Human Compliance with Normative Principles in Argumentation: Effects of Naturalness Bias and Graphical Representation

Marija Petrović, Predrag Teovanović, Danka Purić, Bruno Yun, Caren Al Anaissy, Sébastien Konieczny, Srdjan Vesic

Human Compliance with Normative Principles in Argumentation: Effects of Naturalness Bias and Graphical Representation

Marija Petrović[1,2], Predrag Teovanović[2,3], Danka Purić[1,2], Bruno Yun[5], Caren Al Anaissy[6], Sébastien Konieczny[4] and Srdjan Vesic[4]

[1]University of Belgrade, Faculty of Philosophy, Department of Psychology, 18-20 Čika Ljubina Street, 11000 Belgrade, Serbia

[2]University of Belgrade, Faculty of Philosophy, LIRA Lab, 18-20 Čika Ljubina Street, 11000 Belgrade, Serbia

[3]FASPER - Faculty of Special Education and Rehabilitation, Visokog Stevana 2, 11000 Belgrade, Serbia

[4]CRIL – CNRS – Univ. Artois

[5]Universite Claude Bernard Lyon 1, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Lumière Lyon 2, LIRIS, UMR5205, 69622 Villeurbanne, France

[6]CRIL Université d'Artois & CNRS, France

Abstract

Argumentation theory examines how conclusions are derived or refuted through logical reasoning, playing a crucial role in human interaction and decision-making. In artificial intelligence, computational argumentation leverages formal models to aid in decision-making processes. This paper investigates the influence of argument content (specifically the naturalness bias) and graphical representation on participants' adherence to the simple principles of reinstatement and void precedence principles. We conducted experiments testing three hypotheses related to participants' rationality in evaluating arguments with and without graphical aids and bias-provoking content. Contrary to our expectations, neither the presence of graphical representations nor the type of content significantly impacted participants' compliance with the reinstatement principle. Additionally, the graph did not enhance understanding, suggesting the need for instructional aids. Our findings challenge previous studies and highlight the complexity of factors influencing argument evaluation.

Keywords

Argumentation, Bias, Ranking-Based Semantics, Human Reasoning, Principles

ArgXAI-24: 2nd International Workshop on Argumentation for eXplainable AI

maka.petrovic@gmail.com (M. Petrović); teovanovic@fasper.bg.ac.rs (P. Teovanović); dpuric@f.bg.ac.rs (D. Purić); bruno.yun@univ-lyon1.fr (B. Yun); alanaissy@cril.fr (C. A. Anaissy); konieczny@cril.fr (S. Konieczny); vesic@cril.fr (S. Vesic)

https://lira.f.bg.ac.rs/marija-petrovic-en/ (M. Petrović); https://lira.f.bg.ac.rs/phd-predrag-teovanovic/ (P. Teovanović); https://lira.f.bg.ac.rs/phd-danka-puric/ (D. Purić); https://liris.cnrs.fr/page-membre/bruno-yun (B. Yun); https://www.cril.univ-artois.fr/~alanaissy/ (C. A. Anaissy); https://www.cril.fr/~konieczny/ (S. Konieczny); https://www.cril.univ-artois.fr/~vesic/ (S. Vesic)

0000-0001-7910-4213 (M. Petrović); 0000-0003-3477-6723 (P. Teovanović); 0000-0001-5126-3781 (D. Purić); 0000-0001-9370-3917 (B. Yun); 0000-0002-8750-1849 (C. A. Anaissy); 0000-0002-2590-1222 (S. Konieczny); 0000-0002-4382-0928 (S. Vesic)

1. Introduction

Argumentation theory is the interdisciplinary study of how conclusions can be supported or undermined by premises through logical reasoning. It is an essential part of our daily human reasoning, interaction and communication with one another. Argumentation is often used in decision-making problems, resolving conflicts of opinion through negotiation and deliberation processes, and influencing the thoughts of others through persuasion. The complex and dynamic nature of argumentation is studied in various disciplines.

It stands as an important domain of Artificial Intelligence (AI), especially in knowledge representation and reasoning. Indeed, the use of AI allows for the development of computational models for the exchange of arguments among various agents, facilitating the derivation of valid conclusions from incomplete, inconsistent or conflicting information. Within AI, computational argumentation focuses on creating formal models that support decision-making through the construction and evaluation of arguments. Many works in computational arguments are based on Dung's seminal work [1] where he proposed the *abstract argumentation framework*, which is a general tool for studying relations between the arguments such as attacks (and, more recently, supports [2, 3] or sets of attacking arguments [4, 5]).

