On the Cognitive Plausibility of Abstract Argument Evaluation Criteria: The Case of Argument Reinstatement

Mohammed Iqbal Madakkatel

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science in Information Technology
Faculty of Informatics

The British University in Dubai

June 2009

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Mohammed Iqbal Madakkatel*)

Abstract

Classical reasoning such as reasoning based on propositional logic is monotonic in the sense that adding new information does not remove any previously made conclusion. Common sense suggests that the failure of monotonic reasoning is widespread. Many a time, we jump to conclusions and then we correct our conclusions based on further information as it arrives. Reasoning of this kind is called as nonmonotonic reasoning. Recently, the study of nonmonotonic reasoning has appealed to the powerful notion of argument through the proliferation of so-called argumentation systems. A very influential approach to argumentation systems completely abstracts the origin and the internal structure of the arguments. The focus, instead, is on the relationships between these abstract arguments using defeat relations. The defeat structures can be in different forms such as mutual attack (one argument attacks another argument and the attacked argument attacks the attacker), reinstatement (an argument reinstates another argument by defeating the defeater) and a cycle of attack. The obvious question here is to identify which arguments are rejected, accepted and undecided in such defeat structures. Extension-based abstract argument evaluation criteria (also known as extension-based semantics) can be thought of as criteria for making this decision and have been studied in detail in the literature. These evaluation criteria have been mainly developed for obtaining desirable formal or computational properties, largely based on intuition. The cognitive plausibility of such evaluation criteria has mostly been ignored. However, it is crucial to understand the cognitive plausibility of such evaluation criteria if we are to build software agents capable of interacting persuasively with humans through arguments. This study is an attempt to explore the cognitive plausibility of abstract argument evaluation criteria.

Cognitive plausibility of abstract argument evaluation criteria is explored by conducting psychological experiments. Scenarios of standard reinstatement (an argument reinstating another argument by defeating the defeater) and the floating reinstatement (two mutually conflicting arguments reinstating another argument by defeating the defeater) are studied in detail. The empirical results show that the notion of reinstatement in abstract argumentation is cognitively plausible by supporting both grounded and preferred semantics. The results also show that the notion of floating reinstatement is cognitively plausible by not supporting the grounded semantics. Lack of a significant interaction between the pattern and the reasoners' preference (or lack of preference) for one of the two mutually conflicting defenders in the tests indicates the existence of a different cognitively plausible notion that cannot be explained using the abstract argument evaluation criteria. The notion is that there is no clear endorsement for both credulous preferred and sceptical preferred semantics in floating reinstatement. The results also suggest that a floating reinstatement has an effect that is not significantly different from that of the standard form. That is, the mutually conflicting nature of defenders does not play any role that undermines the job of reinstating the main argument. Importantly, only partial recovery is achieved in both scenarios of reinstatement and the idea of such a partial recovery is not dealt with in the abstract argumentation theory.

To my wife, daughter and parents

Acknowledgements

I wish to express my sincere and deep gratitude to my dissertation supervisor Dr. Iyad Rahwan for motivating me and showing me the path to cognitive science. I wish to thank him for his highly commendable academic support and valuable and thoughtful advice throughout the course of this dissertation. I would like to express my sincere and deep gratitude to my co-supervisor Dr. Sherief Abdallah for his valuable and thoughtful advice.

My sincere and deep gratitude goes to Jean-Francois Bonnefon, cognitive research scientist at the French national research agency (CNRS) for his guidance on experiments and help in analysing the results. His quick and timely advice throughout the course of this dissertation is highly commendable.

It is my pleasure to express my sincere gratitude to Dr. Naz Awan, Faculty of Education at the British University in Dubai for her valuable suggestions and recommendations.

I express my gratitude to all staff and students in the British University in Dubai for their help and support. I must express my gratitude to the students at various universities in Knowledge Village, Dubai, also for participating in the surveys.

Finally, I would like to thank my managers and colleagues in the office (Information Technology Department, Commercial Bank of Dubai, Dubai) for their support.

Contents

List of Figures

List of Tables

Chapter 1

Overview

1.1 Introduction

In classical reasoning, such as reasoning based on propositional logic, inference from a set of premises S to a conclusion c is done through a so-called *inference procedure* $\vdash$, and we write $S \vdash c$. Inference procedures are collections of rules for inferring new information from existing information. An example of an inference procedure is *modus ponens*, in which q is inferred from the premises p and $p \rightarrow q$. An inference procedure implements a (semantic) *consequence relation* that specifies what sentences follows logically from given premises, written $S \models c$. Classical reasoning is *deductive*, meaning that the conclusions necessarily follow from the premises.

Classical reasoning is *monotonic* in the sense that adding new premises does not remove any previously made conclusion. So if we initially have $S \vdash c$, then it cannot be the case that $S \cup S' \vdash \neg c$ for any S'. This contrasts with so-called *non-monotonic* reasoning in which new information may *defeat* previously made conclusions [35]. For instance, we may conclude that an object that looks red to us is of a red colour. We may later learn that the object was illuminated by red light. Since we know that an object that is illuminated by red light may look red to us even though it may not be of a red colour, we could conclude that the object in question may not be red [35].

Many approaches exist for studying non-monotonic reasoning, especially in Artificial Intelligence literature (e.g. default logic [39] and circumscription [26]). More recently, the study of non-monotonic reasoning has appealed to the powerful notion of *argument* through the proliferation of so-called *argumentation systems* [36], as pioneered by philosopher John Pollock and computer scientist Ronald Loui in 1987 [35, 25].

In AI, argumentation systems are used in automating non-monotonic reasoning, such as reasoning from inconsistent information [17, 4]. They are also used to automate the interaction (e.g. negotiation, persuasion) among autonomous software agents [37, 3, 38].

The cognitive plausibility of argument evaluation criteria demands the notion of *argument* to be further explored, particularly in terms of natural language arguments. Consider an example of argument: "*Tweety is a bird. Therefore, Tweety can fly*". It is quite true that bird flies. Now consider another interesting argument against the previously arrived conclusion that Tweety flies: "*Tweety is a penguin. Therefore, Tweety cannot fly*". If arguments are labelled as argument A, argument B and so on and here the former argument as argument A and latter argument as argument B, it can be seen that argument B attacks argument A. Since in the literature on argumentation, the terms *attack* and *defeat* are used interchangeably, we can also say argument B defeats argument A. A third argument (argument C) "*Tweety is a genetically modified penguin. Therefore, Tweety can fly*" defeats argument B and in defeating argument B, it reinstates argument A (the main argument). It is often useful to visualise these defeat structures (also called argumentation frameworks) using *argument graphs*. The reinstatement defeat structure described above is shown in Figure 1.1. Throughout this report we will refer

to this argument reinstatement as *Standard Reinstatement* or simply *Reinstatement* in order to differentiate it from *Floating Reinstatement* which is defined later.

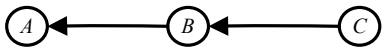

Figure 1.1: Reinstatement

The defeat structures can be cyclic also. Consider another example consisting of three arguments forming a cycle of defeats.

Argument A: "Ilham says that Roshna is not reliable, therefore everything that Roshna says cannot be relied on."

Argument B: "Roshna says that Iqbal is not reliable, therefore everything that Iqbal says cannot be relied on."

Argument C: "Iqbal says that Ilham is not reliable, therefore everything that Ilham says cannot be relied on."

The *argument graphs* depicting the aforementioned example is shown in Figure 1.2.

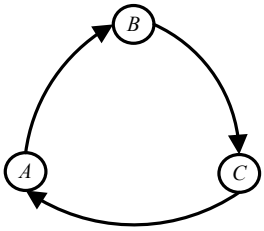

Figure 1.2: Cycle of Attacks - 3 Arguments

While many argumentation systems exist in the literature, a very influential *abstract* approach was suggested by Dung [17]. This approach completely abstracts the internal structure of the argument and the origin of the argument. The focus, instead, is on the relationships between these abstract arguments, as defined by a binary *defeat* (or *attack*) relation.

Consider an argumentation framework of two arguments, mutually attacking each other as shown in Figure 1.3. As long as both argument A and argument B attack each other, they fit into this argumentation framework with no concern on their internal structure and origin. For instance, now we can instantiate this argumentation framework with the following two natural language arguments, often referred to as Nixon Diamond.

Argument A: "Nixon is a Quaker. Therefore, Nixon is a pacifist"

Argument B: "Nixon is a Republican. Therefore, Nixon is not a pacifist'

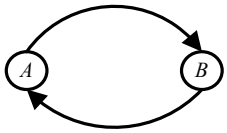

Figure 1.3: Mutual Attack

Another interesting characteristic argumentation framework, often referred to as floating

arguments, is shown in Figure 1.4. Prakken and Vreeswijk [36] explain floating arguments using an example of arguments A, B, C and D such that A defeats B, B defeats A, A defeats C, B defeats C and C defeats D. Since no argument is undefeated, all of them are defensible. Prakken and Vreeswijk observe that, "However, it might be argued that for C and D this should be otherwise: since C is defeated by both A and B, C should be overruled: as far as the status of C is concerned, there is no need to resolve the conflict between A and B: the status of C 'floats' on that of A and B. And if C should be overruled, then D should be justified, since C is its only defeater" [36]. Throughout this report we will refer to the reinstatement in floating arguments as *Floating Reinstatement*.

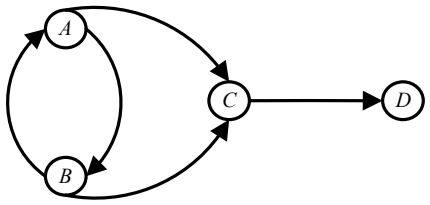

Figure 1.4: Floating Arguments

The abstract approach of studying argumentation is useful in that it allows one to focus on the macro structure of the argumentation scenario, as opposed to the micro structure of individual arguments. Different so-called *extension-based semantics* have been studied in the literature [36]. A particular semantics can be thought as a criterion for assigning status (e.g. accepted, rejected, or undecided) to arguments in a given argument graph, by appealing to the structural properties of that graph. The set of arguments accepted under a particular status assignment is called an *extension* of the argumentation framework. Thus semantics of abstract argumentation frameworks are often called *extension-based semantics*.

A variety of extension-based semantics have been studied in the literature, including Dung's original semantics [17] and many others [2]. In addition to identifying possible extensions, two acceptance *attitudes* have been studied in the literature. Under the *sceptical* attitude, an argument is accepted if it belongs to the intersection of all extensions defined by the given semantics. Under the *credulous* attitude, on the other hand, an argument is accepted if it belongs to at least one extension defined by the semantics used.

1.2 Objectives

In this research programme, we intend to conduct psychological experiments to assess the cognitive plausibility of various abstract argument evaluation criteria (i.e. extension-based semantics) from the literature. Our aim is to understand which semantics (if any) humans use when they evaluate different argumentation structures, and whether entirely new semantics are needed to achieve cognitive plausibility. We believe this work is important for a number of reasons:

1. **Theoretical advancement:** If existing argumentation semantics fail to capture human evaluation of arguments, then observing how humans actually evaluate arguments can inform the design of new semantics based on well-founded empirical data.
2. **Agent-human argumentation:** Understanding how humans evaluate arguments in different argumentation scenarios can inform the design of software agents that engage in argumentation and persuasion with humans. For example, if a set of arguments can be framed using different structures, the agent may choose the structure that is most likely to successfully persuade the user, given its knowledge of the human's approach to evaluation.

3. **Delegation to agents:** Understanding how humans argue can eventually help us program software agents that better represent human interests as they argue on their behalf (with other human or software agents).

In order to explore the cognitive plausibility of argument evaluation criteria in the reinstatements of arguments, we will study standard reinstatement and floating reinstatement. Experiments will test the confidence of human subjects in the conclusion of arguments. The confidence level marked by the subjects in the conclusion of arguments is analysed to relate them to the already established evaluation criteria, to check whether humans follow those evaluation criteria.

1.3 Research Questions

Argument evaluation criteria have been mainly developed for obtaining desirable formal or computational properties, largely based on intuition. The cognitive plausibility of such evaluation criteria has mostly been ignored. However, it is crucial to understand the cognitive plausibility of evaluation criteria if we are to build software agents capable of interacting persuasively with humans through arguments. In this research, we concentrate on two important characteristic frameworks, *Reinstatement* and *Floating Reinstatement*, to understand such cognitive plausibility. Further studies on other characteristic defeat structures should help understand the full spectrum of all such cognitive plausibility.

We address the following research questions in this programme:

- Is the notion of argument reinstatement cognitively plausible?
- What effect, if any, does the presence of conflicting defenders (arguments which reinstate the main argument by defeating the defeater of the main argument) have on argument reinstatement?
- If reinstatement through conflicting defenders is cognitively plausible, what kind of commitment attitude is usually taken with respect to the defenders?
- Is there any difference between standard reinstatement and floating reinstatement?

1.4 Methodology

To answer the first question, we consider the scenarios of *Reinstatement*, which involve three arguments, an argument reinstating another argument by defeating the latter's defeater. In Figure 1.1, argument C reinstates argument A by defeating argument B. We explore empirically whether the main argument (argument A) overcomes its rejection due to the defeat by argument B, with the support of argument C. In other words, we monitor the changes in the acceptance of the main argument when it is alone, when it is attacked by argument B and when it is reinstated by argument C.

To answer the second question, we consider a more complex form of reinstatement, so-called *Floating Reinstatement* as shown in Figure 1.4. We explore empirically whether the main argument overcomes its rejection due to the defeat, with the support of two conflicting defenders. In other words, we explore empirically whether reinstatement in floating arguments follows what logicians refer to as credulous or sceptical attitude to evaluating the argument. We also try to differentiate a possible reinstatement in floating arguments by conflicting defenders from standard reinstatement.

To answer the third question, we again consider the floating argumentation framework. If reinstatement in floating arguments is cognitively plausible, then we check participants' preference (or lack thereof) between the conflicting defenders to explore what logicians refer to as

credulous preferred (choose one of the conflicting defenders over the other) or sceptical preferred (choose none of the conflicting defenders over the other).

The final question is answered by comparing the results from two groups of participants, one for standard reinstatement and another group for floating reinstatement using a mixed design repeated-measure analysis of variance to check the interaction between the problem and the type of reinstatement.

Hypotheses demanding psychological experiments are designed as the first step and then the experiments in the form of surveys focusing on the acceptance of the conclusion of arguments are conducted. Appropriate statistical significance tests like repeated measures of analysis of variance are run on the results to check the overall significance of the effects and then standard contrast analysis and regression analysis to check individual hypotheses. The results of the tests and the status of hypotheses are synthesized to arrive at the conclusions of the study.

1.5 Scope

As mentioned above, research questions were addressed only in the context of reinstatement and floating argumentation frameworks, mainly due to the time constraints. It is possible to extend the study by exploring other characteristic defeat structures of more number of arguments and defeat relations.

1.6 Contribution

The contribution of this dissertation includes the finding that the notion of reinstatement in abstract argumentation is cognitively plausible by supporting both grounded and preferred semantics and that the notion of floating reinstatement is cognitively plausible by not supporting the grounded semantics. The contribution also includes the finding that a floating reinstatement has an effect that is not significantly different from that of the standard form even though the extensions of their semantics are different. That is, the mutually conflicting nature of defenders does not play any role that undermines the job of reinstating the main argument. Study also finds that the lack of a significant interaction between the pattern and the reasoners' preference (or lack of preference) for one of the two mutually conflicting defenders in the tests indicating the existence of a different cognitively plausible notion that cannot be explained using the abstract argument evaluation criteria. The notion is that there is no clear endorsement for both credulous preferred and sceptical preferred semantics in floating arguments. Another contribution is the revealing of the notion of partial recovery in reinstatement. That is, only partial recovery is achieved in both scenarios of reinstatement and the idea of such a partial recovery is not dealt with in the abstract argumentation theory.

1.7 Organization of the Thesis

The rest of this thesis is organised as follows, Chapter 2 is dedicated to the literature review and the background of the dissertation discussing mainly nonmonotonic reasoning and human reasoning. In Chapter 3, we explore the cognitive plausibility of reinstatement by discussing hypotheses, experiment design, questionnaire design and the results obtained. Chapter 4 is dedicated to a study of the cognitive plausibility of floating reinstatement by discussing hypotheses, experiment design, questionnaire design and the results obtained. Chapter 5 discusses and summarises the results of the experiments whereas Chapter 6 concludes the study and lists the possible ways of extending it.

Chapter 2

Background and Literature Review

We begin the literature review with a brief introduction to the fundamentals of logic which are necessary background to the study.

2.1 Logic Fundamentals

We begin the discussion with the notion of *knowledge-base* [41]. Knowledge-base is a set of sentences expressed in a language called knowledge representation language. Knowledge-base can be considered as the central component of a *logic system*. The sentences in the knowledge base must follow the syntax of the knowledge representation language. Essentially, the logic must also define the *semantics* of the sentences. Semantic is the meaning of the sentences which are expressed in the knowledge representation language and it defines the truth of the sentences with respect to each *possible world*(also known as *model*).

A logic system essentially has to be involved in a process called *inference* [41]. Inference is the process of deriving new sentences from the old sentences, already existing in the knowledge-base. The notion of *entailment* in logic is important. While inference is deriving new sentences from the old sentences, entailment gives the idea that one sentence logically follows the other sentence(s). Entailment is written as $\alpha \models \beta$ (α entails β) and it means that in every possible world in which α is true, β is also true. We call an inference algorithm *sound* when only entailed sentences are derived by that inference algorithm. The soundness is also known as *truth-preserving*. The notion of *completeness* is also important to be mentioned. It gives the idea that any sentence derived from the inference algorithm is an entailed sentence.

Complex sentences are constructed from simpler sentences by using logical connectives. The commonly used connectives are $negation, and, or, implies$ and *bidirectional* [41]. The concept of *logical equivalence* is that two sentences are logically equivalent if they are true in the same set of models. One another important concept in logics is *satisfiability* and it means that a sentence can be called satisfiable if it is true in some model.

There are different inference rules that can be used to derive conclusions. These are called patterns of inference. For instance, one best rule is *Modus Ponens* as shown below.

$$\frac{\alpha \Rightarrow \beta, \alpha}{\beta}$$

Another useful inference rule is *And-Elimination*. It gives the idea that from a conjunction any conjuncts can be derived, as shown below.

$\underline{\alpha \wedge \beta}$

α

Monotonicity is one of the most important concepts in logic, particularly where this study is concerned. Monotonicity gives the idea that the set of entailed sentences can only increase as information is added to the knowledge-base [41]. That is, addition of new information does not make any already entailed sentences as no longer entailed. So we can write the following.

