Experiments for Assessing Floating Reinstatement in Argument-based Reasoning

Iyad Rahwan
[1]Masdar Institute of Science & Technology, [2]MIT, [3]University of Edinburgh

Jean-François Bonnefon
CNRS and Université de Toulouse

Mohammed Iqbal Madakkatel, Ruqiyabi Naz Awan
British University in Dubai

Sherief Abdallah
[1]British University in Dubai, [2]University of Edinburgh

Abstract

Various Artificial Intelligence semantics have been developed to predict when an argument can be accepted, depending on the abstract structure of its defeaters and defenders. These semantics can make conflicting predictions, as in the situation known as floating reinstatement. We argue that the debate about which semantics makes the correct prediction can be informed by the collection of experimental data about the way human reasoners handle these critical cases. The data we report show that floating reinstatement yields comparable effects to that of simple reinstatement, thus supporting preferred semantics over grounded semantics. Besides their theoretical value for validating and inspiring argumentation semantics, these results have applied value for developing artificial agents meant to argue with people.

Keywords: Argumentation; Semantics; Nonmonotonic reasoning; Behavioural Experiment;

Introduction

Argumentation has become a very fertile area of research in Artificial Intelligence (Rahwan & Simari, 2009), where a highly influential framework for studying argumentation-based reasoning was introduced by 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.[1]

Textual argument:

***A*:** Tweety flies because it is a bird.
***B*:** Tweety does not fly, because it is a penguin.
***C*:** The observation that Tweety is a penguin is not reliable.

Graphical structure:

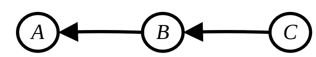

Figure 1: Defeat structure with reinstatement

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). 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.

These semantics typically come from a normative perspective, which relies on intuition and ad hoc hypothetical examples as to what constitutes correct reasoning. We will argue that there are limits to relying solely on this approach, and we will advocate the use of psychological experiments as a methodological tool for informing and validating intuitions about argumentation-based reasoning.

In this paper, we apply this experimental method to the problem of floating reinstatement. We will show that psychological experiments can help to evaluate these various semantics, and can provide unique insights even when all formal semantics are in agreement. Not only can these insights inform current and future semantics, but they are relevant to the design of software agents that can argue persuasively with humans, or provide reliable support to human evaluation of arguments (e.g., on top of argument diagramming tools).

Abstract Argumentation Frameworks

This section contains technical background only, whose outline is the following. Figure 1 displays the canonical graph of simple reinstatement, whereas Figure 2 displays the canonical graph of floating reinstatement. The main question is, in both cases, whether A can be accepted. For simple reinstatement, A is accepted by preferred as well as grounded semantics (to be defined below). For floating reinstatement, A is not accepted by grounded semantics, but is accepted by preferred

[1]While many notions of defeat exist, here we adopt the simple notion of *undercutting*: the defeater's conclusion explicitly negates the defeated argument's premise.

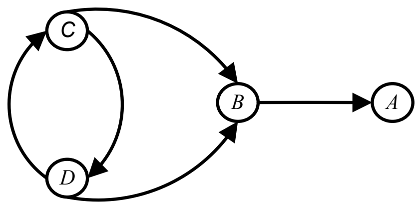

Figure 2: The canonical graph of defeat and floating reinstatement. Argument A is defeated by B, which is itself defeated by C as well as D, although C and D are mutual defeaters.

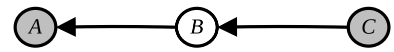

Figure 3: Single (complete, grounded, and preferred) extension in simple reinstatement. Accepted arguments are shaded.

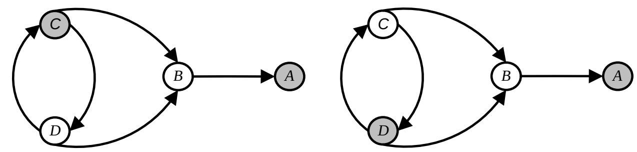

Figure 4: The two (complete, preferred) extensions in floating reinstatement. Accepted arguments are shaded.

semantics. Additionally, preferred semantics also accept C and D in the (formally defined) 'credulous' sense, but not in the 'sceptical' sense.