On a given abstract argumentation framework, one can apply many different semantics to compute sets of acceptable arguments (conflict-free, admissible, ...) [1, 6] or rank arguments from the strongest to the weakest (using ranking-based semantics) [7]. Those semantics are based on many intuitions or normative principles, defined in the literature, which characterise their behaviour. While some work has studied the link between basic intuitions from the argumentation theory (e.g., reinstatement) and human reasoning [8, 9, 10], it is still unknown if most of the principles are intuitive or used by humans.

A recent work by Vesic, Yun, and Teovanovic [11] studied some principles for ranking-based semantics (anonymity, void precedence, maximality, and independence) and showed that the graphical representation influences the normativity of participant responses. Namely, there was a higher level of compliance with the principles when the participants were shown the graphical representation of the arguments. Moreover, they showed that anonymity between tasks principle was not followed by the participants. In other words, for two different sets of textual arguments with same structure (the graphs are isomorphic), the participants differed in their evaluation of the arguments, although this effect was dampened for participants who were shown the graphical representation. This demonstrates that both the content of the arguments and the graphical representation have a significant effect on the evaluation task.

The latter result opens up several important questions that were not studied before. In what way does the content of the arguments (e.g., language, bias-provoking content) affect the principle compliance? Can graphical representations consistently serve as a prescriptive tool for enhancing rationality in individuals?

In this exploratory paper, we will focus on a restricted setting. We will study whether participants comply with the *simple reinstatement principle* and how this is affected by the *naturalness bias* and the *graphical representation.*

On the one hand, the simple reinstatement principle states that in a configuration with three arguments A, B, and C such that C attacks B and B attacks A (see Figure 1), roughly speaking "both C and A must be stronger than B". On the other hand, the naturalness bias (also known

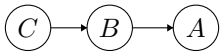

Figure 1: Simple reinstatement setting.

as appeal-to-nature) is a cognitive bias where people have a preference for things perceived as "natural" over those perceived as "unnatural" or artificial. This bias can influence attitudes and decisions in various domains, including food, medicine, or lifestyle choices, among others.

To experimentally check the impact of the content (e.g. bias-provoking text) and the graph, we set up an experiment featuring three hypotheses. Firstly, we expected participants to behave more rationally in the absence of the naturalness bias, regardless of graph presence. Next, we predicted that the graph would make participants more rational (as shown in previous studies) and we also expected that the effect of the graph would be stronger in the presence of naturalness bias than in its absence.

This paper is structured as follows. In Section 2, we recall the computational argumentation background on extension and ranking-based semantics needed to motivate the principles. In Section 3, we formulate our aims and hypothesis with respect to the principles, the graph, and the naturalness bias. In Section 4, we describe the design of our study, the sampling plan, and the instruments. Lastly, we present our results in Section 5 and conclude in Section 6.

2. Argumentation Background

We start this section by recalling the definition of an argumentation framework as defined by Dung in his seminal paper [1] (as an example, see Figure 1).

Definition 1 (Argumentation framework). *An argumentation framework is a pair* $\mathcal{F} = (\mathcal{A}, \mathcal{C})$*, where* $\mathcal{A}$ *is a finite set of arguments and* $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$ *is a set of binary attacks between arguments. The set of attackers of* $a \in \mathcal{A}$ *is* $Att(a) = \{b \in \mathcal{A} \mid (b, a) \in \mathcal{C}\}$.

In the following sub-sections, we explain what is considered as rational with respect to extension-based semantics and ranking-based semantics.

2.1. Extension-based semantics

Given an argumentation framework $\mathcal{F} = (\mathcal{A}, \mathcal{C})$, we say that a set $S \subseteq \mathcal{A}$ is conflict-free iff there is no $x, y \in S$ such that $(x, y) \in \mathcal{C}$. A set $S \subseteq \mathcal{A}$ defends $a \in \mathcal{A}$ iff for every $b \in \mathcal{A}$ such that $(b, a) \in \mathcal{C}$, there exists $s \in S$ such that $(s, b) \in \mathcal{C}$. A set $S \subseteq \mathcal{A}$ is admissible iff it is conflict-free and defends every argument in S. A set S is preferred iff it is a maximal (for set inclusion) admissible set.