If $KB \models \alpha$ then, $KB \wedge \beta \models \alpha$ for any β

In other words, we can say that it does not affect entailing α that the addition of β to the KB if KB already entails α.

2.2 Nonmonotonic Logic

Common sense reasoning suggests that the failures of monotonicity are very widespread. Many a time, we jump to conclusions and then we correct our conclusions based on further information as it arrives. For instance, if we see an object in a red colour then we assume that the object is red. Later, if we come to know that the object was illuminated by a red colour, then we conclude that the object may not be red. That is, we retract our previous conclusion, the object is of a red colour. Reasoning of this kind is called as *nonmonotonic reasoning*. It is called as nonmonotonic because the set of entailed sentences does not grow monotonically over time as new observations are added to the knowledge base. Nonmonotonic logic is devised by modifying the notions of truth and entailment.

Following are some of the important concepts in nonmonotonic logic.

2.2.1 Circumscription

Circumscription is a formalized rule of conjecture which can be used along with the rules of inference of first order logic by a knowledge-based agent or a person to jump to certain conclusions [26]. In simple words, we can say that if we can show objects to have a certain property P by reasoning from certain facts A pertain to the objects, then we can say all those objects satisfy P [26]. Conclusions derived from circumscription are conjectures that A includes all the relevant facts and that the objects whose existence follows from A are all the relevant objects [26].

2.2.2 Default Reasoning

Default assumptions are made quite frequently in our everyday life about incompletely specified worlds. Such default assumptions are best viewed as beliefs which may well be modified or rejected by subsequent observations [39]. *Default reasoning* exists in various forms. One such form is to derive a plausible conclusion in the absence of any information to the contrary. For example, assuming that a bird can fly even though penguin cannot fly. It means, most birds fly except for penguins and ostriches. A great number of things that we know about a world are something that is almost true with few exception. Consider again the case birds fly except a few birds that cannot fly. Now, if we only know that the object is a bird, then it would be impossible to establish that the object can fly if the first order representation of *canfly* lists explicitly all the exceptions as given below [39]:

$\forall x\ bird(x) \wedge \neg penguin(x) \wedge \neg ostrich(x) \wedge ... \Rightarrow canfly(x)$

Default reasoning becomes useful now to assume that the bird can fly by default with the following interpretation "if x is a bird and it is consistent to assume that x can fly, then we infer that x can fly", represented formally as the default rule [39]:

$\underline{isabird(x) : Mcanfly(x)}$

$canfly(x)$

M is read as "it is consistent to assume".

The exceptions to $canfly$ are then given as standard first order representation [39]:

$\forall x\ penguin(x) \Rightarrow \neg canfly(x)$

$\forall x\ ostrich(x) \Rightarrow \neg canfly(x)$

...

2.2.3 Pollock's Defeasible Reasoning

Reasoning can be of two different kinds, *nondefeasible* and *defeasible*. Nondefeasible reasons are those reasons that can logically entail their conclusions. For example, everyone will agree that $P \wedge Q$ is a nondefeasible reason for P. Such reasoning is conclusive. On the other hand, defeasible reasoning also exists in this world. We call P a defeasible reason for Q just in case P is a reason for Q but adding additional information may destroy the reason connection. Defeasible reason can be called as "prima facie reasons".

Pollock [35] defines "prima facie reason" as follows:

Definition 1. *P is a prima facie reason for S to believe in Q if and only if P is a reason for S to believe Q and there is an R such that R is logically consistent with P but $(P\&R)$ is not a reason for S to believe Q.*

As mentioned above, the additional information may destroy the reason connection. Here the additional information R destroys the reason connection and we call R as a *defeater*. There are two kinds of defeaters, namely, *rebutting defeaters* and *undercutting defeaters*.

Pollock [35] defines *rebutting defeater* as follows:

Definition 2. *R is a rebutting defeater for P as a prima facie reason for Q if an only if R is a defeater and R is a reason for believing $\neg Q$.*

The additional information that the object (which is a bird) is a penguin is a rebutting defeater of the prima facie reason that if an object is a bird then it can fly because it is a reason to believe the bird in question cannot fly.

Pollock [35] defines *undercutting defeater* as

Definition 3. *R is an undercutting defeater for P as a prima facie reason for S to believe Q if and only if R is a defeater and R is a reason for denying that P wouldn't be true unless Q were true.*

An undercutting defeater of the reason the *object is red because it looked red* is that the *object was illuminated by a red light* because it attacks the connection between the premise *object looked red* and the conclusion *object is red*.

2.2.4 Loui's Defeasible Logic Programming

An early notable system for programming defeasible logic was devised by Loui [25]. The system presents nonmonotonic reasoning with defeasible rules. The system has the advantage of solving multiple extensions problem without additional explicit knowledge and also has the advantage of ordering competing extensions. The core of the system is the inference rules explicitly mentioning various types of defeats among arguments. Four different kinds of defeaters are used, namely, superior evidence, superior specificity, directness and preferred premise. The system is designed in such a way that more evidence is better than having more specificity, being more direct or having a preferred premise [25]. More evidence prevails over more specificity.

When any two of directness, specificity and preferred premise compete, neither prevails; the arguments interfere with each other [25]. In other words, directness, specificity and preferred premise share the same priority.

2.2.5 Abstract Argumentation

Some argumentation systems for defeasible reasoning assume particular logic, while there are some argumentation systems which leave underlying logic making room for instantiating the system with alternative logics. One such argumentation framework which leaves the internal structure altogether was proposed by Dung [17] allowing focus on the interaction of arguments. The frameworks proposed by Dung [17] plays a major role in this dissertation

In the theory of argumentation proposed by Dung [17], the framework is defined as a pair of a set of arguments and a binary relation representing the attack-relationship between arguments. As mentioned above, the internal structure of the argument is completely abstracted and hence the role of an argument is entirely determined by its relations to other arguments.

Formally, we can define Dung's theory of abstract argumentation as follows.

Definition 4. *An argumentation framework is a pair $AF = \langle AR, \rightharpoonup \rangle$, where AR is a set of arguments, and $\rightharpoonup$ is a binary relation on AR, that is, $\rightharpoonup \subseteq AR \times AR$. For two arguments A and B, the meaning of $\rightharpoonup (A, B)$ is that A attacks B.*

Consider an example with two arguments, Argument A and Argument B that attack each another. The corresponding framework will be $AF = \langle \{A, B\}, \{(A, B), (B, A)\} \rangle$. Often it is useful to visualise such argumentation frameworks using *argument graphs*, as shown in Figure 1.3.

Consider another example in which there are three arguments, Argument A, Argument B and Argument C. B attacks A and C attacks B. In this example, we say that C reinstates A. The framework for this is $AF = \langle \{A, B, C\}, \{(B, A), (C, B)\} \rangle$ as shown in Figure 1.1.

Let us see a few more definitions by Dung [17] before proceeding with the semantics of argumentation.

Definition 5. *A set S of arguments is said to be conflict-free if there are no arguments A,B in S such that A attacks B.*

Definition 6. *An argument $A \in AR$ is acceptable with respect to a set S of arguments if and only if for each argument $B \in AR$: if B attacks A then B is attacked by S*

Definition 7. *A conflict-freeset of arguments S is admissible if and only if each argument in S is acceptable with respect to S.*

The problem of whether we can consider an argument, interacting with other arguments in a framework is overall justified or not is usually handled using the concept of argument extensions. A particular semantics can be thought as a criterion for assigning status (e.g. accepted, rejected, or undecided) to arguments in a given argument graph, by appealing to the structural properties of that graph. The set of arguments accepted under a particular status assignment is called an *extension* of the argumentation framework. Thus semantics of abstract argumentation frameworks are often called *extension-based semantics*.

A variety of extension-based semantics have been studied in the literature, including Dung's original semantics [17] and many others [2]. In addition to identifying possible extensions, two acceptance *attitudes* have been studied in the literature. Under the *sceptical* attitude, an argument is accepted if it belongs to the intersection of all extensions defined by the given semantics. Under the *credulous* attitude, on the other hand, an argument is accepted if it belongs to at least one extension defined by the semantics used.

Some of the common extension-based semantics are defined below.

Complete Semantics

Definition 8. *Let (AR, $\rightharpoonup$) be an argumentation framework and Args be a conflict-free set of arguments. Args is said to be a complete extension if and only if Args = F(Args), where $F : 2^{AR} \rightarrow 2^{AR}$ such that $F(Args) = \{A | A$ is defended by Args$\}$.*

A complete extension is the set of arguments which exactly defends itself. We can find that the complete extensions of the framework shown in Figure 1.3 are {A}, {B}, ϕ. On the other hand, it can be seen that the set {A,C} is a complete extension of the framework shown in Figure 1.1.

Grounded Semantics

There could be more than one complete extension in a framework. However, grounded semantics is defined in such a way that there can be only one set as the grounded semantics. It is actually the smallest complete extension. It is also a conflict-free set of arguments.

Definition 9. *Let (AR, $\rightharpoonup$) be an argumentation framework. The grounded extension is the minimal fixpoint of F.*

We can find that the grounded extension of the framework shown in Figure 1.3 is ϕ. On the other hand, it can be seen that the set {A,C} is the grounded extension of the framework shown in Figure 1.1.

Preferred Semantics

Grounded semantics has the advantage that there exists only one grounded extension for one framework. However, the disadvantage is that it is a very sceptical approach and many times it could be simply ϕ. Preferred semantics is the credulous approach, in which we find extensions which are maximal from the point of admissibility.

Definition 10. *Let (AR, $\rightharpoonup$) be an argumentation framework and Args $\in$ AR. Args is said to be a preferred extension if and only if Args is a maximal (with respect to set-inclusion) admissible set.*

There are two preferred extensions in the framework shown in Figure 1.3. They are {A} and {B}. On the other hand, it can be seen that the set {A,C} is the preferred extension of the framework shown in Figure 1.1.

Stable Semantics

Stable semantics is one of the oldest semantics. The idea is to take only those arguments that defeat each argument which does not belong to the set. Stable semantics is defined as:

Definition 11. *Let (AR, $\rightharpoonup$) be an argumentation framework and Args $\in$ AR. Args is a stable extension if and only if $\{B \mid B \in AR$ and A defeats $B\} = AR \backslash Args$.*

The stable extension is conflict-free. It is also an admissible set. A stable extension is also maximal admissible set, and hence by definition a preferred extension. In Figure 1.3, two stable extensions exist, namely, {A} and {B} and in Figure 1.1 only one stable extension exits and that is {A,C}

Semi-Stable Semantics

The idea of semi-stable semantics is to find a complete extension such that the union of that extension and the set of arguments defeated by the complete extension is maximal.

Definition 12. *Let $(AR , \rightharpoonup)$ be an argumentation framework and $Args \in AR$. Args is said to be a semi-stable extension if and only if Args is a complete extension of which $Args \cup \{B \mid B \in AR$ and A defeats $B\}$ is maximal.*

In figure 1.3, there exist two semi-stable extensions, namely, {A} and {B}. In Figure 1.1 there exists only one semi-stable extension and that is {A,C}.

Semantics Compared

Figure 2.1 shows an overview of the semantics discussed above. A partial ordering of semantics can be seen in the figure, that is, every stable extension is a semi-stable extension, every semi-stable extension is a preferred extension, every preferred extension is a complete extension, and every grounded extension is a complete extension.

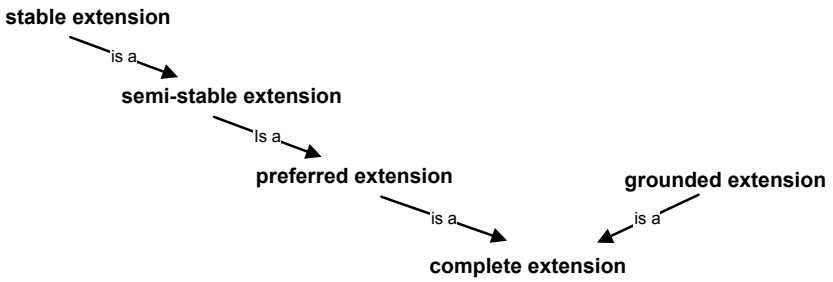

Figure 2.1: Semantics Compared [15]

2.3 Psychology of Reasoning

Human reasoning is often defined as the process of arriving at the conclusions when people solve problems or take decisions. The study of human reasoning involves at least psychology, philosophy, cognitive science, linguistics, artificial intelligence, logic and probability theory. Numerous studies have been conducted in human reasoning in order to clearly understand the reasoning process. Some of the research areas in human reasoning are given below:

1. Reasoning about sentences in natural language. The best example would be the classic Peter Wason's selection task. Peter Wason's selection task tests how people reason the conditional *if p then q*.
2. Relationships between induction, deduction, abduction and analogy from psychology perspective
3. Different approaches to human reasoning, such as mental logic, mental model approach, probabilistic approach and connectionist approach
4. Relating reasoning to intelligence
5. Relating pragmatics and emotion to reasoning
6. Neural correlation of reasoning

2.3.1 Some Types of Reasoning

Conditional Reasoning

Conditional reasoning or conditional inference is reasoning in which the human participant is provided with a conditional *if p then q* and one of four categorical premises p, $\neg p$, q, $\neg q$. Accordingly, we have four basic conditional reasoning arguments, out of which two are deductively valid and two are deductively invalid. We treat here $\rightarrow$ as the everyday conditional of natural language and $\supset$ as the material conditional of standard logic. The conditional reasoning arguments are given below:

Following are the deductively valid conditional reasoning arguments:

Modus Ponens (MP)

Using symbols we can represent Modus Ponens (MP) as given below

$\underline{\alpha \supset \beta, \alpha}$

β

An example is given below:

If it is raining, then it is wet outside

It is raining.

Therefore, it is wet outside

Modus Ponens is one of the conditionals used extensively in forming arguments.

Modus Tollens (MT)

Using symbols we can represent the Modus Tollens as given below

$\underline{\alpha \supset \beta, \neg\beta}$

$\neg\alpha$

An example is given below:

If it is raining, then it is wet outside

It is not wet outside

Therefore, it is not raining

Following are the deductively invalid conditional reasoning arguments:

Affirming the Consequent (AC)

Using symbols we can represent the Affirming the Consequent as given below

$\underline{\alpha \supset \beta, \beta}$

α

An example is given below:

If today is Monday, then I go to work

I go to work.

Therefore, today is Monday

Denying the Antecedent (DA)

Using symbols we can represent Denying the Antecedent (DA) as given below

$$\frac{\alpha \supset \beta, \neg\alpha}{\neg\beta}$$

An example is given below:

If today is Monday, then I go to work

Today is not Monday.

Therefore, I do not go to work

There are studies in the endorsement rates of these four conditional reasoning arguments. Studies have established that the empirical results obtained are substantially different from their standard logic equivalent, in case of MT, DA and AC. People endorse MP significantly more than MT and endorse DA and AC significantly above zero [31]. According to standard logic, participants should endorse MP and MT equally and they should not endorse DA and AC as they are deductively invalid. Also, there are inferential asymmetries among these arguments as observed by Oaksford [28]. Empirical results show that MP is endorsed more than MT, and AC is endorsed more than DA. They cannot be explained by standard logic.

Syllogistic Reasoning

Syllogistic reasoning is reasoning from a pair of premises, each containing a single quantifier. Following is an example of syllogistic reasoning:

All politicians are public speakers.
Some public speakers are paid speakers.
Therefore, some politicians are paid speakers.

Syllogisms were first analyzed by Aristotle [8]. Traditionally, the premises and conclusions of a syllogistic reasoning is one of the four moods [8]:

All X are Y
Some X are Y
No X are Y
Some X are not Y

For a valid deduction two premises must contain at least one term in common, the middle term, designated as B and other two terms designated as A and C should occur in a single premise. Therefore, the three terms (A, B and C) can be in of the four arrangements [8]

1. A-B, B-C
2. B-A, C-B
3. A-B, C-B
4. B-A, B-C

The conclusion is either A-C or C-A. Arrangements of the terms affect syllogistic reasoning and the cause of the effects is controversial [8]. Syllogism has been studied in Psychology for nearly 100 years but a consensus has yet to be reached.

Informal Argument Fallacies

Informal argument fallacies are arguments that are psychologically persuasive but logically incorrect [29]. There are disagreement on the definition of "fallacy". These fallacies have been

dubbed "informal" because it has not been possible to give "a general or synoptic account of the traditional fallacy material in formal terms" [19].

Following are some of the informal argument fallacies:

1. argumentum ad ignorantium (argument from ignorance)
2. petition principii (question begging)
3. ad hominem (arguments from authority)
4. slippery slop arguments

For instance, "Hosts exist because no one has proved that they do not" is an example of argument from ignorance. Walton [43] identified the following three types of arguments from ignorance where fallacies may arise:

1. Negative Evidence (for example, If drug A were toxic, it would produce toxic effects in legitimate tests. Drug A has not produced toxic effects in such tests. Therefore, A is not toxic)
2. Epistemic Closure (if a proposition is false - so its negation is true - because it cannot be proved from the contents of the database).
3. The Burden of Proof (illegitimate attempts to shift the burden of proof. As in the case of the Ghosts example given above, there is an illegitimate attempt to shift the burden of proof to other party instead of providing adequate reason for the existence of ghosts).

"There are more cars outside because there are fewer parking spaces available and less parking availability is due to more cars" is an example of question begging. This is equivalent to saying that "there are more cars outside, because there are more cars outside."

An example of ad hominem fallacy is given below:

Source A makes claim X
There is something objectionable about Source A
Therefore claim X is false

Slippery slop arguments are arguments that urge one to resist A, because allowing it could lead to B which is clearly objectionable [24]. For example "Legalizing cannabis will ultimately lead to increased use of cocaine or heroin, hence it should remain banned" [29].

2.3.2 Theories of Human Reasoning

Mental Logic

The Mental logic approach is one of the theories developed to explain how humans reason. Actually there are two mental-logic theories which have been developed [7]. One was developed by Braine and O'Brien and the other one developed by Rips [7]. They are similar in some aspects but significantly different in others. The theory of mental logic developed by Braine et al has three parts, namely, mental logic per se, a reasoning program and a pragmatic component as explained below.

The mental logic consists of a set of inference schemas like those proposed by Gentzen [7]. As an example, rules like p or q, not p, therefore q. The second part, the *reasoning program* applies to the inference schemas. Reasoning program includes direct reasoning routine as well as more sophisticated indirect-reasoning strategies. The direct reasoning routine is thought to be universally available and is applied with minimal effort in reasoning as well as in comprehension. The basic prediction of the theory is that inferences that are made by the application of basic

schemas with the direct-reasoning routine will be made routinely. The pragmatic part of the theory says that that the basic meaning of each logic particle is in the inferences that are sanctioned by the inference schemas [7].

