificial Intelligence*. Springer.

Rips, L. J. 1994. *The psychology of proof: deductive reasoning in human thinking*. USA: MIT Press.

Stenning, K., and van Lambalgen, M. 2008. *Human Reasoning and Cognitive Science*. USA: MIT Press.

van Benthem, J. 2008. Logic and reasoning: do the facts matter? *Studia Logica* 88(1):67–84.

Appendix: Natural Language Arguments

Argument Set 1

A: "The battery of Alex's car is not working. Therefore, Alex's car will halt."

B: "The battery of Alex's car has just been changed today. Therefore, the battery of Alex's car is working."

C: "The garage was closed today. Therefore, the battery of Alex's car has not been changed today."

Argument Set 2

A: "Louis applied the brake and the brake was not faulty. Therefore, the car slowed down."

B: "The brake fluid was empty. Therefore, the brake was faulty."

C: "The car had just undergone maintenance service. Therefore, the brake fluid was not empty."

Argument Set 3

A: "Mary does not moderate her phone usage. Therefore, Mary has a large phone bill."

B: "Mary has a speech disorder. Therefore, Mary moderates her phone usage."

C: "Mary is a singer. Therefore, Mary does not have a speech disorder."

Argument Set 4

A: "John has no way to know Leila's password. Therefore, Leila's emails are secured from John."

B: "Leila's secret question is very easy to answer. Therefore, John has a way to know Leila's password."

C: "Leila purposely gave a wrong answer to her secret question. Therefore, Leila's secret question is not very easy to answer."

Argument Set 5

A: "Mike's laptop does not have anti-virus software installed. Therefore, Mike's laptop is vulnerable to computer viruses."

B: "Nowadays anti-virus software is always available by default on purchase. Therefore, Mike's laptop has anti-virus software."

C: "Some laptops are very cheap and have minimal software. Therefore, anti-virus software is not always available by default."

Argument Set 6

A: "There is no electricity in the house. Therefore, all lights in the house are off."

B: "There is a working portable generator in the house. Therefore, there is electricity in the house."

C: "The fuel tank of the portable generator is empty. Therefore, the portable generator is not working."