The *preferred semantics* is the function that computes all the preferred sets from an argumentation framework. There are many other acceptability semantics, which we do not introduce in this paper as the details are not of practical interest for the rest of our cognitive study[1].

[1]Please refer to [6] for an introduction to extenion-based semantics.

Example 1. *In Figure 1, there are three admissible sets, which are* $\emptyset, \{C\}$, *and* $\{C, A\}$, *but there is only one preferred set which is* $\{C, A\}$.

In Example 1, the preferred extension shows that both A and C are better than B (as A and C are in the preferred set but B is not). In this case, we can consider that A and C should be equally acceptable. Thus, we define the *extension-based simple reinstatement*, inspired by [12, 13], as follows.

Definition 2 (Extension-based simple reinstatement). *In a configuration with three arguments* $A, B,$ *and* C *such that* C *attacks* B *and* B *attacks* A, *the extension-based simple reinstatement is satisfied iff* C *is equally strong as* A, C *is stronger than* B *and* A *is stronger than* B.

While extension can extract sets of acceptable arguments, an orthogonal approach consists in ranking-arguments from the stronger to the weakest one.

2.2. Ranking-based semantics

Another family of argumentation semantics is called ranking-based. They do not calculate extensions; instead, they rank the arguments from the strongest to the weakest one. In this paper, we focus on a particular family of ranking-based semantics, called gradual semantics, which associates to each argument, a number from 0 (weakest) to 1 (strongest). Those scores naturally induce an order on the set of arguments.

Definition 3 (Gradual semantics). *A gradual semantics is a function* σ *that takes as input any* $\mathcal{F} = (\mathcal{A}, \mathcal{C})$ *and returns a function* $Deg^{\sigma}_{\mathcal{F}} : \mathcal{A} \rightarrow [0, 1]$. *The notation* $Deg^{\sigma}_{\mathcal{F}}(a) \leq Deg^{\sigma}_{\mathcal{F}}(b)$ *means that* b *is at least as acceptable as* a *w.r.t.* σ.

For the rest of our study, we do not need particular semantics. However, just for the sake of illustration, we introduce one semantics, called h-categorizer [14]. We refer the reader to the work of [15] for a comparative study on ranking-based semantics.

Definition 4 (h-categorizer). *The* h*-categorizer semantics is a ranking-based semantics such that for every* $\mathcal{F} = (\mathcal{A}, \mathcal{C})$, *for every argument* $a \in \mathcal{A}$ *we have*

$$Deg^{h}_{\mathcal{F}}(a) = \frac{1}{1 + \sum_{b \in Att(a)} Deg^{h}_{\mathcal{F}}(b)}$$

It was shown [16] that h-categorizer is well-defined for every argumentation framework (i.e., the semantics converge to unique values for each argumentation framework). To illustrate, the values assigned by h-categorizer to the arguments from Figure 1 are: $Deg^{h}_{\mathcal{F}}(C) = 1$, $Deg^{h}_{\mathcal{F}}(B) = \frac{1}{2}$, and $Deg^{h}_{\mathcal{F}}(C) = \frac{2}{3}$.

Many principles were defined for characterising the behaviour of ranking-based semantics. For the purpose of this paper, we do not need to recall them all. We just formalize the ranking-based simple reinstatement and the void precedence principles.

Definition 5 (Ranking-based simple reinstatement). *In a configuration with three arguments A, B, and C such that C attacks B and B attacks A, the ranking-based simple reinstatement is satisfied iff C is stronger than A and A is stronger than B.*

We consider that both extension-based and ranking-based are valid and rational points of view. In the rest of the paper, we allow the rational agent to act based on either of them. Hence, we will say that the reinstatement is satisfied iff either ranking-based simple reinstatement or extension-based simple reinstatement is satisfied.

Definition 6 (Simple reinstatement). *In a configuration with three arguments A, B, and C such that C attacks B and B attacks A, the simple reinstatement is satisfied iff C is equal or stronger than A and A is stronger than B.*

Void precedence states that a non-attacked argument is stronger than an attacked one.