Rip's PSYCOP theory is another mental logic theory. The name PSYCOP comes from PSYchology Of Proof. According to Rips [40]:

A person faced with a task involving deduction, attempts to carry it out through a series of steps that take him or her from an initial description of the problem to its solution. These intermediate steps are licensed by mental inferences rules, such as modus ponens, whose output people find intuitively obvious. The resulting structure thus provides a conceptual bridge between the problem's "givens" and its solution.

Thus inference rules are used in constructing and verifying mental proofs. However, the attempts at the proofs may not be always successful but an attempt is always made. An error could be caused by anything that hinders the application of an appropriate inference rule like working memory restrictions, lack of a requisite rule, or possession of a non-standard rule.

There are studies on whether humans actually follow rules in reasoning. There are a number of controversial theoretical positions in psychology which state that the rule based reasoning is not involved in human reasoning or its involvement is very minimal. Smith, Langston and Nisbett [42] proposed eight criteria to determine whether or not humans use abstract rules in their reasoning. They argue that humans use abstract rules in reasoning quite substantially with evidence. The inferential rules such as modus ponens, contractual rules, causal rules and the law of large numbers are investigated using the eight criteria defined to check whether humans use abstract rules in their reasoning.

The criteria used are given below [42]:

Criteria Stemming from Linguistics

1. Performance on rule-governed items is as accurate with unfamiliar as with familiar material
2. Performance on rule-governed items is as accurate with abstract as with concrete material
3. Early in acquisition, a rule may be applied to an exception (the rule is overextended)

Performance Criteria

4. Performance on a rule-governed item or problem deteriorates as a function of the number of rules that are required for solving the problem
5. Performance on a rule-governed item is facilitated when preceded by another item based on the same rule (application of a rule primes its subsequent use)
6. A rule, or components of it, may be mentioned in a verbal protocol.

Training Criteria

7. Performance on a specific rule-governed problem is improved by training on abstract versions of the rule.
8. Performance on problems in a particular domain is improved as much by training on problems outside the domain as on problems within it, as long as the problems are based on the same rule.

As an example, in discussing the evidence about Modus Ponens, Smith, Langston and Nisbett find that criterion 1 supports the hypothesis that humans use modus ponens and suggest that humans can perform extremely well with Modus Ponens in unfamiliar scenarios [42].

There are studies on how humans reason and their relations with the standard logic (modern prepositional logic and first-order predicate logic) of reasoning. There have been attempts to create logic systems that closely resemble human reasoning. For instance, Braine [6] presents

a system of logic as the logic that people use in propositional reasoning. The logic proposed is guided by two conceptions: a) inference rule schemata (inference rule schema is a formula to define inference rules by specifying their general form) should be preferred to axiom schemata and b) the connectives (such as $if-then, and, or, some$) should capture essential semantic and syntactic properties of the corresponding English words. The connectives form the basis of the logic presented. The logic system defines the place of logic within the natural reasoning process and specifies a set of natural connectives closely corresponding to the meaning of English conjunctions.

How the theory proposed by Braine [6] treats $if-then$ connective requires special mention. According to the theory, natural logic has no connective equivalent with the truth-functional conditional as far as $if-then$ is concerned. $If-then$ in the natural language corresponds to the inference line, in other words, the inference line and $if-then$ are two different notations for indicating the same relationship.

For instance, assume that we have the following inference rule:

$\underline{\alpha}$

β

That is, if α is established then β can be concluded. According to the theorem *if* α *then* β is exactly the same as the above inference rule irrespective what makes α and β. If we apply this inference line theory to Modus Ponens (*if p then q,* $p \setminus q$), one can easily see that the theory fits well with the responses from the subjects. It can be noted that if the second premise is not a p, then inference line theory does not say what to do and the response of the subjects depends upon whether the subject works out the minimum commitments of the major premise(*if p then q*) [6].

Another attempt at building a logic system of plausible reasoning is from Collins and Michalski [16]. They try to develop a formal system based on variable-valued logic calculus characterizing different patterns of plausible inference that humans apply in reasoning [16]. The experiments are conducted as responses to questions for which the participants do not have ready answers with the core objective of discovering recurring patterns that humans use in reasoning and to determine parameters which affect the certainty of these inferences [16]. They believe that there are high payoffs from trying to formalize the patterns of human plausible reasoning. They argue that it helps identify parallels among apparently different inference types and it also makes it possible to see the systematic patterns in which different certainty factors in the psychological literature (e.g., typically, similarity, frequency, dominance) related inference types [16].

Mental Models

Mental model theory states that any kind of reasoning consists of five main processes, namely, construction of mental models out of the premises, integration of the mental models constructed, formulation of conclusions which are in consistent with the integrated models, falsification of conclusions and the production of responses such as linguistic or motor responses [1]. The idea that humans rely on mental models can be traced to Craik's suggestion in 1943 that the mind constructs a "small-scale model" of reality that it uses to anticipate events. Mental models represent the surrounding world. It also represents the relationship between various parts of this surrounding world in the perception of the subject, the way the subject going to behave in this world and the consequences his or her behaviour. So mental models can be considered as internal representation of external reality by saving all perceived stimulations and observation of the subject.

Byrne and Johnson-Laird argue that humans rely on constructing and manipulating mental models in the reasoning process [12]. In this process of manipulating, subjects come up with conclusions which are compatible with the models that are integrated and a deliberate process of trying to falsify the conclusions by constructing alternative models of the premises and finally producing responses such as linguistic or motor responses. The larger the number of

number of models compatible with the problem, the greater the difficulty [12]. On the other hand, according to mental logic approach advocated by Rips [40] and Braine and O'Brien [7] the difficulty of a problem depends on the number of inference rules used and the difficulty weight of inference rule. Following are some of the experiments conducted to prove the validity of the mental model theory in various areas of human reasoning such conditional reasoning, propositional reasoning and fallacies.

Byrne and Johnson-Laird argue that the results obtained in the experiments conducted in [12] support the prediction that the larger the number of models compatible, the greater the difficulty as opposed to the more inferences rules, the greater difficulty.

In another experiment conducted by Byrne [9] on the everyday reasoning with conditional sequences such as "if she leaves home early, she reaches office early", "if she reaches office early, then she gets extra things done" proves wrong the common assumption that people make same amount of inferences in conditional sequences as they do on simple arguments that contain just a single conditional premise. They argue that the number of inferences that people make in everyday conditional reasoning sequences depends on the representation of state of affairs that sequences describe.

According to Byrne and Johnson-Laird, results from Wason's selection task makes sense with the mental model approach [13] stating that individuals will select those cards corresponding to a counterexample only if they construct its explicit model. The particular selection made will depend on the particular counterexample selected by the reasoner, especially in deontic domains, the premises are open to bi-directional interpretation and thus to have two set of counterexamples [13]. They argue that pragmatic reasoning schemas are unable to explain the results obtained from experiments outside the deontic domains.

The Mental model approach has been studied in many cases such as the ones explained before. The approach has been tested in multiply-quantifier (involving more than one quantifiers such as *all*, *some*, *any*, *some* and so on) problems by Byrne, Johnson-Laird and Tabossi [22]. They argue that the reasoner constructs mental models based on their knowledge of the meaning of these quantifiers and support the prediction that that one model problem will be easier than those problems which require more than one model.

In an other study Bara, Bucciarelli and Lombardo [1], provide a single set of procedures to unify the main types of deductive reasoning in humans such as syllogistic, propositional and relational reasoning. In other words, they produce a single unified theory with a formal foundation by bringing together the micro theories developed from mental models. The theory is tested using a computational model (UNICORE, which stands for UNIfied COmputational REasoner) and the model is tested against the experimental subjects to a satisfactory level reproducing both correct and erroneous performance of the subjects. Finally the theory confirms the mental model approach for deductive reasoning.

In a similar study, Bucciarelli and Johnson-Laird [8] present a computer model that implements syllogism based on mental models. They tested the model with 4 experiments on human subjects. The experiments revealed that the humans participants not trained in logic construct counter examples as required by the experiments using similar operations to those implemented in the computer model. However, humans rely on a much greater variety of interpretations of premises and of search strategies compared to the computer model.

Johnson-Laird, Bynre and Schaeken proposed a mental model based theory of propositional reasoning as a semantic process. It is interesting to note that the disjuncts are initially represented as imaging two alternatives models and this can, if necessary, be fleshed out to yield an explicit representation of an exclusive or an inclusive disjunction [23]. There are crucial differences between the mental model approach and the rule based (mental logic) approach as far as propositional reasoning is concerned [21]. Johnson-Laird, Bynre and Schaeken argue that mental model theory provides a decision procedure whereas rule based theories do not. They also argue that the rule theories have fewer domains of reasoning compared to mental theories [21].

Probabilistic Approach

While mental logic and mental models stress deductive relationships the probabilistic approach of human reasoning stresses on the inductive strength of the argument [31]. Oaksford and Hahn argue that humans are more sensitive to inductive strength (also known as argument strength) of the conditional than deductive correctness [31]. The following example clearly shows what argument strength is. If the inductive strength of the argument *if x is a bird, then x flies* = 0.9 and *P(Tweety is a bird) = 1*, then the inductive strength of the conclusion that *Tweety can fly* is 0.9. They interpret that the low inductive strength of modus tollens is due to the fact that proportion of people endorsing an inference is a reflection of the underlying degree of belief that people have about the conditional [31]. They also argue that people are not particularly sensitive to deductive relations is because they are monotonic.

However, Rips argues that trying to make inductive strength does service for deductive correctness and trying to make deductive correctness does service for inductive strength. Oaksford and Hahn argue that this argument is misguided [31]. They point out that if deductive correctness and inductive strength are psychologically one and the same thing, then deductive correctness should track inductive strength. They produce empirical results which show inductive strength and deductive correctness do not track each other [31].

Oaksford and Charter argue that the conditional inference based on inductive strength is compatible with the dynamic view of inference in probability logic [27]. The inference forms include apart from modus pones and modus tollens, two deductively invalid arguments, namely, denying the antecedent and affirming the consequent. These are defined elsewhere in the document.

Oaksford and Charter also propose similar ideas for addressing informal argument fallacies, such as the argument from ignorance, circularity, and the slippery slop argument [32]. They argue that invoking the concept of argument or inductive strength can resolve the paradox of why some instances of informal argument fallacies seem perfectly acceptable and capable of rationally persuading someone of a conclusion. They propose a Bayesian theory of content-dependent argument strength and possible psychological mechanisms of that are identified [32]. Finally they argue that the study on fallacies based on Bayesian theory can be extended to more general human activity of argumentation.

Ferguson argued that the defeasible nature of most everyday informal reasoning might be accommodated without the need to invoke non-monotonic reasoning procedures [30].There are two components to Ferguson's proposal. First, a new rule of inference as a special case of standard modus ponens. Second, he argues that most cases where non-monotonic behaviour is observed are cases where an inference is unsound rather than invalid. However, Oaksford and Hahn counter argue that the conclusions derived from this inference are vacuous, requires a procedural "hack" to break the symmetry between the disjuncts of the tautological conclusions to which the proposal leads [30]. They also argue that Ferguson's proposal amounts to arguing that all everyday inferences are sound by definition [30]. In summary, they argue that Ferguson's proposal to reconcile monotonic logic with defeasibility has some very counterintuitive consequences [30].

2.3.3 Suppression of Conclusion in Conditional Reasoning

Psychological experiments have firmly established the fact that sound inferences can be defeated by the provision of additional information, that is, the conclusion that was arrived previously can be invalidated with the addition of new information to the premise. A number of research have established something called Core Pattern of results, that two conditional arguments, namely, *modus ponens* and *modus tollens* that are deductively valid are defeated by disabling conditions (or rebutters) and the two conditional arguments, namely, *Affirming the Consequent* and *Denying the Antecedent* that are deductively invalid are defeated by alternative causes (or alternative antecedents or undercutters) [34].

This can be explained by using examples of the notions of 'disabling conditions' and 'alternative causes'. Assume that we have the conditional statement "If he is hungry, then he eats food" and the categorical premise "He is hungry". This is an example of modus ponens conditional. We can find that the endorsement ratings of the conclusion "he eats food" should be very high. Now assume that we are also provided with another conditional statement "If food is available, then he eats the food." This statement is a cue to the existence of a disabling condition, that is, food is not available. We can find that the endorsement ratings in the conclusion of "he eats food" reduces with this disabling condition.

Similarly, let us take the case of Affirming the Consequent. Assume that we have the conditional statement "If she wants to reach office early, then she leaves home early" and the categorical premise "she leaves home early". Now assume that we are also provided with another conditional statement "If she wants to pick her colleague up, then she leaves home early". This is an 'alternative cause' of leaving home early and thus reduces the rate of endorsement in the conclusion "she wants to reach office early".

The results of the experiments conducted by Politzer and Bonnefon show that Cummin's observation that one can reverse the core pattern of defeasible inferences in causal conditional reasoning is correct [34]. Cummin's observation was in the covert paradigm and in causal domain. Politzer and Bonnefon extended the result to the overt paradigm. They also tested this reversion beyond the causal domain namely for means-end rules, and found that this is not limited to causal. As explained recently by Politzer and Bonnefon, "One can conclude that Cummin's observation seems to be but phenomenon that finds an explanation in Pollock's view of defeaters, a conceptualization that extends beyond the causal domain. At a descriptive this means that people are responsive to logical necessity and sufficiency as well as to causal necessity and sufficiency. Like disablers and alternatives for causal relationships, rebutters and undercutters modulate the degree of necessity and sufficiency of various relationships, including relations of causality" [34].

Byrne, Espino and Santamaria in their studies of counter examples and suppression of inferences [11] find that inferences are suppressed even when they produce their own conclusions. Their experiments also show that suppression occurs even when the premises are expanded to rule out the opposite and the inferences are suppressed more when the background conditions are expressed in a biconditional. Finally they prove that the fallacy inferences(affirming the consequent and denying the antecedent) can be suppressed when the premises refer to alternatives as well as to additional conditions. In summary they suggest that the suppression rests on the availability of counterexamples [11].

It is interesting to note, as explained by Politzer and Bonnefon, that "according to the mental model approach, the key factor is the existence of a counter example to the conclusion of the argument; but it has been shown that this approach cannot explain participants' responses in the variant of the overt paradigm that consists of a production task. Even more crucially it has been argued that it cannot explain participants' systematic allocation of degree of belief to the conclusion as a function of the credibility of the premises in both paradigm" [34]. Also it is interesting to note the findings by Politzer and Bonnefon, that "it is therefore no wonder that the user of two fundamental arguments by which people draw inferences to particular propositions namely MP and AC can be linguistically signalled to an interlocutor by a special divide, which what the two varieties of inferential conditionals consist in. That is, an inferential conditional can be put to two uses, both of which exploit the same general fundamental property of conditional sentences, namely license modus ponens: One (in the necessary-epistemic case) to make a deductive inference by a 'compacted' MP (if cause then effect, where cause is affirmed and effect is inferred); the other, more subtly (in the possible-epistemic case), to make an adductive inference by a hidden compacted AC argument in the guise of an MP argument (if effect then cause, where effect is affirmed and cause is inferred), which is made possible by the fundamental property just reminded. Thus, human adductive reasoning can be performed by the exploitation of a formally entirely deductive derivation, but one in which a premise is uncertain, that is, a plausible reasoning process (defined as the use of a deductively valid argument in which at least one premise is uncertain)" [34].

Byrne argues that the experiments conducted invalidate the argument of proponents of mental logic that suppression relies on inconsistent and incoherent premises [10].

Ford in her studies of human nonmonotonic reasoning argues that people find monotonic reasoning difficult [18]. She also argues that recognizing differences in the logical strengths of arguments is important for nonmonotonic problems. Ford finds that in some problems in which specificity was irrelevant instead of deriving a 'can't tell' response, participants could give rational conclusions based on differences in the logical consequences of arguments. In other problems, where specificity was relevant, studies suggest that specificity is not paramount. Finally the studies show that the ability to appreciate the differences in the logical strength of simple, non-conflicting statements is what matters in participants' success at responding to nonmonotonic problems involving conflict [18].

Pelletier and Elio argue that unlike many other reasoning formalisms, nonmonotonic or default reasoning is "psychologistic", meaning that it is defined as what people do in circumstances when they are engaged in commonsense reasoning [33]. They also argue that nonmonotonic reasoning should not be discovered by looking at mathematical systems or not by formal considerations such as 'simplicity' or 'computability' and not by the intuitions of the researchers or their fellows [33].

2.3.4 Reinstatement and floating conclusions

The floating conclusions phenomenon was first investigated in the context of defeasible inheritance by Thomason, Touretzky and Horty [20]. A floating conclusion can be explained using a well-known example provided by Ginsberg as shown in Figure 2.2. Here the statements Qn, Rn, Dn, Hn and En represent the propositions that Nixon is Quaker, Nixon is a Republican, Nixon is a Hawk, Nixon is a Dove and Nixon is politically extreme respectively. Statement of the form $A \rightarrow B$ represents logical implication, $A \dashrightarrow B$ represents defeasible implication and T represents truth. What the diagram tells us is that Nixon is both a Quaker and a Republican and hence there is a good possibility that Nixon is Dove (if he is a Quaker) and Nixon is a Hawk (if he is Republican). In either case, Nixon is politically extreme. That is, the conclusion "Nixon is politically extreme" 'floats' above mutually conflicting arguments "Nixon is a Dove" and "Nixon is a Hawk". The investigation by Thomsason, Touretzky and Horty was in connection with the theory developed based on a single argument extension as a contrast to multiple-extension accounts and was thought of as containing the "sceptically acceptable" arguments [20]. However the political extremist example given by Ginsberg showed that the single argument extension cannot contain the "sceptically acceptable" arguments and thereby the sceptical conclusions defined by those sceptically acceptable arguments. Studies by Makinson, Schlechta and Stein also consider floating conclusions in the context of defeasible inheritance reasoning and they support the view of Ginsberg that the conclusions to be derived from the knowledge base are those that are supported by each of its argument extensions [20].