We now lay bare the technical background required to arrive at these conclusions. We begin with Dung's (1995) abstract definition of an argumentation framework.

Definition 1 (Argumentation framework)**.** *An* argumentation framework *is 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. An argument* α defeats *an argument* β *iff* $(\alpha, \beta) \in \rightharpoonup$, *also written* $\alpha \rightharpoonup \beta$.

The directed graphs displayed in Figures 1 and 2 will be our running examples all through the article. The critical issue with these examples is whether argument A can be accepted in spite of being defeated by argument B.

For a given set S of arguments, S^+ is the set of arguments that are defeated by the arguments in S. Formally, $S^+ = \{\beta \in \mathcal{A} \mid \alpha \rightharpoonup \beta \text{ for } \alpha \in S\}$. Conversely, for a given argument α, the set α^- is the set of all arguments that defeat α. Formally, $\alpha^- = \{\beta \in \mathcal{A} \mid \beta \rightharpoonup \alpha\}$.

Definition 2 (Conflict-freedom)**.** *Let* $\langle \mathcal{A}, \rightharpoonup \rangle$ *be an argumentation framework and let* $S \subseteq \mathcal{A}$. S *is* conflict-free *iff* $S \cap S^+ = \emptyset$.

In other terms, a set of arguments is *conflict free* if and only if no argument in that set defeats another.

Definition 3 (Defence)**.** *Let* $\langle \mathcal{A}, \rightharpoonup \rangle$ *be an argumentation framework, let* $S \subseteq \mathcal{A}$, *and let* $\alpha \in \mathcal{A}$. S defends α *if and only if* $\alpha^- \subseteq S^+$. *We also say that argument* α *is* acceptable with respect to S.

In other terms, a set of arguments *defends* a given argument if and only if it defeats all its defeaters.

Example 1. *In the graph displayed in Figure 1, the set* $\{A, C\}$ *is conflict free, but the set* $\{A, B\}$ *is not, and neither is the set* $\{B, C\}$. *Because the set* $\{C\}$ *defeats all the defeaters of* A, *we can say that the set* $\{C\}$ *defends argument* A. *In the graph displayed in Figure 2, the only conflict-free sets (apart from trivial ones containing single arguments) are* $\{A, C\}$ *and* $\{A, D\}$. *Either one of the sets* $\{C\}$, $\{D\}$, *or* $\{C, D\}$, *defends* A *against all its defeaters.*

We now define the *characteristic function* of an argumentation framework.

Definition 4 (Characteristic function)**.** *Let* $AF = \langle \mathcal{A}, \rightharpoonup \rangle$ *be an argumentation framework. The* characteristic function *of* AF *is* $\mathcal{F}_{AF}$: $2^{\mathcal{A}} \to 2^{\mathcal{A}}$ *such that, given* $S \subseteq \mathcal{A}$, *we have* $\mathcal{F}_{AF}(S) = \{\alpha \in \mathcal{A} \mid S \text{ defends } \alpha\}$.

Applied to an argument set S, the characteristic function returns the set of all arguments defended by S. Because we are only dealing in this article with one argumentation framework at a time, we will use the notation $\mathcal{F}$ instead of $\mathcal{F}_{AF}$.

We now turn to various so-called *extensions* that can characterise the collective acceptability of a set of arguments. Essentially, these extensions provide different possible ways to group self-defending arguments together. These extensions will be used subsequently to define the argument evaluation criteria that we study empirically in this paper.

Definition 5 (Complete/grounded/preferred extensions)**.** *Let* S *be a conflict-free set of arguments in framework* $\langle \mathcal{A}, \rightharpoonup \rangle$.

- *S is a* complete *extension iff* $S = \mathcal{F}(S)$.
- *S is a* grounded *extension iff it is the minimal complete extension with respect to set inclusion.*
- *S is a* preferred *extension iff it is a maximal complete extension with respect to set inclusion.*

S is a complete extension if and only if *all* arguments defended by S are also in S (that is, if S is a fixed point of the operator $\mathcal{F}$). There may be more than one complete extension, each corresponding to a particular consistent and self-defending viewpoint.