Definition 7 (Void precedence [15]). *We say that a ranking-based semantics σ satisfies void precedence iff for every argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{C})$ and $a, b \in \mathcal{A}$ such that $Att(a) = \emptyset$ and $Att(b) \neq \emptyset$ then $Deg^{\sigma}_{\mathcal{F}}(a) > Deg^{\sigma}_{\mathcal{F}}(b)$.*

Now that we have defined what is considered rational in abstract argumentation, we formally describe our aims and hypotheses.

3. Aim and Hypotheses

In the current study, we examined if people comply with normative principles more readily when presented with arguments with neutral content in comparison to arguments that contain bias-provoking content. To do so, we explored people's response patterns in a simple reinstatement scenario. As an example of bias-provoking content, we relied on the naturalness bias i.e., the tendency to prefer natural things over artificial ones, all other things being equal. Furthermore, we aimed to replicate previous findings that graphical representation of arguments leads to higher compliance with normative principles i.e., the presence of an argumentation graph facilitates rational inference.

We expected participants to comply more with normative principles in the neutral condition than in the naturalness bias condition, regardless of graph presence (H1). Moreover, we expected participants to comply more with normative principles in the graph condition relative to the no-graph condition, regardless of content (H2). Finally, we expected an interaction between the two factors, i.e., that the effect of the graph will be stronger in the naturalness bias condition relative to the neutral condition (H3).

On an exploratory level, we also checked whether people tend to follow extension-based or ranking-based argumentation semantics more, depending on the rating of different arguments (the non-attacked one and the reinstated one). The extension-based semantics predict that both arguments will have the same acceptability level, whereas most of the ranking-based semantics predict that the non-attacked one will be stronger.

4. Method

4.1. Open science

The materials and data to reproduce the findings of this study can be found on the project OSF page: https://osf.io/kce7q/.

4.2. Study design

We employed a 2x2 design with two independent factors: content (neutral vs. bias-provoking) and graph (present vs. absent), resulting in four experimental groups of participants.

4.3. Sampling plan and data exclusions

The sample size was based on an a priori power analysis. Since we considered several possible effects, we opted to base the power analysis on the smallest effect size of interest - i.e., a small interaction effect size (Cohen's $\omega = .10$). For power .80, the apriori analysis indicated we would need $n = 197$ per group (788 in total) to detect this effect size. For power .90, for the same effect size, we would need $n = 263$ per group (1052 in total). For further details on the power analysis, see preregistration - https://aspredicted.org/9T9_NG4.

As per preregistration, we included an attention check ("Please choose Completely agree to indicate you are paying attention"), and all participants who failed the attention check were automatically excluded from the sample and did not count towards the final sample size.

4.4. Final sample and procedure

The final sample size consisted of a total of $N = 1048$ UK and US Prolific participants. All respondents were compensated for their participation using the standard Prolific 9 £/hour rate. Most participants were from the UK (85.6%), and 61.8% were female, with an average age of $Mage = 41.45$ ($SDage = 13.71$)[2]. Due to a technical error in the questionnaire setup, there was an unequal distribution between the four experimental groups (116 for neutral content without a graph, 398 for neutral content with a graph, 400 for bias-provoking content without a graph, and 133 for bias-provoking content with a graph).

The questionnaire was administered online via Prolific and hosted on the SoSci Survey platform [17]. The participants were randomly allocated first to either the neutral or biased-content condition, and then to either the graph or no-graph condition. The participants then saw the arguments (and the graph, depending on the condition) and rated them, after which they answered an additional question on their strategy when rating the arguments. The planned duration of the study was four minutes, but on average, participants completed it in 100 seconds ($SD = 43.3s$).

[2]Based on the data from $N = 1046$ participants, since two participants' demographic Prolific data could not be matched.

Consider the arguments below.
- Argument A: It doesn't matter if you drink natural or lab-produced water.
- Argument B: Natural things are healthier than lab-produced ones, so you should drink natural water.
- Argument C: They are chemically identical, so they have the same impact on your health.

Figure 2: Textual description of the arguments shown to the bias-provoking groups.

Consider the arguments below.
- Argument A: It's not going to rain today.
- Argument B: But look at all these clouds, it will rain.
- Argument C: Those clouds are Cumulus, hence they do not produce rain.

Figure 3: Textual description of the arguments shown to the neutral groups.

4.5. Study instruments

As mentioned in Section 4.2, we had four experimental groups of participants (neutral vs. bias-provoking / graph vs. no-graph).