Bonnefon in his studies of reinstatement, floating conclusions and the credulity of mental model reasoning, argues that the mental model theory (MMT) of propositional logic is somewhere between "credulous" and "sceptical" with respect to conclusions that the theory draws on default reasoning problems [5]. He claims that that this is supported by the consideration two of these situations, namely problems involving reinstatement and floating conclusions [5]. This study is important from the perspective that the mental model theory has been criticised in the past for its inability to explain defeasible reasoning, even in some cases to give wrong answers to defeasible reasoning problem. This criticism is countered by proponents of mental model theory by pointing out that MMT may provide a more psychological approach to defeasible reasoning [14]. Interestingly, Bonnefon suggests that, "Now if indeed the mild credulity of MMT is seen as a problem to be fixed, one may consider two routes thereto: (a) to change the way MMT derives conclusions in order to make it more sceptical; (b) to refine the pragmatic dimension of the theory so that it is able to differentiate between situations where floating conclusions or reinstated arguments are acceptable and situations where they are not."

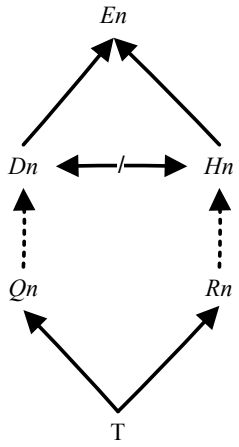

Figure 2.2: Floating Conclusion [20]

Chapter 3

The Cognitive Plausibility of Reinstatement

In this part of the research we intend to study the cognitive plausibility of argument evaluation criteria for two characteristics argumentation frameworks, the simplest framework involving a defeat, $AF = \langle\{A, B\}, \{(B, A)\}\rangle$ in which argument B attacks argument A and the reinstatement 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.

3.1 Hypotheses

There are three hypotheses to test. The hypothesis $H1$ is on the first framework whereas other hypotheses are on the reinstatement framework. Hypotheses $H1$, $H2$ and $H3$ will test the change in confidence when attacking and reinstating arguments are presented to the subject in the specified order.

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.

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 by argument C.

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 need to test another hypothesis, hypothesis $H4$. This hypothesis is to make sure that in a reinstatement scenario, the attacking argument is 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 3.1.

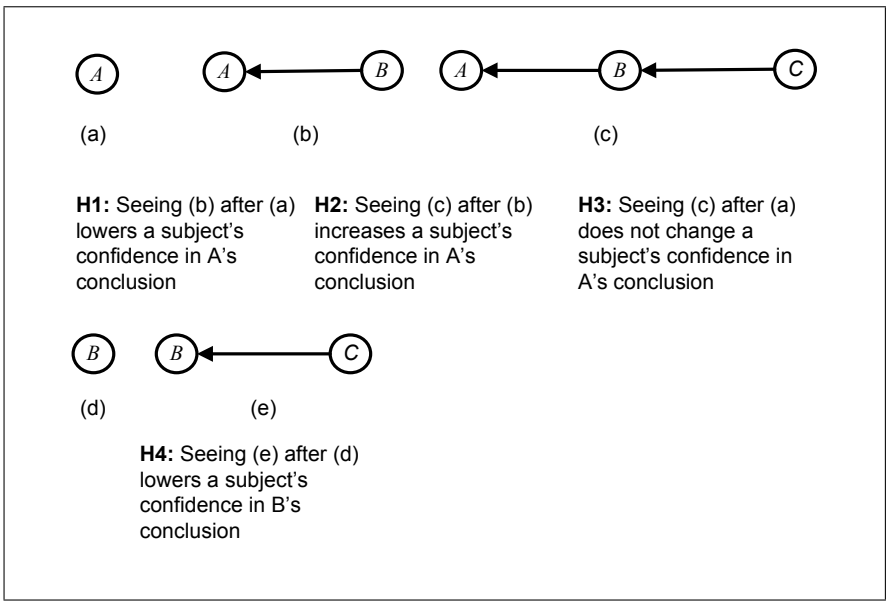

Figure 3.1: Summary of Hypotheses - Reinstatement

3.2 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 A, $A \leftarrow B$, $A \leftarrow B \leftarrow C$, B, $B \leftarrow C$.

Natural language arguments are used to instantiate the independent variable.

The dependent variable (DV) is the confidence marked by the participant in the conclusion of arguments, argument A and argument B on a given interval scale. We used the following scale as used by Politzer and Bonnefon [34] to mark the confidence:

1 means the conclusion is certainly false,
2 means that it is much more likely to be false than to be true,
3 means that it is slightly more likely to be false than to be true,
4 means that it is as likely to be false as to be true,
5 means that it is slightly more likely to be true than to be false,
6 means that it is much more likely to be true than to be false, and
7 means that it is certainly true.

A visual representation of the scale with BLACK representing false and WHITE representing true is given in Figure 3.2

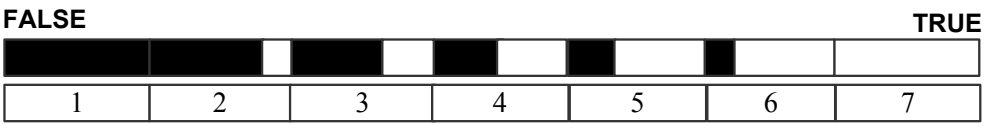

Figure 3.2: Visualization of the Scale

3.3 Experiment Location & Management

The survey was conducted using questionnaires in paper form rather than interviews or telephone surveys. Random sample of people from offices, shopping malls and open spaces in Dubai, UAE were surveyed.

3.4 Participants Groups

There were two participant groups in this survey. These groups were labelled $G1$ and $G2$. Group $G1$ had 20 participants and group $G2$ had eighteen participants totalling 38 participants in the survey. Group $G1$ tested hypothesis $H1$, $H2$ and $H3$ whereas group $G2$ tested hypothesis $H4$. Following were the IV value(s) assigned to each group. For group $G1$, there were three assignments and for group $G2$, there were two assignments. The first assignment is in condition 1, the second assignment in condition 2 and the third in condition 3.

$$G1 : A, \quad A \leftarrow B, \ A \leftarrow B \leftarrow C$$
$$G2 : B, \ B \leftarrow C$$

3.5 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 should have taken 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, in the given scale.

All questions were of the type 'what'. There was no question of type 'why'. The questionnaire was scrutinized to avoid double-barrelled questions (questions with more than one question embedded within it) and leading questions (questions that suggest the answer or answer that is intended). 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. Each set of arguments had 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 problems in each argument set was fixed and corresponded to the order of the IV assignments $A, \quad A \leftarrow B, \quad A \leftarrow B \leftarrow C$. 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.

Group $G2$ used the same argument sets that were used for group $G1$. Group $G2$ tested 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. The participants were not allowed to see later problems before answering the previous problems.

The following personal information about the participants was collected at the end of the survey:

- gender (male, female)
- education level (high school, college diploma, college bachelor, master's degree, doctoral)

- nationality
- age (18-21, 22-25,26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70, above 70)
- proficiency in English language (Expert, Very Good, Good, Competent, Modest, Limited, Extremely Limited, Intermittent)

The experiments were conducted twice (pilot and final). Appendix A describes the pilot survey. A final survey was conducted with an entirely new set of arguments and participants. The following were the argument sets used in the final survey:

Argument Set 1

- Argument A : "The battery of Alex's car is not working. Therefore, Alex's car will halt."
- Argument B : "The battery of Alex's car has just been changed today. Therefore, the battery of Alex's car is working."
- Argument C : "The garage was closed today. Therefore, the battery of Alex's car has not been changed today."

Argument Set 2

- Argument A : "Louis applied the brake and the brake was not faulty. Therefore, the car slowed down."
- Argument B : "The brake fluid was empty. Therefore, the brake was faulty."
- Argument C : "The car had just undergone maintenance service. Therefore, the brake fluid was not empty."

Argument Set 3

- Argument A : "Mary does not moderate her phone usage. Therefore, Mary has a large phone bill."
- Argument B : "Mary has a speech disorder. Therefore, Mary moderates her phone usage."
- Argument C : "Mary is a singer. Therefore, Mary does not have a speech disorder."

Argument Set 4

- Argument A : "John has no way to know Leila's password. Therefore, Leila's emails are secured from John."
- Argument B : "Leila's secret question is very easy to answer. Therefore, John has a way to know Leila's password."
- Argument 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

- Argument A : "Mike's laptop does not have anti-virus software installed. Therefore, Mike's laptop is vulnerable to computer viruses."
- Argument B : "Nowadays anti-virus software is always available by default on purchase. Therefore, Mike's laptop has anti-virus software."
- Argument C : "Some laptops are very cheap and have minimal software. Therefore, anti-virus software is not always available by default."

Argument Set 6

- Argument A : "There is no electricity in the house. Therefore, all lights in the house are off."
- Argument B : "There is a working portable generator in the house. Therefore, there is electricity in the house."
- Argument C : "The fuel tank of the portable generator is empty. Therefore, the portable generator is not working."

3.6 Hypothesis Analysis

3.6.1 Hypotheses $H1$ through $H3$

$H1$, $H2$, 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.

3.6.2 Hypothesis $H4$

$H4$ 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.

$H1$, $H2$ and $H3$ are one-tailed and $H3$ is two-tailed.

3.7 Results

We found that the base (when argument A is presented alone) acceptance rating of the conclusion when we average the scores across the 6 contents as 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 defeater* acceptance ratings of defeaters was 5.1 (SD = 0.8) and the acceptance ratings for the *attacked defeater* 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.

Chapter 4

The Cognitive Plausibility of Floating Reinstatement

The framework $AF = \langle\{A, B, C, D\}, \{(D, C), (C, D), (D, B), (C, B), (B, A)\}\rangle$ as shown in Figure 4.1 is studied in this part of the experiment. The framework contains four arguments. Arguments C and D attack each other, in addition to attacking argument B. However, argument B does not attack argument C or argument D. Argument B attacks argument A. The framework is interesting in the sense that someone can argue that argument B has to be overruled without resolving the conflict between argument C and argument D. If argument B can be overruled, then argument A has to be justified. In other words, it says that the mutually attacking arguments C and D reinstate argument A. Below is an example for this argumentation framework.

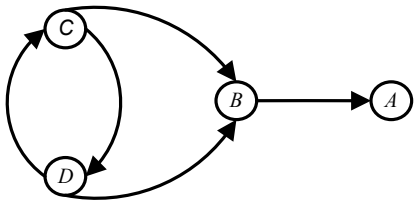

Figure 4.1: Floating Arguments

- Argument A : "Mary does not spend money on phone calls. Therefore, Mary is not a profitable phone customer."
- Argument B : "Mary makes a lot of international phone calls. Therefore, Mary has a huge telephone bill"
- Argument C : "Mary is a minor and uses her parents' phone a lot. Therefore Mary makes a lot of international phone calls on her parents' account"
- Argument D : "Mary uses her husband's phone a lot. Therefore, Mary makes a lot of international phone calls on her husband's account"

We can see that argument B attacks argument A because having a huge telephone bill invalidates the argument that Mary does not spend money on telephone calls. However, argument C and argument D are undercutting defeaters of argument B because it tells us that making a lot of international phone calls does not necessarily produce a huge telephone bill for a caller's

account. We can also see that argument C and D are in conflict with each other in premise and conclusion of both arguments. In premise, we see a married person cannot be a minor. In the conclusion, we see that in argument C, Mary makes phone calls on her parents' account whereas in argument D, these calls are booked under her husband's account.

We need to explore various argument based semantics and corresponding extensions of this framework to understand what objectives are to be achieved in this study. These semantics are defined elsewhere in this paper.

The complete extensions of the framework are $\{C, A\}$, $\{D, A\}$ and $\emptyset$. The grounded extension of the framework is $\emptyset$. Now let us consider preferred extensions. There are two preferred extensions and they are $\{C, A\}$ and $\{D, A\}$. The Stable extensions of this framework are again $\{C, A\}$ and $\{D, A\}$. The semi-stable extensions are the same as the stable extensions.

Following are the objectives of this study:

1. We would like to explore the cognitive plausibility of reinstatement in a floating arguments framework. We know that the reinstatement works in the framework $AF = \langle\{A, B, C\}, \{(B, A), (C, B)\}\rangle$. If reinstatement works in the floating arguments framework, then it means that people tend to follow preferred semantics rather than grounded semantics in the case of floating arguments.

2. If reinstatement works in a floating arguments framework, then we need to explore whether reinstatement in a floating arguments framework is weaker than in the straightforward reinstatement framework $AF = \langle\{A, B, C\}, \{(B, A), (C, B)\}\rangle$.

3. If reinstatement works in a floating arguments framework, then we need to explore the reasoners' preference (or lack thereof), whether argument C and D. This will reveal whether reasoners prefer *credulous preference* to *sceptical preference* or not.

4.1 Hypotheses

This part of the study tests the following hypotheses on the floating arguments framework. These entire hypotheses test acceptance of the conclusion of a single argument (argument A). The following are the hypotheses:

H1: The confidence in the conclusion of an argument (argument A) rises significantly when it is attacked by argument B but reinstated by two mutually attacking arguments (argument C and argument D), from the level where it was attacked by another argument (argument B) but not reinstated by argument C and argument D. That is, we hypothesize that reinstatement works in the case of floating arguments.

H2: The confidence in the conclusion of an argument (argument A) when it is attacked by another argument (argument B) but reinstated by two mutually attacking arguments (argument C and argument D) is significantly lower than the confidence in the conclusion of an argument (argument A) when it is attacked by another argument (argument B) but reinstated by another argument (argument C). That is, we hypothesize that the reinstatement in the case of floating arguments is not as strong as in the straightforward reinstatement framework.

The hypotheses are summarised in Figure 4.2.

4.2 Dependent And Independent Variables

The independent variable (IV) in the experiment is the argumentation framework presented to the participant. Different values for this independent variable are:

- A

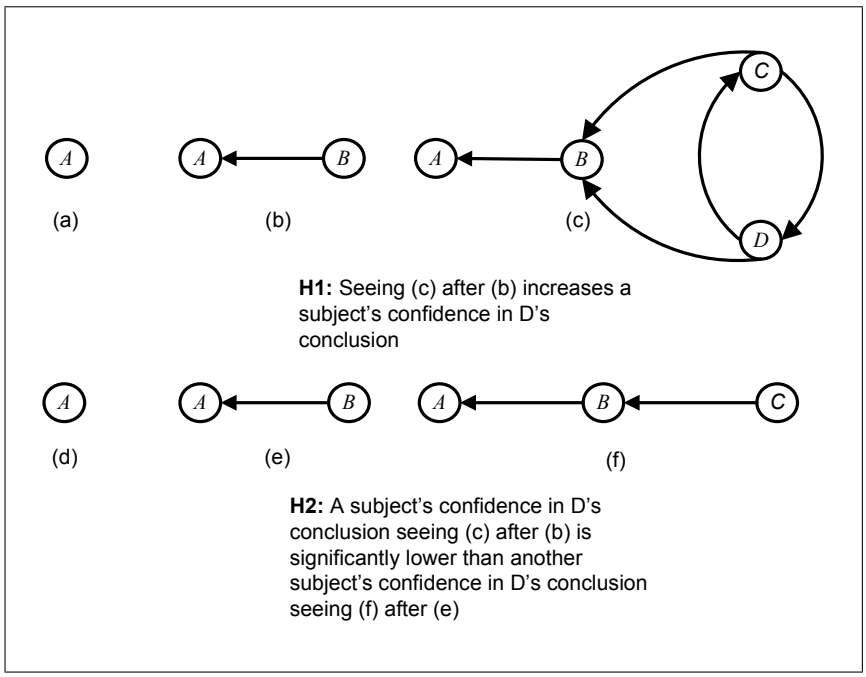

Figure 4.2: Summary of Hypotheses - Floating Arguments

- $A \leftarrow B$
- $A \leftarrow B$,$B \leftarrow C$
- $A \leftarrow B$, $B \leftarrow C$, $B \leftarrow D$, $C \leftrightarrow D$

In the experiments natural language arguments are used to instantiate the independent variable.

The important dependent variable (DV) is the confidence marked by the participant in the conclusion of argument A on a given interval scale. We used the following scale as used by Politzer and Bonnefon [34] to mark the confidence:

1 means the conclusion is certainly false,
2 means that it is much more likely to be false than to be true,
3 means that it is slightly more likely to be false than to be true,
4 means that it is as likely to be false as to be true,
5 means that it is slightly more likely to be true than to be false,
6 means that it is much more likely to be true than to be false, and
7 means that it is certainly true.

In addition to the dependent variable of confidence in the conclusion of argument A, the participants marked their answers on two other dependent variables. The first one was "How clearly did you understand the problem?". The scale was again an interval scale from 1 to 7 as shown below:

1 - Not at all
2 - Very little
3 - Little
4 - Somewhat
5 - Well
6 - Very well

7 - Completely

Second one was to record reasoners' preference (or lack thereof) between arguments C and D. The question was framed as "Do you think that:" and measured using the nominal scale:

1 - C is a better argument than D
2 - D is a better argument than C
3 - C and D are about equally good

4.3 Experiment Type & Management

The survey involved two groups of participants and it was conducted using questionnaires in paper form rather than interviews or telephone surveys. The survey was conducted by drawing random sample of people from offices, shopping malls and open spaces in Dubai.

4.4 Participants Groups

There were two groups of participants in the survey. Each group contained a minimum of 20 participants. These groups were labelled $G1$ and $G2$. Group $G1$ tested the confidence in the conclusion of argument A, when it was presented alone, when it was presented as argument B attacking it and when it was presented as argument B attacking it but reinstated by mutually attacking argument C and argument D. Group $G2$ tested the confidence in the conclusion of argument A, when it was presented alone, when it was presented as argument B attacking it and when it was presented as argument B attacking it but reinstated by argument C.

The following were the IV value(s) assignments. A semicolon separates the values of the independent variable.

$G1 : A; A \leftarrow B; A \leftarrow B, B \leftarrow C, B \leftarrow D, C \leftrightarrow D$
$G2 : A; A \leftarrow B; A \leftarrow B, B \leftarrow C$

4.5 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 humans think. They were informed how much time it should have taken on an average to solve them. 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 in the given scale and that their name or telephone numbers were not asked in the survey.

All questions were of type 'what'. There were no questions of type 'why'. The questionnaire was scrutinized to avoid double-barrelled questions (questions with more than one question embedded within it) and leading questions (questions that suggest the answer or answer that is intended). The assignment of participants to questionnaire 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 four sets of natural language arguments. Each set of arguments had four arguments. Both group $G1$ and $G2$ tested three conditions per argument set. So each participant in both groups responded to 3 (conditions) x 4 (argument sets) = 12 problems. Each problem was presented on a separate sheet of paper in order to make sure that the participant did not read the later problems before answering the earlier ones. The participants marked confidence in the conclusion of argument A in all the problems.