Example 2. *In the graph displayed in Figure 1, the set* $\{C\}$ *is not a complete extension, because it defends* A *without including it. The set* $\{B\}$ *is not a complete extension because it includes* B *without defending it against* C *–see Figure 3. The only complete extension is* $\{A, C\}$. *The graph displayed in Figure 2 has two complete extensions,* $\{A, C\}$ *and* $\{A, D\}$ *–see Figure 4.*

A grounded extension contains all the arguments in the graph that are not defeated, as well as all the arguments which are defended directly or indirectly by non-defeated arguments. This can be seen as a non-committal view (characterised by the *least* fixed point of $\mathcal{F}$). As such, there always exists a unique grounded extension.

Example 3. *The graph in Figure 1 has only one complete extension, $\{A,C\}$, which is also its grounded extension. The graph in Figure 2 has two complete extensions $\{A,C\}$ and $\{A,D\}$, but none of this is the grounded extension, because there is no node in the graph that is initially undefeated. In that case, the grounded extension is the empty set.*

A preferred extension is a bolder, more committed position that cannot be extended (by accepting more arguments) without causing inconsistency. Thus a preferred extension can be thought of as a maximal consistent set of hypotheses. There may be multiple preferred extensions, and the grounded extension is included in all of them.

Example 4. *The graph in Figure 1 has only one complete extension, $\{A,C\}$, which is also a preferred extension. The graph displayed in Figure 2 has two complete extensions $\{A,C\}$ and $\{A,D\}$, and both qualify as preferred extensions.*

Now we can define the status of an individual argument within the graph, that is, we can define criteria for accepting or not each individual argument. The main question in this paper is whether people evaluate a reinstated argument sceptically or credulously in accordance with the definition below.

Definition 6 (Argument status)**.** *Let $\langle \mathcal{A}, \rightharpoonup \rangle$ be an argumentation framework, and $\mathcal{E}_1, \ldots, \mathcal{E}_n$ its extensions under a given semantics. Let $\alpha \in \mathcal{A}$ and $i = 1, \ldots, n$.*

- α *is accepted in the sceptical sense iff* $\alpha \in \mathcal{E}_i$, $\forall \mathcal{E}_i$.
- α *is accepted in the credulous sense iff* $\exists \mathcal{E}_i$ *where* $\alpha \in \mathcal{E}_i$.
- α *is rejected iff* $\nexists \mathcal{E}_i$ *such that* $\alpha \in \mathcal{E}_i$.

Under the grounded semantics, any argument that belongs to the unique grounded extension is accepted both in the credulous and the sceptical sense, and any argument that does not belong to the unique grounded extension is rejected. Under the preferred semantics, an argument is sceptically accepted if it belongs to all preferred extensions; but it can also be credulously accepted if it belongs to at least one preferred extension. If an argument is neither sceptically nor credulously accepted, it is rejected.

Example 5. *The graph displayed in Figure 1 has only one complete extension, $\{A,C\}$, which is grounded as well as preferred. As a consequence, arguments A and C are accepted by grounded as well as preferred semantics, both in the credulous and sceptical sense. The graph displayed in Figure 2 has an empty grounded extension, which means that no argument should be accepted under a grounded semantics. Under a preferred semantics, though, two extensions are identified, $\{A,C\}$ and $\{A,D\}$. From these extensions, only A can be accepted in a sceptical sense, but A, C, and D can all be accepted in a credulous sense.*

What Validates a Semantics?

As established above, different semantics can have different takes on which arguments can be accepted within a given argumentation framework. The question then arises of evaluating the different claims made by different semantics.