Each participant was first shown three arguments in their textual forms. Depending on their group (neutral vs. bias-provoking), the textual content was different (see Figures 2 and 3). We can observe that in both cases, argument B attacks argument A which consists only of a claim, argument C attacks argument B since C attacks B' s premise.

The textual content of the arguments was inspired by structured argumentation which allows for constructing arguments from a knowledge base using a formal language. In this context, arguments are characterized as structured because they reveal their underlying premises and conclusions clearly, the connection between them is formally established, and attacks among them are formally defined. There exist several approaches in structured argumentation for formalizing arguments such as the ASPIC$^+$ framework [18], deductive argumentation [19] and the assumption-based framework [20]. We represented arguments in their simplest forms which consists of a premise that supports a conclusion (claim) using an inference rule. In some cases, the premise can be hidden (e.g., in argument A), hence the argument consists only of a claim. Based on the internal structure of the arguments, we can have different types of attack relations between the arguments. For example, an argument can attack another argument' s premise; this type of attack is called undermining attack (e.g., attack from C to B). Another type of attack, called rebutting attack, is an attack on the conclusion of an argument (e.g., attack from B to A).

Only for the participants in the graph groups, we displayed the graphical visualisation of the arguments and attacks (see Figure 1).

Lastly, the participants were asked to assess the strength of each argument using a 5-point Likert scale from 1 (Weak) to 5 (Strong), with 3 being "Neutral".

5. Results

Overall, participants complied with simple reinstatement to a low degree (see Figure 4), with the percentage of participants who followed these principles in either of the groups ranging from 10.3% to 16.5%.

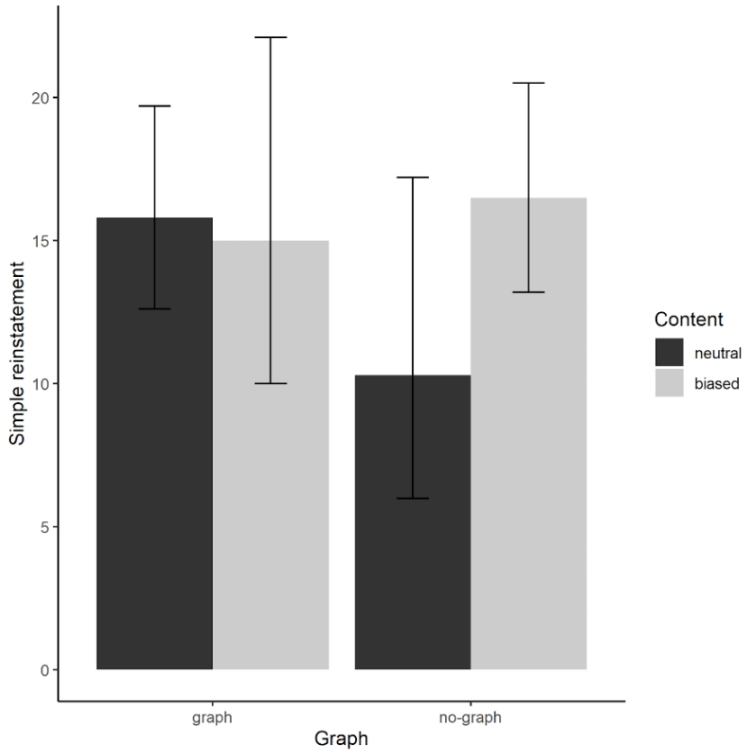

Figure 4: Percentages of responses that align with simple reinstatement across the four groups with 95% confidence intervals.

We conducted a logistic regression analysis to investigate how well the two factors (content and graph) and their interaction predict compliance with simple reinstatement. The goodness-of-fit of the logistic regression model was assessed using a likelihood ratio test, which yielded a non-significant chi-square statistic, $\chi^2(3) = 2.96, p = .398$. Additionally, the R^2 value indicated that the model accounts for only 0.3% of the variability in compliance with simple reinstatement. This suggests that the model with argument content and graphical representation did not provide a significantly better fit to the data than a model without them, as shown in Table 1.

Regarding void precedence, the participants in our sample followed this principle to a larger degree (see Figure 5) - the lowest percentage in either of the groups was 44.4% while the highest was 69.8%.