In order to reduce the order effect, we followed the following strategy for group $G1$. Half (12) of the questionnaires for group $G1$ had the argument sets presented in reverse order while maintaining the order of problems within the argument set intact. Each of the two sets of 12 questionnaires was divided into two sets of 6 questionnaires. So there were four sets of 6 questionnaires. The order of argument sets and the order of presenting argument C and argument D were managed using these 4 sets of questionnaires. In the case of group $G2$, we followed a different strategy. There were 4 sets of questionnaires, each set containing 6 questionnaires. These sets were used to manage the order of argument sets and which of argument C and argument D was taken to reinstate argument A.

The participants were allowed to answer the problems only in the order they appeared on the questionnaire. It was mandatory to answer all problems. The participants were not allowed to see later problems in the questionnaire before answering the previous problems.

Following were the arguments sets and individual arguments used in the survey:

Argument Set 1

- Argument A : "Cody does not fly. Therefore, Cody is unable to escape by flying."
- Argument B : "Cody is a bird. Therefore, Cody flies."
- Argument C : "Cody is a rabbit. Therefore, Cody is not a bird."
- Argument D : "Cody is a cat. Therefore, Cody is not a bird."

Argument Set 2

- Argument A : "Smith does not follow American spelling. Therefore, Smith writes 'colour' instead of 'color'."
- Argument B : "Smith speaks American English. Therefore, Smith follows American spelling."
- Argument C : "Smith was born and brought up in England. Therefore, does not speak American English."
- Argument D : "Smith was born and brought up in Australia. Therefore, does not speak American English ."

Argument Set 3

- Argument A : "The car did not slow down. Therefore, the car approached the signal at the same speed or higher."
- Argument B : "Louis applied the brake. Therefore, the car slowed down."
- Argument C : "Louis applied the accelerator instead. Therefore, Louis did not apply the brake."
- Argument D : "Louis applied the clutch instead. Therefore, Louis did not apply the brake."

Argument Set 4

- Argument A : "Stephen is not guilty. Therefore, Stephen is to be free from conviction."
- Argument B : "Stephen was seen at the crime scene at the time of the crime. Therefore, Stephen is guilty."

	base	*defeated*	*reinstated*
Standard	5.4 (1.0)	4.3 (1.0)	4.9 (0.9)
Floating	5.1 (1.3)	3.8 (1.2)	4.7 (1.1)

Table 4.1: Average confidence in the conclusion in *base*, *defeated*, and *reinstated* conditions

- Argument 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."
- Argument 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."

The following personal information about the participants was collected at the end of the survey:

- gender (male, female)
- education level (high school, college diploma, college bachelor, master's degree, doctoral)
- nationality
- age (18-21, 22-25,26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70, above 70)
- proficiency in English language (Expert, Very Good, Good, Competent, Modest, Limited, Extremely Limited, Intermittent)

Post-test interviews were run for one third of the participants in which we asked them to comment on their general experience in answering the problems, the approach that they used to solve the problems, and any difficulties faced.

4.6 Analysis

A mixed-design ANOVA was used to analyse the data, with the acceptance of conclusion as the dependent variable (DV), problem as a 3-level within-subject predictor(*base*, *defeated*, *reinstated*), argument set as a 4-level measure, and type of reinstatement as a 2-level group variable (*standard*, *floating*). The ANOVA (and the post-hoc contrast analysis) looked for a main effect of problem, as well as for an interaction between problem and type of reinstatement. Hypothesis $H1$ was directly investigated by the contrasts between the three levels of the problem variable in the *floating* reinstatement group. Hypothesis $H2$ was investigated through the interaction between problem and type of reinstatement, especially as concerning the contrast between the *defeated* and *reinstated* levels of the problem variables.

In order to test the participants' preference (or lack thereof) between arguments C and D, we used the following Mixed-design ANOVA: 3-level within-subject predictor (*base*, *defeated*, *reinstated*) by 2-level between-group variable (*preference*, *no preference*), and to check the p value of the interaction effect.

In addition to the above analyses, we also analysed the effect of problem understanding in standard reinstatement and floating reinstatement using a regression analysis seeking to predict acceptance of reinstated arguments on the basis of problem understanding, type of reinstatement and the interaction term between these predictors.

4.7 Results

Table 4.1 displays the average confidence in the conclusion of *base*, *defeated*, and *reinstated* conditions, in the standard and floating reinstatement groups.

A visual inspection of Table 4.1 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 variable, pattern as a 3-level within-subject predictor (*base*, *defeated*, *reinstated*), the type of reinstatement as a 2-level between-group variable (standard, floating), and 4 measures corresponding to the 4 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, $F(8, 38) = 1.2$, $p = .32$, $\eta_p^2 = .20$.

As in the study of the cognitive plausibility of reinstatement described in Chapter 3, 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, $F(1, 45) < 1$, $p = .60$, $\eta_p^2 < .01$. As in study of the cognitive plausibility of reinstatement described in Chapter 3, reinstatement is not perfect here also, 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 standard reinstatement problems, this difference appeared to play no role in the ratings they gave for conclusion acceptance. The average understanding rating was 4.6 ($SD = 1.1$) for standard 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, $t = .32$, $p = .75$.

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.

Chapter 5

Discussion

This chapter discusses the results obtained from the experiments described in Chapter 3 and Chapter 4 on the cognitive plausibility of standard reinstatement and floating reinstatement, summarising them in response to the research questions that were posted at the beginning of this paper.

5.1 Cognitive Plausibility of Reinstatement

Reinstatement argument defeat structure is very prevalent in human argumentation. In fact, many arguments in real life scenarios simply follow a iterative reinstatement structure when two persons or two groups argue about the fate of a main argument between them. The results of the experiments that were designed and tested to explore the cognitive plausibility of reinstatement involving a total of 38 participants show that that the notion of reinstatement in abstract argumentation is cognitive plausible by supporting hypotheses $H1$ ($F(1,19) = 26.8, p < 0.001, n_p^2 = 0.59$) and $H2$ ($F(1,19) = 9.9, p < 0.005, n_p^2 = 0.34$), particularly hypotheses $H2$. Hypothesis $H1$ tests the success of a defeat whereas hypothesis $H2$ tests the success of regaining the confidence in the main argument when it is reinstated after a defeat. It is interesting to note that this finding of cognitive plausibility is reinforced by the fact that the experiment results support the manipulation check hypothesis $H4$ ($F(6,12) = 3.8, p = 0.02, n_p^2 = 0.66$) proving that the reinstating argument effectively defeats the defeater of the reinstated argument. That is, the reinstatement is achieved through defeating the defeater rather than supporting the main argument without defeating the defeater. The high confidence in the conclusion of the main argument thus deteriorates significantly when it is defeated by another argument and then again rises significantly when the defeater of it is defeated by the reinstating argument

From the perspective of abstract argument evaluation criteria, the reinstatement framework has the same extensions for complete, grounded, preferred, semi-stable and stable semantics hence the empirical results do not indicate the preference of one semantic over the others. However, we can conclude that the empirical results show that the abstract argumentation criteria are cognitively plausible. From an agent-human argumentation perspective, cognitive plausibility of reinstatement intuitively tells us that the need and the potential to think ahead whenever possible about the probable defeaters and their reinstaters. On the other hand, we believe that some of the empirical results obtained from the experiments cannot be explained using abstract argument evaluation criteria. As mentioned in Chapter 3, the experiment results do not support hypothesis $H3$, meaning that the recovery from a defeat, by a reinstatement is not perfect. According to the two-tailed hypothesis $H3$, the confidence in the main argument should not have a significant difference when we compare the *base* condition and the *reinstated* condition. However, the results show an opposite effect, a significance difference between the levels of confidence ($F(1,19) = 9.1, p = 0.007, n_p^2 = 0.32$) in the *base* condition and the *rein-*

stated condition. We believe that this aspect of the result which is not captured in the abstract argument evaluation criteria, needs to be considered as an important factor in agent-human argumentation. We believe that further experiments concentrating on this part of the results may have to be conducted in the light of the inductive strength (also known as argument strength) of arguments to furnish a theoretical explanation for the partial recovery.

5.2 Cognitive Plausibility of Floating Reinstatement

Floating arguments defeat structure is an interesting defeat structure, both from the theoretical perspective of abstract argumentation and practical point of view. Many times we observe situations where two conflicting arguments defeat a common 'enemy'. An example would be to reason that "person P could be in place A or person P could be in place B (P cannot be at place A and place B at the same time) and hence he is not the one who committed the crime".

As explained in Chapter 4, the framework is characteristic in the sense that someone can argue that the defeater of the main argument can be overruled without resolving the conflict between two mutually conflicting defenders of the main argument and if the defeater of the main argument is overruled then the main argument is justified. Unlike reinstatement, in floating reinstatement, different extension-based semantics do not have the same extensions, as explained in Chapter 4. The very sceptical semantics, the grounded semantics is interestingly an empty set. The preferred extensions are {C, A} and {D, A}. That is, if someone follows preferred semantics, he or she would choose either {C, A} or {D, A}. Argument A is chosen in either case. A more sceptical variant of these preferred extensions called sceptical preference would be to choose only {A}. That is, to choose only argument A and not to show any preference between argument C and D.

The results of the experiments that were designed and tested to explore the cognitive plausibility of floating reinstatement involving a total of 47 participants show that that the notion of floating reinstatement in abstract argumentation is cognitively plausible by supporting hypotheses $H1(F(1, 45) = 13.7,\ p < .001,\ \eta_p^2 = .23)$. Empirical results support hypothesis $H1$ of the floating arguments study which states that reinstatement is cognitively plausible in floating arguments. This result is an important one because it shows that people tend not to follow grounded semantics, which is a very sceptical approach. If people were to follow grounded semantics, then the confidence in the main argument would not have risen after it had been defeated by another argument but reinstated by two conflicting arguments. Thus we could say that preferred semantics was chosen over grounded semantics as far as the main argument is considered. It is very important to know that in reinstating the main argument, whether the participants' had any preference (or lack thereof) between the conflicting defenders, argument C and D. Results from the experiments show that the reinstatement in floating arguments does not appear to result from participants tended a preference for one the mutually defeated arguments. It suggests that the effectiveness of floating reinstatement appears to be genuine even when allowing for the possibility that subjects might have a personal preference for one of the mutually defeating arguments.

The experiments on floating arguments with respect to reinstatement thus show that the empirical results do not support grounded semantics. On the other hand, we do not find a significant interaction between the problem and the preference meaning that we cannot say either the empirical results supports credulous preferred or sceptical preferred fully. We believe this aspect of the results is to be further explored to establish the existence of such a different cognitively plausible idea.

We modified the experiments on floating reinstatement slightly from those on reinstatement by introducing a new dependent variable to measure how well the participants understood the problem or in other words, how hard the problem was for the participant to understand. We found that even though floating arguments are lightly harder to understand compared to standard reinstatement, it did not have any significant role in the acceptance of reinstated arguments.

From the perspective of the theoretical advancement of abstract argumentation theory, the study indicates that the lack of a significant interaction between the pattern and the preference needs to be further explored. From an agent-human argumentation perspective, it is important to understand that the conflicting arguments can be used to reinstate the confidence in the main argument as long as they defeat the defeater of the main argument.

5.3 Comparison Between Standard Reinstatement and Floating Reinstatement

We were also interested in comparing standard reinstatement and floating reinstatement and there are some interesting results here as well. The empirical results do not support hypothesis $H2$ ($F(4, 42) < 1$, $p = .79$, $\eta_p^2 = .04$) in the study of the cognitive plausibility of floating reinstatement intended to check the difference between standard reinstatement and floating. So the first and most important finding here is that floating reinstatement has an effect that is exactly comparable to classic reinstatement. Floating reinstatement seems to do as good a job as standard reinstatement, no more, no less. That is, reinstated arguments in floating reinstatement do partly recover from their defeat and do so to the same extent as in standard reinstatement. Secondly, although floating reinstatement is slightly harder to process than standard reinstatement, this difficult does not impact its effectiveness.

5.4 Summary

This research programme has produced certain interesting insights. To start with, we verified the cognitive plausibility of the notion of reinstatement in abstract argumentation. We also verified the cognitive plausibility of reinstatement in floating arguments. We found that the floating reinstatement has an effect that is exactly comparable to classic reinstatement which is quite remarkable. The mutually conflicting nature of the defenders does not play any role that undermines the job of reinstating, crucially, even when this conflicting nature of the defenders is acknowledged by the participants. The finding that there is no significant interaction between the pattern and preference in floating reinstatement cannot be explained using the abstract argument evaluation criteria. This finding needs to be further explored to establish the existence of such a different cognitively plausible notion. Partial recovery of confidence in both types of reinstatement is also important to mention and it needs further exploration using a probabilistic approach.

Chapter 6

Conclusion and Future Work

Classical reasoning such as reasoning based on propositional logic is monotonic in the sense that adding new information does not remove any previously made conclusion. Common sense suggests that the failure of monotonic reasoning is widespread. Many a time, we jump to conclusions and then we correct our conclusions based on further information as it arrives. Reasoning of this kind is called as *nonmonotonic* reasoning. Recently, the study of nonmonotonic reasoning has appealed to the powerful notion of argument through the proliferation of so-called argumentation systems. A very influential approach proposed by Dung to argumentation systems completely abstracts the origin and the internal structure of the arguments. The focus, instead, is on the relationships between these abstract arguments using defeat relations. The defeat structures can be in different forms such as mutual attack (one argument attacks another argument and the attacked argument attacks the attacker), reinstatement (an argument reinstates another argument by defeating the defeater) and a cycle of attack. The obvious question here is to identify which arguments are rejected, accepted and undecided in such defeat structures.

Extension-based abstract argument evaluation criteria (also known as extension-based semantics) can be thought of as criteria for making the decision of which arguments can be accepted, rejected and cannot be accepted or rejected. Extension-based argument evaluation criteria have been studied in detail in the literature. These evaluation criteria have been mainly developed for obtaining desirable formal or computational properties, largely based on intuition. The cognitive plausibility of such evaluation criteria has mostly been ignored. However, it is crucial to understand the cognitive plausibility of such evaluation criteria if we are to build software agents capable of interacting persuasively with humans through arguments. Theoretical advancement of abstract argumentation theory is also benefitted from exploring the cognitive plausibility of evaluation criteria. If existing argumentation semantics fail to capture human evaluation of arguments, then observing how humans actually evaluate arguments can inform the design of new semantics based on well-founded empirical data.

Cognitive plausibility of abstract argument evaluation criteria was explored by conducting psychological experiments. Scenarios of standard reinstatement (an argument reinstating another argument by defeating the defeater) and the floating reinstatement (two mutually conflicting arguments reinstating another argument by defeating the defeater) were studied in detail. Reinstatement argument defeat structure is very prevalent in human argumentation. In fact, many arguments in real life scenarios simply follow a iterative reinstatement structure when two persons or two groups argue about the fate of a main argument between them. The empirical results show that the notion of reinstatement in abstract argumentation is cognitively plausible by supporting both grounded and preferred semantics. Reinstatement argumentation framework (also called defeat structure) has the same extension for complete, grounded, preferred, stable and semi-stable semantics and hence empirical results cannot indicate the preference of one semantic over the others. However, we can conclude that the empirical results show that the abstract argumentation criteria are cognitively plausible and they can be used by intelligent

agents in reinstatement scenarios. On the other hand, some of the empirical results obtained from the experiments cannot be explained using abstract argument evaluation criteria. The recovery from a defeat, by a reinstatement is not perfect and the idea of such a partial recovery is not dealt with in the abstract argumentation theory, yet it is an important factor if we are to build software agents capable of interacting persuasively with humans through arguments.

Unlike standard reinstatement, in floating reinstatement, different extension-based semantics do not have the same extensions. The very sceptical semantics, the grounded semantics is interestingly an empty set. The preferred extensions are {C, A} and {D, A}. That is, if someone follows preferred semantics, he or she would choose either {C, A} or {D, A}. Argument A is chosen in either case. A more sceptical variant of these preferred extensions called sceptical preference would be to choose only {A}. That is, to choose only argument A and not to show any preference between argument C and D. The results show that the notion of floating reinstatement is cognitively plausible by not supporting the grounded semantics. It means humans do not follow a sceptical approach in floating reinstatement. However, lack of a significant interaction between the pattern and the reasoners' preference (or lack of preference) for one of the two mutually conflicting defenders in the tests indicates the existence of a different cognitively plausible notion that cannot be explained using the abstract argument evaluation criteria. The idea is that there is no clear endorsement for both credulous preferred and sceptical preferred semantics. The results also suggest that a floating reinstatement has an effect that is not significantly different from that of the standard form even though the extensions of semantics are different. That is, the mutually conflicting nature of defenders does not play any role that undermines the job of reinstating the main argument. As in the case of standard reinstatement, the floating reinstatement also offers only partial recovery. From the perspective of the theoretical advancement of abstract argumentation theory, the study indicates that the sceptical variant of preferred semantics is an important semantics to be considered. From an agent-human argumentation perspective, it is important to understand that conflicting arguments can be used to reinstate the confidence in the main argument as long as they defeat the defeater of the main argument. Another result of the experiment, which is not as important as the previous ones is that, although floating reinstatement is slightly harder to process than standard reinstatement, this difficulty does not impact its effectiveness.

This study can be extended at least in the following ways:

- Test the cognitive plausibility of abstract argument evaluation criteria on other characteristic defeat structures of higher complexity considering more complicated extension-based semantics. This should help understand the full spectrum of all such cognitive plausibility of abstract argumentation criteria.
- Study the partial recovery of confidence in reinstatement using a probabilistic approach. Partial recovery is an important factor to be considered in agent-human argumentation. It can be considered as a deviation from abstract argumentation in which defeats are perfect. The study of confidence levels with respect to abstract argumentation can also be done. This study explores the threshold of confidence at which one can say that the defeat is successful.
- Explore the role of negations in the premise and conclusions of individual arguments to see the effect on the confidence in the arguments
- Conduct experiments with arguments employing inference patterns other than modus ponens as their internal structure.

Appendix A

The Cognitive Plausibility of Reinstatement - Pilot Survey Analysis

We conducted the reinstatement survey two times (pilot and final). In the pilot survey we found that the reinstatement did not work. Following is the analysis of the pilot survey.

We found that the base (when argument A is presented alone) acceptance rating of the conclusion when we average the scores across the 6 contents as 6.3 (SD = 0.60) whereas acceptance rating of defeated conclusion (when argument A is attacked by argument B) as 4.6 (SD = 1.5). The acceptance rating of reinstated (when argument A is attacked by argument B but reinstated by argument C) was 4.6 (SD = 1.30).