Most semantics for argumentation-based reasoning in Artificial Intelligence are based on intuition as to what constitutes correct reasoning. This intuition is informed by specific (hypothetical or real) argumentation scenarios in which a particular semantics draws the desired intuitive answer. This *example-based approach* (to borrow a term from Baroni & Giacomin, 2007) is problematic, since one can often construct other examples with the same logical structure, in which the proposed semantics draws counter-intuitive conclusions. For example, Horty (2002) famously devoted a whole paper to demonstrate counter-intuitive results with floating conclusions in default reasoning. Baroni and Giacomin (2007) made a compelling case for the limitations of the example-based approach, noting that even in relatively simple examples, there might not be a consensual intuition on what should be the correct conclusion. In parallel, Prakken (2002) observed that intuitions about given examples were helpful for generating new investigations, but less helpful as critical tests between different semantics.

To overcome the limitations of the example-based approach, a number of authors recently advocated a more systematic, axiomatic, *principle-based* approach. In this approach, alternative semantics are evaluated by analysing whether they satisfy certain principles, or quality postulates. Such postulates include the *reinstatement criterion*, according to which an argument must be included in any extension that reinstates it, and *directionality criterion* which requires that an argument's status should only be affected by the status of its defeaters (Baroni & Giacomin, 2007).

The principle-based approach provides a significant improvement over the basic example-based approach, since it enables claims that transcend individual examples and characterise semantics more generally. The source of the general postulates, however, is still the researcher's intuition as to what correct reasoning ought to be. In sum, most of the extent validation of various argumentation semantics, example-based or principle-based, relies on normative claims based on intuition. We now suggest that this normative-intuitive perspective could be adequately complemented with descriptive, *experimental* evidence about how people actually reason from conflicting arguments.

The Experiment-based Approach

There is a growing concern within the Artificial Intelligence community that logicians and computer scientists ought to give serious attention to cognitive plausibility when assessing formal models of reasoning, argumentation and decision-making. For example, Benthem (2008) strongly supports the rise of a *new psychologism* in logic at large, arguing that although logicians and computer scientists have tended to go by intuition and anecdotal evidence, formal theories can be modified under pressure from evidence obtained though careful experimental design.

Pelletier and Elio (2005) also argued extensively for the importance of experimental data when formalizing default and inheritance reasoning, arguing that default reasoning is particularly psychologistic in that it is *defined* by what people do. Their own results have been complemented by a dynamic experimental literature consisting of controlled tests of human default reasoning (e.g., Bonnefon, Da Silva Neves, Dubois, & Prade, 2008; Ford & Billington, 2000; Pfeifer & Kleiter, 2009).

Finally, and in close relation to the problems of simple and floating reinstatement that we have introduced in the previous section, Horty (2002) implicitly appealed to descriptive validation when highlighting the issues that floating conclusions raise for sceptical semantics:

> There is a vivid practical difference between the two skeptical alternatives. [. . .] Which alternative is correct? I have not done a formal survey, but most of the people to whom I have presented this example are suspicious of the floating conclusion (p.64).

We believe that the field of computational argumentation can indeed benefit from the same kind of formal surveys that have been conducted in the field of default reasoning, and that have been generally called for in Artificial Intelligence. To our knowledge, only very few articles have explicitly sought to inform formal models of argumentation with experimental evidence, and these experimental data have only been collected in relation to the specific issue of argumentation-based decision making (e.g., Dubois, Fargier, & Bonnefon, 2008). What we offer in this article is an experimental investigation of the basic issue of how people reason from the complex argument structure corresponding to floating reinstatement, and whether one of the current available semantics can capture their reasoning.

Recently, we conducted experiments on the simple reinstatement structure, across a varied set of linguistic contents (Madakkatel, Rahwan, Bonnefon, Awan, & Abdallah, 2009). Our study revealed that participants reasoned in a way that reflected the formal notions of defeat and reinstatement: Their confidence in an argument A decreased when it was attacked by an argument B, but bounced back up when B itself was attacked by a third argument C. These findings are in agreement with grounded as well as preferred semantics (and others). What neither semantics could predict, though, is the finding that the recovery of argument A was not complete when reinstated by argument C: Confidence in A in presence of B and C did not raise back to its former level, when A was presented alone.