To explore if the two factors and their interaction predict compliance with void precedence, we conducted another logistic regression analysis. The goodness-of-fit of the model was assessed using a likelihood ratio test, which yielded a significant chi-square statistic, $\chi^2(3) = 53.66, p <$

Predictor	B	SE_B	Wald(1)	p
Intercept	-3.78	1.30	8.45	.004
Content	1.14	.72	2.48	.116
Graph	1.09	.72	2.26	.133
Content * Graph	-0.60	.43	1.90	.168

Table 1
Results of Binary Logistic Regression Analysis on Compliance with Simple Reinstatement.

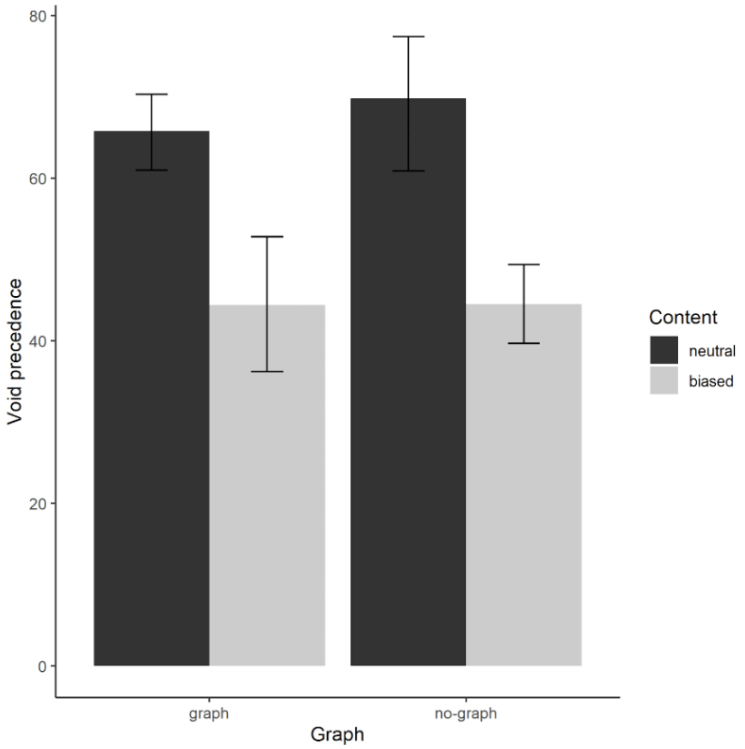

Figure 5: Percentages of responses that align with void precedence across the four groups with 95% confidence intervals.

.001, and R^2 value indicated that the model accounts for 5% of the variability in compliance with void precedence. As shown in Table 2, content influenced rational behavior, while the graph and interaction terms were non-significant. Presenting arguments with naturalness bias-provoking content decreased compliance with void precedence by 71% ($exp(B) = 0.29, 95\%CI = [0.11, 0.77]$).

Further analyses revealed that the effect of content is due to higher ratings of argument B ($t(1057) = 2.66, p = .008, d = 0.16$) and especially lower ratings of argument C ($t(1057) = 7.69, p < .001, d = 0.47$) in the group presented with bias-provoking content, regardless of the presence of graph (see Table 3). Ratings for argument A did not differ significantly between bias-provoking and neutral content groups ($p = .28$).

Predictor	B	SE_B	Wald(1)	p
Intercept	-2.26	0.88	6.63	.010
Content	-1.24	0.50	6.23	.013
Graph	-0.36	0.50	0.52	.47
Content * Graph	0.18	0.30	0.34	.56

Table 2
Results of Binary Logistic Regression Analysis on Compliance with Void Precedence.

Condition	Argument A	Argument B	Argument C
Neutral no-graph	2.30 (0.088)	3.16 (0.079)	4.14 (0.084)
Neutral graph	2.34 (0.053)	3.17 (0.045)	4.12 (0.045)
Biased no-graph	2.36 (0.047)	3.32 (0.055)	3.67 (0.053)
Biased graph	2.47 (0.090)	3.32 (0.091)	3.57 (0.102)

Table 3
Mean ratings (with standard errors) for three arguments in four experimental conditions.

6. Discussion

We now summarize and discuss all of our hypotheses and results.

First, we expected participants to comply more with normative principles in the neutral condition than in the naturalness bias condition, regardless of graph presence (H1). For this hypothesis, our results are mixed. The experiment showed that the change in content (i.e. neutral vs biased) did not impact the degree of compliance with the simple reinstatement principle. However, the content influenced the degree to which participants complied with the void precedence principle. Namely, in the presence of arguments incorporating naturalness bias, the ratings of the non-attacked argument C were very low while the ratings of the non-defended argument B were high (probably due to the naturalness bias increasing participants' ratings of B).