Acceptance ratings were analysed with a repeated-measure analysis of variance, with pattern as a 3-level predictor *base*, *defeated* and *reinstated*), and 6 measures corresponding to 6 contents. The repeated-measure analysis of variance detected a significant effect of pattern, $F(2, 18) = 20.6, p < 0.001, n_p^2 = 0.70$. This overall effect, however, was mostly due to the contrast between the *base* condition and other two conditions. The ratings in the *base* conditions were significantly higher than ratings in the *defeated* condition, $F(1, 19) = 32.7, p < 0.001, n_p^2 = 0.63$, and significantly higher than ratings in the *reinstated* condition, $F(1, 19) = 34.6, p = 0.001, n_p^2 = 0.65$. However, and critically, ratings in the *defeated* condition were not significantly different from ratings in the *reinstated* condition, $F(1, 19) = 1.0, p = 0.33, n_p^2 = 0.05$.

This lack of reinstatement must be related to the results of the manipulation check. Averaging across the 6 contents, the *base defeater* acceptance rating of the defeaters was 5.2 (SD = 1.0), and the acceptance ratings for the *attacked defeater* was 4.8 (SD = 1.3). The acceptance ratings for the manipulation check were analyzed with a repeated-measure analysis of variance, with pattern as a 2-level predictor(*base defeater*, *attacked defeater*), and 6 measures corresponding to the 6 contents. The multivariate test failed to detect a significant effect of pattern, $F(6, 12) = 1.5, p = 0.27, n_p^2 = 0.42$. It thus appears that the arguments chosen as reinstaters failed to properly attack the defeaters, preventing reinstatement from occurring.

The surprising outcome that the survey results did not support the hypothesis $H4$ but supported the very similar hypothesis $H1$ led us to think about the subtle differences between them. It can be noticed that in case of $H1$, the conclusion of the attacked argument (namely argument A) is a positive statement. On the other hand, in the case of $H4$, the conclusion of the attacked argument (namely argument B) is a negated statement.

To clarify the above point, let us redraw the Figure 3.1 summarizing our hypotheses by unpacking the internal structure of the arguments slightly. Instead of writing an argument abstractly as a letter A, B, etc., we expand each argument as follows. If X is an argument with a positive (i.e. non-negated) premise and conclusion, we write its general "form" as follows:

$(P_X \rightarrow C_X)$

For example, an actual argument taking the above form is:

P_A = "Louis applied the brake and the brake was effective." C_A = "Therefore, the car slowed down."

If the premise or conclusion is a negated statement, we write these as $\neg P_X$ and $\neg C_X$ respectively. For example, an actual argument taking the form $(P_B \rightarrow \neg C_B)$ is:

P_B = "Mary is mute." C_B = "Therefore, Mary does not talk a lot to her friends."

With the above new notation in place, we recall that in most argument sets used in the pilot experiment (with the exception of argument set 1), arguments took the following forms (note how the conclusion of each attacker negates the premise of the attacked):

- argument A took the form $(P_A \rightarrow C_A)$,
- argument B took the form $(P_B \rightarrow \neg P_A)$,
- argument C took the form $(P_C \rightarrow \neg P_B)$.

In the Figure A.1, we rewrite the hypotheses using this new notation to highlight the structure of negations in the majority of our argument sets.

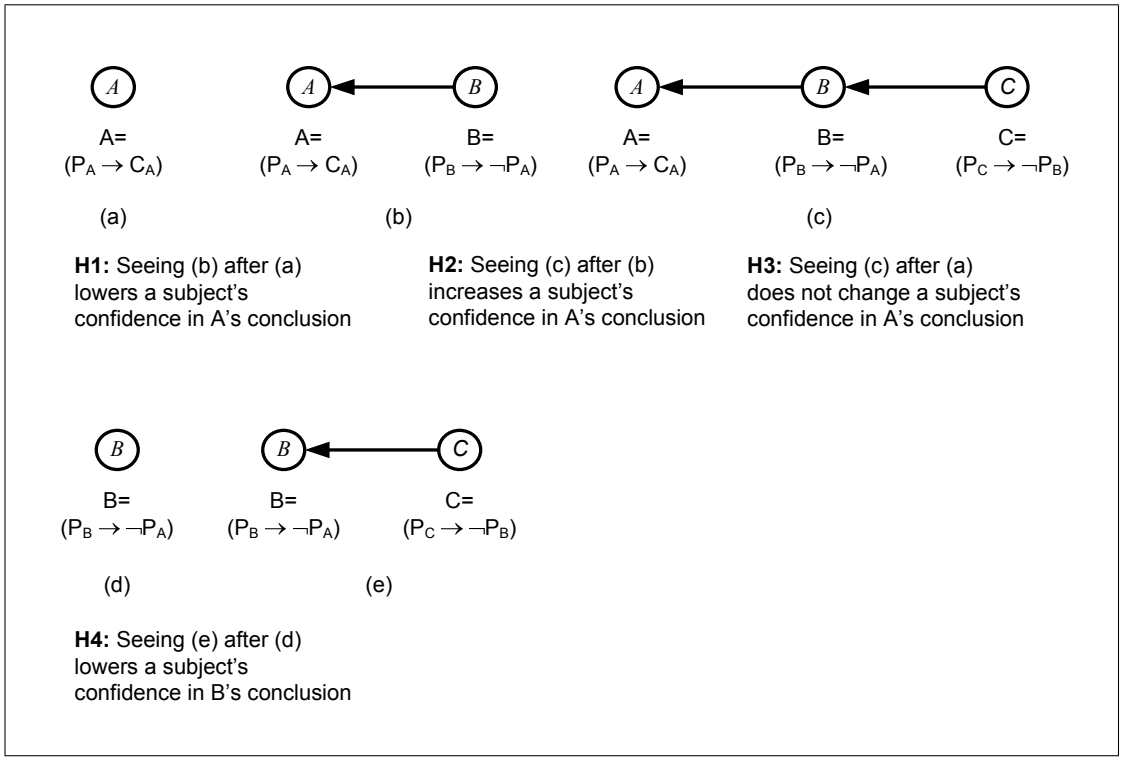

Figure A.1: Summary of Hypotheses (With Argument Internal Structure)

With this explicit structure in place, we are ready to highlight a possible explanation for our negative results (i.e. for failing to satisfy hypotheses $H2$ and $H4$). This may be due to the failure of the defeat of argument B by argument C (as revealed by our manipulation check using $H4$).

Notice that under the original abstraction, the structural relationship in $H1$ and $H4$ were identical. In $H1$, participants saw argument A, followed the defeater B of A. Likewise, in $H4$, participants saw argument B, followed by the defeater C of B. Yet, while $H1$ is verified, $H4$ was not. Looking at the above new figure provides a possible explanation. By looking "into the

arguments," we notice that argument sets (b) and (e) have different forms. When presented with set (e), the participant needed to consider two negations in the conclusions of both the attacker (argument C) and the attacked (argument B). In contrast, when presented with set (b), the participant only faced a single negation of the conclusion of A. Thus, the sort of reasoning involved in evaluating $H4$ is less trivial, which may have overloaded the participants.

Given the above, one possible direction was to explore whether this specific structure of negation is the reason behind the results we obtained.

We designed a new experiment in which we flipped the negations. We redefined the internal structure of arguments, while maintaining the overall defeat structure, as follows:

- argument A takes the form $(\neg P_A \rightarrow C_A)$,
- argument B takes the form $(P_B \rightarrow P_A)$,
- argument C takes the form $(P_C \rightarrow \neg P_B)$.

This way, we slightly isolated the negations. When viewing B after A, only one negation is present, and likewise when viewing C after B. Of course, when participants view all arguments A, B and C, they still need to deal with two negations, but these are more isolated than in the original case.

Appendix B

The Cognitive Plausibility of Reinstatement - Pilot Survey Questionnaires

Following are the questionnaires. Questionnaires are categorised as $QN1$, $QN2$, $QN3$ and $QN4$. The questionnaires $QN1$ and $QN2$ are for participant group $G1$. The questionnaires $QN3$ and $QN4$ are for participant group $G2$. $QN2$ is obtained by simply presenting sections in $QN1$ in the reverse order while preserving the order of questions within the section. $QN4$ is obtained by simply presenting sections in $QN3$ in the reverse order while preserving the order of questions within the section. In the actual survey each problem was given on a separate sheet of paper.

Introduction

Thank you for participating in this simple survey. The purpose of the survey is to collect information on how human beings argue. Please note that there are no trick questions in this survey. All you need to do is to read the question, understand it, and choose an answer that you most agree with. Please note that all answers are right; so you choose the answer that you believe in. All of these questions require only common sense and what we are looking for is the obvious answer that you feel. Some of the questions might be very simple, so no need to guess for any implicit meaning.

Each questionnaire contains 6 sections of questions and each section contains 3 (for group $G2$, 2) questions. Each section is totally independent of other sections as there are no connections among sections.

Each question should take a maximum of 1 minute, but it may vary question to question and participant to participant. Mark your answer by circling the number against the answer. For any clarification, please ask the person in charge of the questionnaire.

Please ensure the following before you choose an answer:

1. You understand the question and the answer
2. You believe in the answer that you chose

Questionnaire QN1

Section 1, Question 1

- "The battery of Alex's car is dead. Therefore, Alex's car will not start."

Alex's car will not start is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 1, Question 2

- "The battery of Alex's car is dead. Therefore, Alex's car will not start."
- "The battery of Alex's car has just been changed. Therefore, the battery of Alex's car is not dead."

Alex's car will not start is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 1, Question 3

- "The battery of Alex's car is dead. Therefore, Alex's car will not start."
- "The battery of Alex's car has just been changed. Therefore, the battery of Alex's car is not dead."
- "The garage is still closed. Therefore, the battery of Alex's car has not just been changed."

Alex's car will not start is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 1

- "Louis applied the brake and the brake was effective. Therefore, the car slowed down."

The car slowed down is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 2

- "Louis applied the brake and the brake was effective. Therefore, the car slowed down."
- "The brake fluid was empty. Therefore, the brake was not effective."

The car slowed down is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 3

- "Louis applied the brake and the brake was effective. Therefore, the car slowed down."
- "The brake fluid was empty. Therefore, the brake was not effective."
- "The car had just been undergone maintenance service. Therefore, the brake fluid was not empty."

The car slowed down is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 1

- "Mary talks a lot to her friends on the phone. Therefore, Mary has a large phone bill."

Mary has a large phone bill is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 2

- "Mary talks a lot to her friends on the phone. Therefore, Mary has a large phone bill."
- "Mary is mute. Therefore, Mary does not talk a lot to her friends."

Mary has a large phone bill is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 3

- "Mary talks a lot to her friends on the phone. Therefore, Mary has a large phone bill."
- "Mary is mute. Therefore, Mary does not talk a lot to her friends."
- "Marys likes to sing. Therefore, Mary is not mute."

Mary has a large phone bill is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 1

- "John jumped the red signal. Therefore, John will be punished."

John will be punished is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 2

- "John jumped the red signal. Therefore, John will be punished."
- "The green signal was on for John. Therefore, John did not jump the red signal."

John will be punished is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 3

- "John jumped the red signal. Therefore, John will be punished."
- "The green signal was on for John. Therefore, John did not jump the red signal."
- "Other vehicles were moving across in the signal junction. Therefore, green signal was not on for John."

John will be punished is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 1

- "Sophie wants to learn swimming. Therefore, Sophie takes swimming lessons."

Sophie takes swimming lessons is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 2

- "Sophie wants to learn swimming. Therefore, Sophie takes swimming lessons."
- "Sophie already swims well. Therefore, Sophie does not want to learn swimming."

Sophie takes swimming lessons is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 3

- "Sophie wants to learn swimming. Therefore, Sophie takes swimming lessons."
- "Sophie already swims well. Therefore, Sophie does not want to learn swimming."
- "Recently, Sophie almost drowned in a pool. Therefore, Sophie does not swim well."

Sophie takes swimming lessons is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 1

- "Alice eats a lot of ice cream. Therefore, Alice loves ice cream."

Alice loves ice cream is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 2

- "Alice eats a lot of ice cream. Therefore, Alice loves ice cream."
- "Alice is on a sugar-free diet. Therefore, she does not eat a lot of ice cream."

Alice loves ice cream is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 3

- "Alice eats a lot of ice cream. Therefore, Alice loves ice cream."
- "Alice is on a sugar-free diet. Therefore, she does not eat a lot of ice cream."
- "Alice eats a lot of fruit. Therefore, she is not on a sugar-free diet."

Alice loves ice cream is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Participant Information

Gender

1. Male
2. Female

Education

1. High school
2. College diploma
3. College bachelor
4. Master's degree
5. Doctoral

Age

1. 18-21
2. 22-25
3. 26-30
4. 31-35
5. 36-40
6. 41-45
7. 46-50
8. 51-55
9. 56-60
10. 61-65
11. 66-70
12. Above 70

Proficiency In English

1. Expert
2. Very Good
3. Good
4. Competent
5. Modest
6. Limited
7. Extremely Limited
8. Intermittent

Nationality:

Questionnaire QN3

Section 1, Question 1

- "The battery of Alex's car has just been changed. Therefore, the battery of Alex's car is not dead."

Battery of Alex's car is not dead is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 1, Question 2

- "The battery of Alex's car has just been changed. Therefore, the battery of Alex's car is not dead."
- "The garage is still closed. Therefore, the battery of Alex's car has not just been changed."

Battery of Alex's car is not dead is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 1

- "The brake fluid was empty. Therefore, the brake was not effective."

The brake was not effective is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 2

- "The brake fluid was empty. Therefore, the brake was not effective."
- "The car had just been undergone maintenance service. Therefore, the brake fluid was not empty."

The brake was not effective is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 1

- "Mary is mute. Therefore, Mary does not talk a lot to her friends."

Mary does not talk a lot to her friends is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 2

- "Mary is mute. Therefore, Mary does not talk a lot to her friends."
- "Marys likes to sing. Therefore, Mary is not mute."

Mary does not talk a lot to her friends is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 1

- "The green signal was on for John. Therefore, John did not jump the red signal."

John did not jump the red signal is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 2

- "The green signal was on for John. Therefore, John did not jump the red signal."
- "Other vehicles were moving across in the signal junction. Therefore, green signal was not on for John."

John did not jump the red signal is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 1

- "Sophie already swims well. Therefore, Sophie does not want to learn swimming."

Sophie does not want to learn swimming is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 2

- "Sophie already swims well. Therefore, Sophie does not want to learn swimming."
- "Sophie almost drowned recently in a pool. Therefore, Sophie does not know swimming."

Sophie does not want to learn swimming is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 1

- "Alice is on a sugar-free diet. Therefore, Alice does not eat a lot of ice cream."

Alice does not eat a lot of ice cream is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 2

- "Alice is on a sugar-free diet. Therefore, Alice does not eat a lot of ice cream."
- "Alice eats a lot of fruit. Therefore, Alice is not on a sugar-free diet.

Alice does not eat a lot of ice cream is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Participant Information

Gender

1. Male
2. Female

Education

1. High school
2. College diploma
3. College bachelor
4. Master's degree
5. Doctoral

Age

1. 18-21
2. 22-25
3. 26-30
4. 31-35
5. 36-40
6. 41-45
7. 46-50
8. 51-55
9. 56-60
10. 61-65
11. 66-70
12. Above 70

Proficiency In English

1. Expert
2. Very Good
3. Good
4. Competent
5. Modest
6. Limited
7. Extremely Limited
8. Intermittent

Nationality:

Appendix C

The Cognitive Plausibility of Reinstatement - Final Survey Questionnaires

Following are the questionnaires. Questionnaires are categorised as $QN1$, $QN2$, $QN3$ and $QN4$. The questionnaires $QN1$ and $QN2$ are for participant group $G1$. The questionnaires $QN3$ and $QN4$ are for participant group $G2$. $QN2$ is obtained by simply presenting sections in $QN1$ in the reverse order while preserving the order of questions within the section. $QN4$ is obtained by simply presenting sections in $QN3$ in the reverse order while preserving the order of questions within the section. In the actual survey each problem was given on a separate sheet of paper.

Introduction

Thank you for participating in this simple survey. The purpose of the survey is to collect information on how human beings argue. Please note that there are no trick questions in this survey. All you need to do is to read the question, understand it, and choose an answer that you most agree with. Please note that all answers are right; so you choose the answer that you believe in. All of these questions require only common sense and what we are looking for is the obvious answer that you feel. Some of the questions might be very simple, so no need to guess for any implicit meaning.

Each questionnaire contains 6 sections of questions and each section contains 3 (for group $G2$, 2) questions. Each section is totally independent of other sections as there are no connections among sections.

Each question should take a maximum of 1 minute, but it may vary question to question and participant to participant. Mark your answer by circling the number against the answer. For any clarification, please ask the person in charge of the questionnaire.

Please ensure the following before you choose an answer:

1. You understand the question and the answer
2. You believe in the answer that you chose

Questionnaire QN1

Section 1, Question 1

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

Alex's car will halt is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 1, Question 2

- "The battery of Alex's car is not working. Therefore, Alex's car will halt."
- "The battery of Alex's car has just been changed today. Therefore, the battery of Alex's car is working."

Alex's car will halt is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 1, Question 3

- "The battery of Alex's car is not working. Therefore, Alex's car will halt."
- "The battery of Alex's car has just been changed today. Therefore, the battery of Alex's car is working."
- "The garage was closed today. Therefore, the battery of Alex's car has not been changed today."

Alex's car will halt is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 1

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

The car slowed down is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 2

- "Louis applied the brake and the brake was not faulty. Therefore, the car slowed down."
- "The brake fluid was empty. Therefore, the brake was faulty."

The car slowed down is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 3

- "Louis applied the brake and the brake was not faulty. Therefore, the car slowed down."
- "The brake fluid was empty. Therefore, the brake was faulty."
- "The car had just undergone maintenance service. Therefore, the brake fluid was not empty."

The car slowed down is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 1

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

Mary has a large phone bill is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 2

- "Mary does not moderate her phone usage. Therefore, Mary has a large phone bill."
- "Mary has a speech disorder. Therefore, Mary moderates her phone usage."

Mary has a large phone bill is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 3

- "Mary does not moderate her phone usage. Therefore, Mary has a large phone bill."
- "Mary has a speech disorder. Therefore, Mary moderates her phone usage."
- "Mary is a singer. Therefore, Mary does not have a speech disorder."

Mary has a large phone bill is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 1

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

Leila's emails are secured from John is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 2

- "John has no way to know Leila's password. Therefore, Leila's emails are secured from John."
- "Leila's secret question is very easy to answer. Therefore, John has a way to know Leila's password."