Our present study offers an experimental comparison of the simple reinstatement structure to the more complex structure known as floating reinstatement, shown in Figure 2.The present study seeks to answer the following questions: Does floating reinstatement restore the confidence in the conclusion of argument A, and does it do so to the same extent as simple reinstatement? (A 'yes' to both questions would go against the predictions of grounded semantics.) If so, does the effectiveness of floating reinstatement require that participants manifest a preference for either C over D or D over C? (A 'yes' would provide support to the predictions of credulous preferred semantics, a 'no' would provide support to the predictions of sceptical preferred semantics.)

Method

Fourty-seven participants were randomly approached in offices, shopping malls, and open spaces in Dubai. Participants read an introduction to the task, informing them that the purpose of the experiment was to collect information about how people thought, that the task included no trick question, and that they simply had to mark the answer that they felt correct. They were randomly assigned to two experimental groups corresponding to simple and floating reinstatement, respectively, then solved 12 problems, following a 3-level, 4-measure within-participant design.

The 3-level independent variable was the *Pattern* of the problem (Base, Defeated, Reinstated). In the Base pattern, participants were only presented with argument A; in the Defeated pattern, participants were presented with arguments A and B; finally, in the Reinstated pattern, participants were presented with the three arguments A, B, and C (in the simple reinstatement group) or with the four arguments A, B, C, and D (in the floating reinstatement group).

The linguistic contents of arguments A, B, C, and D were taken from four different argument sets (see Appendix). All participants saw each argument set in its Base, Defeated, and Reinstated versions. The order of argument sets within the questionnaire was counterbalanced across participants (two different orders), but the order of Pattern within each argument set was fixed across the experiment. Participants had to answer every problem, in the order they appeared in the questionnaire, without looking at the next problem in the questionnaire. For each problem, participants had to assess the conclusion of argument A, using a 7-point scale anchored at *certainly false* and *certainly true*.

In addition, participants rated their understanding of each problem ('How clearly did you understand the problem?') on a 7-point scale anchored at *Not at all* and *Completely*. Lastly, participants in the floating reinstatement group answered the following question about the four reinstated problems: Do you think that (i) C is a better argument than D, (ii) D is a better argument than C, or (iii) C and D are about equally good?

Results

Figure 5 displays the average confidence in the conclusion of A, as a function of Pattern and Type of reinstatement, averaged across the contents and participants. The visual inspection of Figure 5 already suggests that the results are very similar for the two groups. This preliminary intuition was confirmed by the results of a mixed-design analysis of variance, using the confidence in the conclusion as a dependent

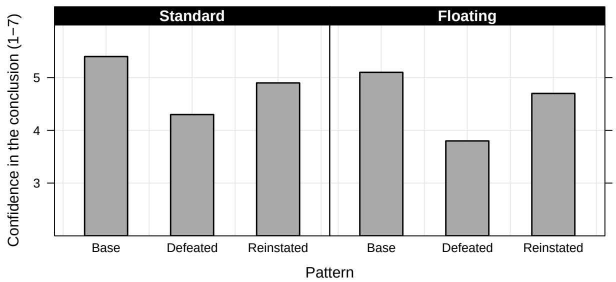

Figure 5: Reinstatement is as effective in its floating form as in its simple form. Confidence in the conclusion of an argument decreases when the argument is defeated, and is then imperfectly restored when its defeater is itself defeated, whether by a single argument (simple reinstatement) or by two mutually defeating arguments (floating reinstatement).

variable, pattern as a 3-level within-subject predictor (Base, Defeated, Reinstated), the type of reinstatement as a 2-level between-group variable (Simple, Floating), and four measures corresponding to the four linguistic contents.

The multivariate test detected a significant effect of Pattern, $F(8,38) = 6.1$, $p < .001$, $\eta_p^2 = .56$. It did not, however, detect a significant main effect of Type of reinstatement $F(4,42) < 1$, $p = .79$, $\eta_p^2 = .04$, nor a significant interaction between Pattern and Type, $F(8,38) = 1.2$, $p = .32$, $\eta_p^2 = .20$.