Second, we expected participants to comply more with normative principles in the graph condition relative to the no-graph condition, regardless of content (H2). This hypothesis was disconfirmed - our experiment showed that the graph did not provide a significantly better fit to the data. This contrasts the recent results [11] where the presence of the graphical representation was shown to enhance compliance with principles for graded argumentation semantics. Note, however, that the participants in that study went through a tutorial and several tasks before evaluating the strength of the arguments. This hints that a tutorial may be necessary to teach participants the meaning of the graphical representation and how it is linked to argument ratings.

Third, we expected an interaction between the two factors, i.e., that the effect of the graph will be stronger in the naturalness bias condition relative to the neutral condition (H3). Contrary to our hypothesis, the experiment showed that the interaction effect was not significant in our model.

In our experiments, compliance with the reinstatement principle was very low (less than 20%). Our results differ from the existing empirical work on reinstatement [21]. However, the task that participants were presented with is also different. While we study the individual satisfaction of

the principle considering all arguments simultaneously, they compared the average ratings of argument A (before B is introduced, after B is introduced, and then when C is introduced).

Note that according to the average scores of arguments in our experiment, C is stronger than B and B is stronger than A. One of the possible explanations for this is that the argument A is just a statement and contains no justification. Argument B has a general justification (e.g., there are clouds so it will rain) and might be seen as more acceptable for this reason. Finally the argument C has the strongest hypothesis that fully justifies its conclusion and might seem difficult to attack (e.g., the cumulus clouds do not produce rain).

Another reason why compliance with simple reinstatement might be low is that participants' intuition about the notion of defense diverges significantly from that of the researchers. More precisely, it might be that the participants misunderstood the meaning of attacks and associated attacks solely with a reduction in argument B strength, ignoring the increase in strength brought by reinstatement for argument A.

Importantly, several factors limit the generalizability of our findings. Firstly, due to a programming error, the four groups in our study were not equal in size, somewhat reducing statistical power. Next, since participants only responded to one presented scenario, it is difficult to say how they would respond to other scenarios, be it simple reinstatement scenarios with different content or more complex scenarios. Finally, we presented all arguments at the same time, which may have an impact on participants' perception of attacks and, consequently, their ratings of arguments' strength.

In future work, we plan to introduce the arguments step-by-step (as it was done by Rahwan et al. [21]) while monitoring the ratings of all arguments (instead of only A). We believe that this will give us a better insight on the factors at play behind the participants' ratings of the arguments.

Acknowledgments

This work benefited from the support of the project SATTORI in the framework of the Pavle Savic - Hubert Curien funding scheme, and from the support of the project AGGREEY ANR-22-CE23-0005 of the French National Research Agency (ANR).

References

[1] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (1995) 321–358. URL: https://doi.org/10.1016/0004-3702(94)00041-X. doi:10.1016/0004-3702(94)00041-X.

[2] C. Cayrol, M. Lagasquie-Schiex, Bipolarity in argumentation graphs: Towards a better understanding, Int. J. Approx. Reason. 54 (2013) 876–899. URL: https://doi.org/10.1016/j.ijar.2013.03.001. doi:10.1016/J.IJAR.2013.03.001.

[3] A. Cohen, S. Gottifredi, A. J. García, G. R. Simari, A survey of different approaches to support in argumentation systems, Knowl. Eng. Rev. 29 (2014) 513–550. URL: https://doi.org/10.1017/S0269888913000325. doi:10.1017/S0269888913000325.

[4] B. Yun, S. Vesic, M. Croitoru, Ranking-based semantics for sets of attacking arguments, in: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, AAAI Press, 2020, pp. 3033–3040. URL: https://doi.org/10.1609/aaai.v34i03.5697. doi:10.1609/AAAI.V34I03.5697.

[5] S. H. Nielsen, S. Parsons, A generalization of dung's abstract framework for argumentation: Arguing with sets of attacking arguments, in: N. Maudet, S. Parsons, I. Rahwan (Eds.), Argumentation in Multi-Agent Systems, Third International Workshop, ArgMAS 2006, Hakodate, Japan, May 8, 2006, Revised Selected and Invited Papers, volume 4766 of *Lecture Notes in Computer Science*, Springer, 2006, pp. 54–73. URL: https://doi.org/10.1007/978-3-540-75526-5_4. doi:10.1007/978-3-540-75526-5_4.