Leila's emails are secured from John is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 3

- "John has no way to know Leila's password. Therefore, Leila's emails are secured from John."
- "Leila's secret question is very easy to answer. Therefore, John has a way to know Leila's password."
- "Leila purposely gave a wrong answer to her secret question. Therefore, Leila's secret question is not very easy to answer."

Leila's emails are secured from John is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 1

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

Mike's laptop is vulnerable to computer viruses is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 2

- "Mike's laptop does not have anti-virus software installed. Therefore, Mike's laptop is vulnerable to computer viruses."
- "Nowadays anti-virus software is always available by default on purchase. Therefore, Mike's laptop has anti-virus software."

Mike's laptop is vulnerable to computer viruses is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 3

- "Mike's laptop does not have anti-virus software installed. Therefore, Mike's laptop is vulnerable to computer viruses."
- "Nowadays anti-virus software is always available by default on purchase. Therefore, Mike's laptop has anti-virus software."
- "Some laptops are very cheap and have minimal software. Therefore, anti-virus software is not always available by default."

Mike's laptop is vulnerable to computer viruses is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 1

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

All lights in the house are off is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 2

- "There is no electricity in the house. Therefore, all lights in the house are off."
- "There is a working portable generator in the house. Therefore, there is electricity in the house."

All lights in the house are off is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 3

- "There is no electricity in the house. Therefore, all lights in the house are off."
- "There is a working portable generator in the house. Therefore, there is electricity in the house."
- "The fuel tank of the portable generator is empty. Therefore, the portable generator is not working."

All lights in the house are off is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Participant Information

Gender

1. Male
2. Female

Education

1. High school
2. College diploma
3. College bachelor
4. Master's degree
5. Doctoral

Age

1. 18-21
2. 22-25
3. 26-30
4. 31-35
5. 36-40
6. 41-45
7. 46-50
8. 51-55
9. 56-60
10. 61-65
11. 66-70
12. Above 70

Proficiency In English

1. Expert
2. Very Good
3. Good
4. Competent
5. Modest
6. Limited
7. Extremely Limited
8. Intermittent

Nationality:

Questionnaire QN3

Section 1, Question 1

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

Battery of Alex's car is working is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 1, Question 2

- "The battery of Alex's car has just been changed today. Therefore, the battery of Alex's car is working."
- "The garage was closed today. Therefore, the battery of Alex's car has not been changed today."

Battery of Alex's car is working is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 1

- "The brake fluid was empty. Therefore, the brake was faulty."

The brake was faulty is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 2, Question 2

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

The brake was faulty is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 1

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

Mary moderates her phone usage is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 3, Question 2

- "Mary has a speech disorder. Therefore, Mary moderates her phone usage."
- "Mary is a singer. Therefore, Mary does not have a speech disorder."

Mary moderates her phone usage is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 1

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

John has a way to know Leilas password is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 4, Question 2

- "Leila's secret question is very easy to answer. Therefore, John has a way to know Leila's password."
- "Leila purposely gave a wrong answer to her secret question. Therefore, Leila's secret question is not very easy to answer."

John has a way to know Leilas password is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 1

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

Mike's laptop has anti-virus software is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 5, Question 2

- "Nowadays anti-virus software is always available by default on purchase. Therefore, Mike's laptop has anti-virus software."
- "Some laptops are very cheap and have minimal software. Therefore, anti-virus software is not always available by default."

Mike's laptop has anti-virus software is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 1

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

There is electricity in the house is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Section 6, Question 2

- "There is a working portable generator in the house. Therefore, there is electricity in the house."
- "The fuel tank of the portable generator is empty. Therefore, the portable generator is not working."

There is electricity in the house is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Participant Information

Gender

1. Male
2. Female

Education

1. High school
2. College diploma
3. College bachelor
4. Master's degree
5. Doctoral

Age

1. 18-21
2. 22-25
3. 26-30
4. 31-35
5. 36-40
6. 41-45
7. 46-50
8. 51-55
9. 56-60
10. 61-65
11. 66-70
12. Above 70

Proficiency In English

1. Expert
2. Very Good
3. Good
4. Competent
5. Modest
6. Limited
7. Extremely Limited
8. Intermittent

Nationality:

Appendix D

The Cognitive Plausibility of Floating Reinstatement - Questionnaires

This survey is intended to test hypotheses $H1$ and $H2$. The same argument sets are used to test all hypotheses in both groups.

Following are the questionnaires. Questionnaires are categorised as $QN1$, $QN2$, $QN3$, $QN4$, $QN5$, $QN6$, $QN7$ and $QN8$. The questionnaires $QN1$, $QN2$, $QN3$ and $QN4$ are for participant group $G1$. The questionnaires $QN5$, $QN6$, $QN7$ and $QN8$ are for participant group $G2$. $QN2$, $QN3$ and $QN4$ are obtained by simply presenting sections in $QN1$ in the reverse order while preserving the order of questions within the section as well as changing the order of argument A and B. $QN6$, $QN7$ and $QN8$ are obtained by simply presenting sections in $QN3$ in the reverse order while preserving the order of questions within the section as well as choosing whether argument C or argument D to reinstate argument A. Following 'Introduction' section was there in all the questionnaires. In the actual survey each problem was given on a separate sheet of paper.

Introduction

Thank you for participating in this simple survey. Please note that this is not a test but a survey. The purpose of this survey is to collect information on how human beings argue. Please note that there are no trick questions in this survey. All you need to do is to read the question, understand it, and choose an answer that you most agree with. Please note that all answers are right; so you choose the answer that you believe in. All of these questions require only common sense and what we are looking for is the obvious answer that you feel. Some of the questions might be very simple, so no need to guess for any implicit meaning. Please also note that your name and telephone numbers are not asked in this survey.

Each questionnaire contains 4 sections of questions and each section contains 3 questions. Each section is totally independent of other sections as there are no connections among sections.

Each question should take a maximum of 2 minute, but it may vary question to question and participant to participant. Mark your answer by circling the number against the answer. Please note that you must answer a question before reading the next question. For any clarification, please ask the person in charge of the questionnaire.

Please ensure the following before you choose an answer:

1. You understand the question and the answer
2. You believe in the answer that you chose

Questionnaire QN1

Section 1, Question 1

- "Cody does not fly. Therefore, Cody is unable to escape by flying."

Cody is unable to escape by flying is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 1, Question 2

- "Cody does not fly. Therefore, Cody is unable to escape by flying."
- "Cody is a bird. Therefore, Cody flies."

Cody is unable to escape by flying is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 1, Question 3

- "Cody does not fly. Therefore, Cody is unable to escape by flying."
- "Cody is a bird. Therefore, Cody flies."
- "Cody is a rabbit. Therefore, Cody is not a bird."
- "Cody is a cat. Therefore, Cody is not a bird."

Cody is unable to escape by flying is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Do you think that:

1. C is a better argument than D
2. D is a better argument than C
3. C and D are about equally good

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 2, Question 1

- "Smith does not follow American spelling. Therefore, Smith writes 'colour' instead of 'color'."

Smith writes 'colour' instead of 'color' is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 2, Question 2

- "Smith does not follow American spelling. Therefore, Smith writes 'colour' instead of 'color'."
- "Smith speaks American English. Therefore, Smith follows American spelling."

Smith writes 'colour' instead of 'color' is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 2, Question 3

- "Smith does not follow American spelling. Therefore, Smith writes 'colour' instead of 'color'."
- "Smith speaks American English. Therefore, Smith follows American spelling."
- "Smith was born and brought up in England. Therefore, does not speak American English."
- "Smith was born and brought up in Australia. Therefore, does not speak American English."

Smith writes 'colour' instead of 'color' is:

1. certainly false
2. much more false than true
3. slightly more false than true

4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Do you think that:

1. C is a better argument than D
2. D is a better argument than C
3. C and D are about equally good

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 3, Question 1

- "The car did not slow down. Therefore, the car approached the signal at the same speed or higher."

The car approached the signal at the same speed or higher is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 3, Question 2

- "The car did not slow down. Therefore, the car approached the signal at the same speed or higher."
- "Louis applied the brake. Therefore, the car slowed down."

The car approached the signal at the same speed or higher is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 3, Question 3

- "The car did not slow down. Therefore, the car approached the signal at the same speed or higher."
- "Louis applied the brake. Therefore, the car slowed down."
- "Louis applied the accelerator instead. Therefore, Louis did not apply the brake."
- "Louis applied the clutch instead. Therefore, Louis did not apply the brake."

The car approached the signal at the same speed or higher is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Do you think that:

1. C is a better argument than D
2. D is a better argument than C
3. C and D are about equally good

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 4, Question 1

- "Stephen is not guilty. Therefore, Stephen is to be free from conviction."

Stephen is to be free from conviction is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 4, Question 2

- "Stephen is not guilty. Therefore, Stephen is to be free from conviction."
- "Stephen was seen at the crime scene at the time of the crime. Therefore, Stephen is guilty."

Stephen is to be free from conviction is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 4, Question 3

- "Stephen is not guilty. Therefore, Stephen is to be free from conviction."
- "Stephen was seen at the crime scene at the time of the crime. Therefore, Stephen is guilty."
- "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."
- "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."

Stephen is to be free from conviction is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

Do you think that:

1. C is a better argument than D
2. D is a better argument than C
3. C and D are about equally good

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Participant Information

Gender

1. Male
2. Female

Education

1. High school
2. College diploma
3. College bachelor
4. Master's degree
5. Doctoral

Age

1. 18-21
2. 22-25
3. 26-30
4. 31-35
5. 36-40
6. 41-45
7. 46-50
8. 51-55
9. 56-60
10. 61-65
11. 66-70
12. Above 70

Proficiency In English

1. Expert
2. Very Good
3. Good
4. Competent
5. Modest
6. Limited
7. Extremely Limited
8. Intermittent

Nationality:

Questionnaire QN5

Section 1, Question 1

- "Cody does not fly. Therefore, Cody is unable to escape by flying."

Cody is unable to escape by flying is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 1, Question 2

- "Cody does not fly. Therefore, Cody is unable to escape by flying."
- "Cody is a bird. Therefore, Cody flies."

Cody is unable to escape by flying is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 1, Question 3

- "Cody does not fly. Therefore, Cody is unable to escape by flying."
- "Cody is a bird. Therefore, Cody flies."
- "Cody is a rabbit. Therefore, Cody is not a bird."

Cody is unable to escape by flying is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 2, Question 1

- "Smith does not follow American spelling. Therefore, Smith writes 'colour' instead of 'color'."

Smith writes 'colour' instead of 'color' is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 2, Question 2

- "Smith does not follow American spelling. Therefore, Smith writes 'colour' instead of 'color'."
- "Smith speaks American English. Therefore, Smith follows American spelling."

Smith writes 'colour' instead of 'color' is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 2, Question 3

- "Smith does not follow American spelling. Therefore, Smith writes 'colour' instead of 'color'."
- "Smith speaks American English. Therefore, Smith follows American spelling."
- "Smith was born and brought up in England. Therefore, does not speak American English."

Smith writes 'colour' instead of 'color' is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 3, Question 1

- "The car did not slow down. Therefore, the car approached the signal at the same speed or higher."

The car approached the signal at the same speed or higher is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 3, Question 2

- "The car did not slow down. Therefore, the car approached the signal at the same speed or higher."
- "Louis applied the brake. Therefore, the car slowed down."

The car approached the signal at the same speed or higher is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 3, Question 3

- "The car did not slow down. Therefore, the car approached the signal at the same speed or higher."
- "Louis applied the brake. Therefore, the car slowed down."
- "Louis applied the accelerator instead. Therefore, Louis did not apply the brake."

The car approached the signal at the same speed or higher is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 4, Question 1

- "Stephen is not guilty. Therefore, Stephen is to be free from conviction."

Stephen is to be free from conviction is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 4, Question 2

- "Stephen is not guilty. Therefore, Stephen is to be free from conviction."
- "Stephen was seen at the crime scene at the time of the crime. Therefore, Stephen is guilty."

Stephen is to be free from conviction is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Section 4, Question 3

- "Stephen is not guilty. Therefore, Stephen is to be free from conviction."
- "Stephen was seen at the crime scene at the time of the crime. Therefore, Stephen is guilty."
- "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."

Stephen is to be free from conviction is:

1. certainly false
2. much more false than true
3. slightly more false than true
4. as false as true
5. slightly more true than false
6. much more true than false
7. certainly true

How clearly did you understand the problem?

1. Not at all
2. Very little
3. Little
4. Somewhat
5. Well
6. Very well
7. Completely

Participant Information

Gender

1. Male
2. Female

Education

1. High school
2. College diploma
3. College bachelor
4. Master's degree
5. Doctoral

Age

1. 18-21
2. 22-25
3. 26-30
4. 31-35
5. 36-40
6. 41-45
7. 46-50
8. 51-55
9. 56-60
10. 61-65
11. 66-70
12. Above 70

Proficiency In English

1. Expert
2. Very Good
3. Good
4. Competent
5. Modest
6. Limited
7. Extremely Limited
8. Intermittent

Nationality:

Appendix E

Experiment Results Tables

Following are tables of the results obtained in the surveys. Figure E.1 shows the confidence scores from group $G1$ testing the cognitive plausibility of reinstatement whereas Figure E.2 shows the confidence scores from group $G2$ testing the cognitive plausibility of reinstatement. Figure E.3 shows the confidence scores from group $G1$ testing the cognitive plausibility of floating reinstatement whereas Figure E.4 shows the confidence scores from group $G2$ testing the cognitive plausibility of floating reinstatement.

The confidence scores are given in the order of problems as they appeared in the questionnaires. They are provided here taking into consideration the changes made to the order of sections to reduce the order effect. As we have different sections in the questionnaires, the scores are also given according to the sections. Each section corresponds to an argument set and each question in a section corresponds to a problem. Sections are referred to as 'Sec 1', 'Sec 2', etc and questions are referred to as 'Q1', 'Q2', etc.

The level of understanding a problem by the participants was measured in the surveys intended to test the cognitive plausibility of floating reinstatement. These are provided in Figure E.5 and Figure E.6 for participant group $G1$ and $G2$ respectively. In addition, the preference scores (1 means participant prefers argument C, 2 means participant prefers argument D and 3 means no preference between argument C and argument D) is given in Figure E.7. Preference scores were only available from participant group $G1$, by design.

Survey Results (Confidence Scores) - Cognitive Plausibility of Reinstatement																							
Participant Group G1																							
Participant	Sec 1, Q1	Sec 1, Q2	Sec 1, Q3	Sec 2, Q1	Sec 2, Q2	Sec 2, Q3	Sec 3, Q1	Sec 3, Q2	Sec 3, Q3	Sec 4, Q1	Sec 4, Q2	Sec 4, Q3	Sec 5, Q1	Sec 5, Q2	Sec 5, Q3	Sec 6, Q1	Sec 6, Q2	Sec 6, Q3	Gender	Education	Age	Proficiency in English	Nationality
1	7	6	7	6	7	7	7	7	4	6	7	7	6	7	6	7	7	7	Female	high school	18-21	Good	Syrian
2	5	6	3	4	6	2	7	6	1	7	1	2	5	6	4	2	5	7	Female	college diploma	18-21	Good	Iranian
3	7	7	7	7	5	5	7	5	4	7	5	5	7	5	4	7	5	5	Male	college bachelor	26-30	Very Good	Nigerian
4	7	7	7	7	7	7	7	1	1	7	4	7	1	1	7	7	7	6	Female	college diploma	18-21	Very Good	Iranian
5	4	4	4	6	5	6	6	6	6	6	5	4	7	4	4	4	4	7	Male	Master's degree	31-35	Very Good	Palestine
6	7	1	7	7	1	7	7	1	7	7	1	7	7	1	7	7	1	7	Male	college bachelor	31-35	Expert	Jordanian
7	7	7	7	7	1	6	7	1	4	7	3	7	7	3	7	7	1	1					
8	7	1	4	7	2	3	7	2	1	7	6	6	7	2	4	7	6	2	Male	college bachelor	41-45	Very Good	Mauritarian
9	4	4	1	4	4	7	4	4	7	4	4	7	4	4	7	7	7	7					
10	7	2	7	7	2	5	5	4	5	4	2	4	4	4	4	4	4	4	Female	college bachelor	22-25	Expert	Pakistani
11	7	4	7	6	5	4	4	4	4	4	4	4	7	4	5	6	4	6					
12	7	7	1	6	6	6	6	6	5	6	4	6	7	4	5	7	7	7	Female	college diploma	26-30	Good	Syrian
13	6	2	6	7	3	7	6	3	2	6	6	6	7	6	7	1	7	7	Female	college bachelor	18-21	Expert	Iranian
14	7	1	4	5	7	4	7	3	7	7	2	6	7	1	4	4	1	7	Male	college bachelor	18-21	Very Good	Indian
15	7	2	7	7	1	6	4	4	4	6	2	6	7	2	4	6	2	6	Female	college bachelor	26-30	Very Good	British
16	4	4	4	4	4	4	4	3	4	6	5	6	7	5	4	4	4	4	Female	college bachelor	26-30	Expert	Irish
17	1	2	1	7	2	7	6	2	6	7	4	5	7	2	6	2	1	2	Female	Master's degree	41-45	Very Good	
18	4	5	4	7	4	4	5	4	4	4	4	5	7	6	4	6	5	7	Female	college bachelor	22-25	Very Good	UAE
19	7	4	7	6	5	4	4	4	4	4	4	4	7	4	5	6	4	6	Female	college bachelor	31-35	Expert	Swiss
20	7	4	7	7	4	7	7	4	7	7	4	7	6	7	7	7	6	7	Male	Master's degree	31-35	Very Good	Pakistani