The overall effect of Pattern reflected a successful defeat followed by a successful reinstatement. As shown by contrast analysis, confidence ratings in the Defeated condition were significantly lower than ratings in the Base condition, $F(1,45) = 34.9$, $p < .001$, $\eta_p^2 = .44$, and this difference was not moderated by the Type of reinstatement (there is indeed no reason that it should be), $F(1,45) < 1$, $p = .67$, $\eta_p^2 < .01$. The confidence ratings in the Reinstated condition were significantly greater than in the Defeated condition, $F(1,45) = 13.7$, $p < .001$, $\eta_p^2 = .23$, and this difference (more interestingly this time) was not moderated by the Type of reinstatement, $F(1,45) < 1$, $p = .60$, $\eta_p^2 < .01$. Just as in our earlier study (Madakkatel et al., 2009), reinstatement is not perfect, as ratings in the Reinstated condition remain significantly lower than in the Base condition, $F(1,45) = 9.0$, $p < .01$, $\eta_p^2 = .17$. Again, there is no evidence whatsoever of a moderation by Type of reinstatement, $F(1,45) < 1$, $p = .92$, $\eta_p^2 < .01$.

So far, results suggest that floating reinstatement has an effect that is identical to classic reinstatement. We further note that although subjects found the floating reinstatement problems slightly harder to understand than the simple reinstatement problems, this difference appeared to play no role in the ratings they gave for their confidence in the conclusion. The average understanding rating was 4.6 (SD $= 1.1$) for simple reinstatement problems, compared to 4.0 (SD $= 0.9$) for floating reinstatement problems, $t(45) = 2.0$, $p = .05$. However, a regression analysis seeking to predict acceptance of reinstated arguments on the basis of problem understanding, Type of reinstatement (dummy coded, 1 for floating), and the interaction term between these two predictors, failed to find any significant effect. The interaction term in particular achieved a standardized β of .19, non-reliably different from zero, $t = 0.32$, $p = .75$.

The effectiveness of floating reinstatement does not appear to result from the subjects manifesting a preference for one of the mutually defeated arguments. We conducted four repeated-measure analyses of variance, one for each argument set, with conclusion acceptance as a dependent variable, pattern as a 2-level predictor (Defeated, Reinstated), and preference as a dummy coded between-group variable (0 for subjects who said the two mutually defeating arguments were equally good, 1 otherwise). The interaction term between pattern and preference did not achieve statistical significance in any of the four analyses, all Fs in the $0.5 - 1.5$ range, all ps in the $.23 - .48$ range.

Discussion

We applied the experimental approach to understand how people deal with floating reinstatement in argument-based reasoning, a case that has puzzled theoreticians for many years. Our results suggest that, empirically speaking, floating reinstatement works exactly as well as simple reinstatement. Participants' confidence in an argument A decreased when it was attacked by an argument B, but bounced back up when B itself was attacked by two mutually defeating arguments C and D. These results clearly speak in favour of preferred semantics. Results also suggest that the sceptical version of preferred semantics might be more cognitively plausible than the credulous version, since the effect of floating reinstatement was not dependent on participants showing a preference for one of the two mutually defeating arguments. This question is not yet settled, though, since the data do not make it clear whether participants would be willing to commit to accepting one of the mutually defeating arguments C and D. This aspect requires further investigation.

Besides their theoretical value, our results also have applied value for developing agents that are meant to argue with human users. We already know that artificial agents can achieve better negotiation results with human users when they do not play normative equilibrium strategies, but rather adopt boundedly rational strategies inspired from human behavioural data (Lin, Kraus, Wilkenfeld, & Barry, 2008). Generally speaking, we may expect that artificial agents may similarly be more successful when arguing with human users, if they can anticipate human reactions to various abstract argumentation frameworks. With that goal in mind, our results suggest that artificial agents may be better off avoiding discussion that may reveal a defeater, even if the agent has a counter-argument to that defeater; but should be ready to use floating reinstatement as well as simple reinstatement in order to neutralise a defeater raised by the human user. These kinds of heuristics can be incorporated into a decision-theoretic model of a persuasive agent that interacts with users using natural language (e.g. to promote a healthy diet (Mazzotta, Rosis, & Carofiglio, 2007). Going beyond our specific results, by building up a corpus of argument structures and how they are evaluated, it may be possible to use machine learning techniques to build models that predict how people will react to novel argument structures.