Figure E.1: Confidence Scores, Participant Group $G1$ - Standard Reinstatement

Survey Results (Confidence Scores) - Cognitive Plausibility of Reinstatement																	
Participant Group G2																	
Participant	Sec 1, Q1	Sec 1, Q2	Sec 2, Q1	Sec 2, Q2	Sec 3, Q1	Sec 3, Q2	Sec 4, Q1	Sec 4, Q2	Sec 5, Q1	Sec 5, Q2	Sec 6, Q1	Sec 6, Q2	Gender	Education	Age	Proficiency in English	Nationality
21	7	1	3	4	7	5	2	5	7	2	5	6	Female	college diploma	18-21	Good	Indian
22	6	2	7	4	3	5	6	4	4	5	6	6					
23	7	3	7	2	7	2	5	2	4	7	6	6	Female	college bachelor	36-40	Very Good	UAE
24	7	1	7	1	7	1	7	1	7	7	7	7	Female	college bachelor	26-30	Very Good	Iran
25	5	4	4	5	5	5	5	3	4	5	4	6					
26	7	1	2	5	2	5	5	3	7	5	6	7	Female	master's degree	31-35	Good	Filipino
27	7	2	6	2	5	4	5	6	7	5	5	5	Male	master's degree	51-55	Very Good	Pakistani
28	4	4	4	4	4	4	4	4	4	4	4	4	Male	college bachelor	26-30	Very Good	Jordanian
29	7	4	4	4	6	4	3	4	7	4	7	5	Female	college diploma	22-25	Good	Syrian
30	7	1	1	5	3	7	2	1	7	1	7	4	Male	college bachelor	22-25	Very Good	Nigerian
31	6	2	2	7	6	6	6	7	5	7	7	4	Female	high school	18-21	Very Good	Indian
32	6	4	6	6	6	2	2	5	3	7	7	7	Male	high school	18-21	Competent	Indian
33	4	4	4	5	6	4	4	3	7	6	6	7	Female	college bachelor	41-45	Expert	UAE
34	5	4	3	6	3	5	2	4	4	4	7	7	Female	master's degree	22-25	Competent	UAE
35	4	4	4	4	3	1	2	2	4	4	4	5					
36	4	4	6	1	4	4	4	1	4	4	6	4	Male	master's degree	26-30	Very Good	America
37	7	1	7	7	6	1	7	7	1	7	7	1	Male	master's degree	36-40	Expert	
38	6	4	5	5	5	4	5	4	5	4	7	4	Male	master's degree	31-35	Very Good	Indian

Figure E.2: Confidence Scores, Participant Group $G2$ - Standard Reinstatement

Survey Results (Confidence Scores) - Cognitive Plausibility of Floating Reinstatement																	
Participant Group G1																	
Particip ant	Group	Sec 1, Q1	Sec 1, Q2	Sec 1, Q3	Sec 2, Q1	Sec 2, Q2	Sec 2, Q3	Sec 3, Q1	Sec 3, Q2	Sec 3, Q3	Sec 4, Q1	Sec 4, Q2	Sec 4, Q3	Gender	Education	Age	Proficiency in English	Nationality
1	1	5	5	5	5	4	5	5	5	5	6	5	6	Female	College bachelor	18-21	Very Good	Indian
2	1	2	1	2	7	6	6	5	7	6	7	6	7	Female	High school	18-21	Very Good	Indian
3	1	3	4	5	5	7	5	6	4	3	5	3	4	Male	College bachelor	22-25	Good	Kyzgyz
4	1	2	4	3	5	6	4	7	4	1	7	7	3	Male	College bachelor	22-25	Good	Yzbek
5	1	4	4	4	2	5	4	4	1	3	2	2	3	Male	Doctoral	41-45	Expert	Pakistani
6	1	6	6	5	7	7	5	7	5	6	6	3	5	Male	High school	18-21	Good	Indian
7	1	1	3	4	1	3	4	4	3	1	1	3	3	Female	College bachelor	18-21	Expert	Indian
8	1	7	4	5	7	7	4	5	3	4	7	4	4	Female	College diploma	36-40	Very Good	Iranian
9	1	1	2	3	1	5	4	7	4	3	3	3	4	Male	High school	18-21	Limited	
10	1	7	1	1	1	4	3	4	2	1	1	1	6	Male	College bachelor	18-21	Competent	Indian
11	1	3	4	3	7	6	5	4	5	6	3	2	6	Female	College diploma	18-21	Modest	Korean
12	1	3	3	7	7	1	7	7	3	3	4	3	7	Female	College diploma	18-21	Competent	Korean
13	1	6	4	5	4	5	7	5	3	6	3	5	5	Female	College bachelor	18-21	Very Good	Indian
14	1	7	1	7	7	1	7	3	2	6	7	3	5	Female	Master's degree	22-25	Very Good	
15	1	7	1	7	7	1	1	6	3	6	7	4	4	Female	Master's degree	22-25	Very Good	Indian
16	1	7	2	2	5	1	2	5	6	7	4	4	5	Female	College bachelor	18-21	Very Good	Kazakh
17	1	7	2	2	7	6	6	1	1	2	7	6	6	Male	College bachelor	26-30	Very Good	Japanese
18	1	7	7	7	7	1	4	1	1	4	3	2	6	Female	College diploma	26-30	Good	Filipino
19	1	6	4	4	7	4	5	6	6	4	5	5	6	Female	College bachelor	18-21	Good	Indian
20	1	6	6	6	4	4	6	7	6	6	6	6	6	Female	Master's degree	31-35	Very Good	Kyrgyz
21	1	6	2	6	6	3	4	6	2	3	6	2	5	Female	Master's degree	36-40	Very Good	Egyptian
22	1	6	5	7	6	6	6	6	6	5	7	5	7	Male	College bachelor	26-30	Good	Jordan
23	1	7	4	4	3	4	3	7	7	4	7	4	5	Female	High school	18-21	Expert	Indian
24	1	7	1	2	4	7	7	7	1	7	7	7	7	Male	High school	18-21	Expert	Pakistani

Figure E.3: Confidence Scores, Participant Group $G1$ - Floating Reinstatement

Survey Results (Confidence Scores) - Cognitive Plausibility of Floating Reinstatement																		
Participant Group G2																		
Participant	Group	Sec 1, Q1	Sec 1, Q2	Sec 1, Q3	Sec 2, Q1	Sec 2, Q2	Sec 2, Q3	Sec 3, Q1	Sec 3, Q2	Sec 3, Q3	Sec 4, Q1	Sec 4, Q2	Sec 4, Q3	Gender	Education	Age	Proficiency in English	Nationality
25	2	2	2	4	7	7	7	4	3	4	2	3	2	Female	College bachelor	18-21	Good	Kazakh
26	2	4	4	4	7	1	5	7	7	5	7	1	7	Female	College bachelor	26-30	Very Good	UAE
27	2	7	7	7	7	1	7	6	3	7	7	7	7	Female	Master's degree	22-25	Very Good	
28	2	3	2	2	6	6	7	7	7	7	6	6	6	Male	Master's degree	26-30	Good	uae
29	2	4	7	5	3	5	3	3	6	6	6	6	6	Male		26-30	Good	
30	2	7	1	7	7	4	4	1	1	1	6	6	6	Male	Master's degree	36-40	Modest	Indian
31	2	7	6	7	7	4	6	7	4	4	4	4	4	Female	College diploma	22-25	Expert	
32	2	7	4	4	7	4	7	7	4	2	7	6	6	Female	College bachelor	22-25	Very Good	Jordan
33	2	7	4	6	7	7	4	7	5	2	5	6	6	Male	College bachelor	22-25	Modest	
34	2	7	7	6	7	7	7	1	7	4	7	4	7	Female	Master's degree	22-25	Good	Indian
35	2	6	4	3	6	4	1	7	1	7	7	4	7	Female	Master's degree	31-35	Good	Filipino
36	2	7	1	5	2	7	5	7	1	5	4	7	6					
37	2	7	7	7	6	4	2	7	1	7	7	7	7	Male	College bachelor	18-21	Expert	Pakistani
38	2	5	7	3	7	3	7	4	6	6	4	4	5	Male	College bachelor	18-21	Competent	Indian
39	2	1	5	1	6	5	1	4	1	7	5	2	7	Male	College bachelor	18-21	Very Good	Pakistani
40	2	5	4	1	5	7	1	4	2	4	5	4	4	Female	High school	18-21	Good	Kazakh
41	2	6	6	4	4	5	5	7	4	7	7	5	7	Female	Master's degree	26-30	Very Good	Syrian
42	2	5	3	4	6	3	3	4	1	5	4	1	5	Male	College bachelor	36-40	Good	Indian
43	2	4	4	4	7	4	4	4	4	4	4	4	4					
44	2	4	4	4	5	3	5	4	3	6	4	4	5	Male	College bachelor	31-35	Competent	Indian
45	2	4	5	1	5	5	5	4	3	5	4	3	6	Male	College bachelor	18-21	Competent	Indian
46	2	3	7	2	7	6	5	5	5	4	7	3	5	Male	Master's degree	22-25	Good	Uzbekistan
47	2	5	3	4	5	3	6	5	3	7	5	4	5	Female	Master's degree	26-30	Very Good	Jordan

Figure E.4: Confidence Scores, Participant Group $G2$ - Floating Reinstatement

Survey Results (Scores on Understanding) - Cognitive Plausibility of Floating Reinstatement

Participant Group G1

Participant	Group	Sec 1, Q1	Sec 1, Q2	Sec 1, Q3	Sec 2, Q1	Sec 2, Q2	Sec 2, Q3	Sec 3, Q1	Sec 3, Q2	Sec 3, Q3	Sec 4, Q1	Sec 4, Q2	Sec 4, Q3
1	1	5	5	3	2	4	3	3	4	3	3	3	4
2	1	7	5	5	5	5	5	4	5	4	5	5	5
3	1	3	4	3	3	5	4	4	3	3	4	5	5
4	1	2	1	1	5	3	4	5	4	4	5	5	5
5	1	3	2	1	3	2	4	1	2	1	2	1	1
6	1	4	3	3	4	5	4	4	4	4	5	4	4
7	1	1	4	2	2	2	3	1	6	3	4	1	4
8	1	5	5	5	5	5	5	3	3	5	4	4	4
9	1	2	3	3	5	6	7	5	4	3	4	4	4
10	1	7	7	4	7	4	3	1	2	2	1	4	5
11	1	4	4	4	5	5	5	4	4	3	3	5	1
12	1	4	4	5	6	6	5	5	4	1	3	4	2
13	1	5	6	4	5	4	2	3	5	4	3	1	1
14	1	4	5	4	5	5	5	5	5	4	5	5	4
15	1	6	3	5	6	3	3	5	3	3	7	3	2
16	1	5	3	4	4	3	4	3	3	4	3	3	3
17	1	6	6	6	6	6	5	6	6	6	6	6	6
18	1	7	7	7	7	7	5	6	6	3	3	5	5
19	1	4	4	4	6	4	4	6	5	4	4	5	5
20	1	5	5	5	5	5	5	5	5	5	5	5	5
21	1	5	5	5	6	3	5	5	5	5	2	3	4
22	1	5	6	4	5	7	5	5	4	5	7	4	5
23	1	6	4	4	6	7	3	5	5	4	7	4	5
24	1	6	7	4	7	7	6	7	7	5	5	6	6

Figure E.5: Scores on Understanding, Participant Group $G1$ - Floating Reinstatement

Survey Results (Scores on Understanding) - Cognitive Plausibility of Floating Reinstatement													
Participant Group G2													
Participant	Group	Sec 1, Q1	Sec 1, Q2	Sec 1, Q3	Sec 2, Q1	Sec 2, Q2	Sec 2, Q3	Sec 3, Q1	Sec 3, Q2	Sec 3, Q3	Sec 4, Q1	Sec 4, Q2	Sec 4, Q3
25	1	4	3	3	7	7	6	5	4	3	2	1	2
26	1	1	3	1	7	5	5	7	7	4	7	7	7
27	1	5	3	1	7	3	7	4	1	7	7	7	7
28	1	2	2	3	5	5	7	5	7	7	6	6	6
29	1	6	4	5	3	2	4	4	3	7	3	7	1
30	1	7	7	7	7	7	7	7	7	7	7	7	7
31	1	4	5	3	5	3	4	3	4	3	3	3	3
32	1	6	5	4	6	4	4	6	4	4	6	6	4
33	1	6	4	6	6	6	4	4	6	4	6	5	5
34	1	5	5	5	5	5	5	5	5	5	5	5	5
35	1	5	2	2	5	4	4	6	5	6	6	6	6
36	1	7	1	3	4	7	1	7	6	5	7	6	5
37	1	7	4	3	6	5	4	7	6	5	4	3	5
38	1	5	6	3	5	4	6	4	6	6	3	5	4
39	1	7	4	6	6	6	7	5	7	6	5	6	7
40	1	3	5	4	5	6	4	2	2	1	4	4	1
41	1	6	6	6	6	5	5	5	5	6	6	6	6
42	1	5	3	4	6	3	3	3	2	3	3	3	3
43	1	4	4	4	5	4	4	4	4	4	5	4	5
44	1	1	2	3	2	2	4	2	3	5	2	2	3
45	1	4	5	7	4	4	7	1	2	4	4	4	5
46	1	4	5	5	6	5	4	5	5	4	5	4	4
47	1	5	5	4	5	4	5	5	4	7	4	4	4

Figure E.6: Scores on Understanding, Participant Group $G2$ - Floating Reinstatement

Survey Results (Preference) - Cognitive Plausibility of Floating Reinstatement				
Participant Group G1				
Participant	Section 1	Section 2	Section 3	Section 4
1	3	2	2	2
2	3	3	3	3
3	3	2	1	3
4	2	1	2	3
5	2	1	2	1
6	3	3	3	3
7	1	3	1	2
8	1	3	3	3
9	2	3	1	1
10	1	2	2	1
11	3	3	3	3
12	3	3	3	2
13	3	1	2	1
14	3	1	1	3
15	3	3	1	3
16	3	1	1	3
17	3	1	1	2
18	1	1	1	1
19	3	3	3	3
20	3	1	3	3
21	3	3	1	3
22	1	3	3	2
23	1	3	2	1
24	1	1	2	2

Figure E.7: Scores on Preference, Participant Group $G1$ - Floating Reinstatement

Bibliography

[1] Bruno G. Bara, Monica Bucciarelli, and Vincenzo Lombardo. Model theory of deduction: a unified computational approach. *Cognitive Science*, 25:839–901, 2001.

[2] Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10–15):675–700, 2007.

[3] Trevor J. M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10–15):619–641, 2007.

[4] Philippe Besnard and Anthony Hunter. *Elements of Argumentation*. MIT Press, Cambridge MA, USA, 2008.

[5] Jean-Francois Bonnefon. Reinstatement, floating conclusions, and the credulity of mental model reasoning. *Cognitive Science*, 21:621–631, 2004.

[6] Martin D. Braine. On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85:1–21, 1978.

[7] Martin D. Braine and D. P. O'Brien. *Mental Logic*. Lawrence Erlbaum Associates, Mahwah NJ, USA, 1998.

[8] Monica Bucciarelli and P. N. Johnson-Laird. Strategies in syllogistic reasoning. *Cognitive Science*, 23(3):247–303, 1999.

[9] Ruth Byrne. Everyday reasoning with conditional reasoning. *Quarterly Experimental Psychology*, pages 141–165, 1989.

[10] Ruth Byrne. Can valid inferences be suppressed? *Cognition*, 39:71–78, 1991.

[11] Ruth Byrne. Counterexamples and the suppression of inferences. *Journal of Memory and Language*, 40:347–373, 1999.

[12] Ruth Byrne and Johnson-Laird. Syllogistic inference. *Cognition*, 16:1–61, 1984.

[13] Ruth Byrne and Johnson-Laird. A model point of view. *Thinking and Reasoning*, 1:339–350, 1995.

[14] Ruth Byrne and Johnson-Laird. Conditionals: A theory of meaning pragmatics and inference. *Psychological Review*, 109:646–678, 2002.

[15] Martin W. A. Caminada. A gentle introduction to argumentation semantics. Technical report, University of Luxembourg, Summer 2008.

[16] Allan Collins and Ryszard Michalski. The logic of plausible reasoning: A core theory. *Cognitive Science*, 13:1–49, 1989.

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

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

[19] C. L. Hamblin. *Fallacies*. Vale Press, Newport News, Virginia, 1970.

[20] John F. Horty. Skepticism and floating conclusions. *Artificial Intelligence*, 135(2002):55–72, 2001.

[21] P. N. Johnson-Laird, Ruth Byrne, and Walter Schaeken. Why models rather than rules give a better account of prepositional reasoning: A reply to bonatti and to o'brien, braine, and yang. *Psychological Review*, 101(4):734–739, 1994.

[22] P. N. Johnson-Laird, Ruth Byrne, and Patrizia Tabossi. Strategies in syllogistic reasoning. *Cognitive Science*, 23(3):247–303, 1999.

[23] P. N. Johnson-Laird, Walter Schaeken, and Ruth Byrne. Prepositional reasoning by model. *Psychological Review*, 99(3):418–439, 1992.

[24] Eric Lode. Slippery slope arguments and legal reasoning. *California Law Review*, 87:1469–1544, 1999.

[25] R. Loui. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 3:100–106, 1987.

[26] John McCarthy. Circumscription – a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.

[27] Mick Oaksford and Nich Charter. *Bayesian rationality: The probabilistic approach & reasoning.* Oxford University Press, 2007.

[28] Mike Oaksford. Probability logic and the modus ponens - modus tollens asymmetry in conditional inference. *Manuscript submitted for publication*, 2005.

[29] Mike Oaksford and Ulrike Hahn. A bayesian approach to informal argument fallacies. *Synthese*, 152(2):207–236, 2006.

[30] Mike Oaksford and Ulrike Hahn. Non-monotonicity and informal reasoning: Comment on ferguson (2003). *Argumentation*, 20(2):245–251, 2006.

[31] Mike Oaksford and Ulrike Hahn. Induction, deduction, and argument strength in human reasoning and argumentation. *Inductive reasoning*, pages 269–301, 2007.

[32] Mike Oaksford and Ulrike Hahn. The rationality of informal argumentation: A bayesian approach to reasoning fallacies. *Psychological Review*, 114(3):704–732, 2007.

[33] Francis Jefert Pelletier and Renee Elio. The case for psychologism in default and inheritance reasoning. *Synthese*, pages 1–29, 2005.

[34] Guy Politzer and Jean-Francois Bonnefon. Two varieties of conditionals and two kinds of defeaters help reveal two fundamental types of reasoning. *Mind & Language*, 21(4):484–503, 2006.

[35] John L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[36] Henry Prakken and Gerard Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer Academic Publishers, Dordrecht, Netherlands, second edition, 2002.

[37] Iyad Rahwan and Peter McBurney. Guest editors' introduction: Argumentation technology. *IEEE Intelligent Systems*, 22(6):21–23, 2007.

[38] Iyad Rahwan and Guillermo R. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, New York, USA, 2009.

[39] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[40] Lance J. Rips. *The psychology of proof: deductive reasoning in human thinking.* MIT Press, Cambridge MA, USA, 1994.

[41] Stuart J. Russell and Peter Norvig. *Artificial intelligence: a modern approach.* Prentice Hall, Upper Saddle River, NJ, USA, second edition, 2003.

[42] Edward E. Smith, Christopher Langton, and Richard E. Nisbett. The case for rules in reasoning. *Cognitive Science*, 16:1–40, 1992.

[43] Douglas Walton. *Arguments from Ignorance.* Cambridge University Press, New York, USA, 2006.