Independently of our specific results, we hope to have convinced the reader that the wealth of scientific methodology from psychology can give a new perspective on the problems raised when formalising argumentation and developing argument evaluation semantics. We hope that our claims and findings can prompt researchers working on the computational modelling of argument to explore new avenues of investigation inspired by, and validated against, empirical evidence from psychology and cognitive science.

We also hope to have excited cognitive scientists about the growing literature on formal models of argumentation. These models, and their associated normative properties, have great potential in complementing existing research on human reasoning, and providing conceptual means for dealing with highly complex inference structures.

References

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

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

Bonnefon, J. F., Da Silva Neves, R. M., Dubois, D., & Prade, H. (2008). Predicting causality ascriptions from background knowledge: Model and experimental validation. *International Journal of Approximate Reasoning*, *48*, 752–765.

Dubois, D., Fargier, H., & Bonnefon, J. F. (2008). On the qualitative comparison of decisions having positive and negative features. *Journal of Artificial Intelligence Research*, *32*, 385–417.

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., & Billington, D. (2000). Strategies in human nonmonotonic reasoning. *Computational Intelligence*, *16*, 446–468.

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

Lin, R., Kraus, S., Wilkenfeld, J., & Barry, J. (2008). Negotiating with bounded rational agents in environments with incomplete information using an automated agent. *Artificial Intelligence*, (accepted).

Madakkatel, M. I., Rahwan, I., Bonnefon, J.-F., Awan, R. N., & Abdallah, S. (2009). Formal argumentation and human reasoning: The case of reinstatement. In T. Bench-Capon, S. Parsons, & H. Prakken (Eds.), *The uses of computational argumentation: Papers from the aaai fall symposium.* AAAI Press. (Technical Report FS-09-06)

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

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

Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability logic. *Journal of Applied Logic*, *7*, 206–217.

Prakken, H. (2002). Intuitions and the modelling of defeasible reasoning: some case studies. In S. Benferhat & E. Giunchiglia (Eds.), *Proceedings of the 9th international workshop on non-monotonic reasoning* (p. 91102).

Rahwan, I., & Simari, G. R. (Eds.). (2009). *Argumentation in artificial intelligence*. Springer.

Materials

Argument Set 1

(A) Cody does not fly. Therefore, Cody is unable to escape by flying.

(B) Cody is a bird. Therefore, Cody flies.

(C) Cody is a rabbit. Therefore, Cody is not a bird.

(D) Cody is a cat. Therefore, Cody is not a bird.

Argument Set 2

(A) Smith does not follow American spelling. Therefore, Smith writes ‘colour’ instead of ‘color’.

(B) Smith speaks American English. Therefore, Smith follows American spelling.

(C) Smith was born and brought up in England. Therefore, does not speak American English.

(D) Smith was born and brought up in Australia. Therefore, does not speak American English .

Argument Set 3

(A) The car did not slow down. Therefore, the car approached the signal at the same speed or higher.

(B) Louis applied the brake. Therefore, the car slowed down.

(C) Louis applied the accelerator instead. Therefore, Louis did not apply the brake.

(D) Louis applied the clutch instead. Therefore, Louis did not apply the brake.

Argument Set 4

(A) Stephen is not guilty. Therefore, Stephen is to be free from conviction.

(B) Stephen was seen at the crime scene at the time of the crime. Therefore, Stephen is guilty.

(C) Stephen was having dinner with his family at the time of crime. Therefore, Stephen was not seen at the crime scene at the time of the crime.

(D) Stephen was watching football with his friends in the stadium at the time of the crime. Therefore, Stephen was not seen at the crime at the time of the crime.