[6] P. Baroni, M. Caminada, M. Giacomin, An introduction to argumentation semantics, Knowl. Eng. Rev. 26 (2011) 365–410. URL: https://doi.org/10.1017/S0269888911000166. doi:10.1017/S0269888911000166.

[7] L. Amgoud, J. Ben-Naim, Ranking-based semantics for argumentation frameworks, in: W. Liu, V. S. Subrahmanian, J. Wijsen (Eds.), Scalable Uncertainty Management - 7th International Conference, SUM 2013, Washington, DC, USA, September 16-18, 2013. Proceedings, volume 8078 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 134–147. URL: https://doi.org/10.1007/978-3-642-40381-1_11. doi:10.1007/978-3-642-40381-1_11.

[8] M. Cramer, M. Guillaume, Empirical study on human evaluation of complex argumentation frameworks, in: F. Calimeri, N. Leone, M. Manna (Eds.), Logics in Artificial Intelligence - 16th European Conference, JELIA 2019, Rende, Italy, May 7-11, 2019, Proceedings, volume 11468 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 102–115. URL: https://doi.org/10.1007/978-3-030-19570-0_7. doi:10.1007/978-3-030-19570-0_7.

[9] M. Cramer, M. Guillaume, Empirical cognitive study on abstract argumentation semantics, in: S. Modgil, K. Budzynska, J. Lawrence (Eds.), Computational Models of Argument - Proceedings of COMMA 2018, Warsaw, Poland, 12-14 September 2018, volume 305 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2018, pp. 413–424. URL: https://doi.org/10.3233/978-1-61499-906-5-413. doi:10.3233/978-1-61499-906-5-413.

[10] E. Bezou Vrakatseli, H. Prakken, C. Janssen, L. Amgoud, R. Booth, et al., New experiments on reinstatement and gradual acceptability of arguments, in: Proceedings of the 19th International Workshop on Nonmonotonic Reasoning, 2021, pp. 109–118.

[11] S. Vesic, B. Yun, P. Teovanovic, Graphical representation enhances human compliance with principles for graded argumentation semantics, in: P. Faliszewski, V. Mascardi, C. Pelachaud, M. E. Taylor (Eds.), 21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022, Auckland, New Zealand, May 9-13, 2022, International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2022, pp. 1319–1327. URL: https://www.ifaamas.org/Proceedings/aamas2022/pdfs/p1319.pdf. doi:10.5555/3535850.3535997.

[12] L. van der Torre, S. Vesic, The principle-based approach to abstract argumentation semantics, FLAP 4 (2017). URL: http://www.collegepublications.co.uk/downloads/ifcolog00017.pdf.

[13] P. Baroni, M. Giacomin, On principle-based evaluation of extension-based argumentation

semantics, Artif. Intell. 171 (2007) 675–700. URL: https://doi.org/10.1016/j.artint.2007.04.004. doi:10.1016/J.ARTINT.2007.04.004.

[14] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, Artificial Intelligence 128 (2001) 203–235.

[15] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, A comparative study of ranking-based semantics for abstract argumentation, in: D. Schuurmans, M. P. Wellman (Eds.), Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA, AAAI Press, 2016, pp. 914–920. URL: https://doi.org/10.1609/aaai.v30i1.10116. doi:10.1609/AAAI.V30I1.10116.

[16] F. Pu, J. L. andY. Zhang, G. Luo, Argument ranking with categoriser function, in: International Knowledge Science, Engineering and Management Conference KSEM'14, 2014, pp. 290–301.

[17] D. J. Leiner, Sosci survey (version 3.2. 31)[computer software], München: SoSci Survey GmbH (2021).

[18] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artificial Intelligence 195 (2013) 361–397.

[19] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, Artificial Intelligence 128 (2001) 203–235.

[20] P. M. Dung, R. A. Kowalski, F. Toni, Assumption-based argumentation, Argumentation in artificial intelligence (2009) 199–218.

[21] I. Rahwan, M. I. Madakkatel, J. Bonnefon, R. N. Awan, S. Abdallah, Behavioral experiments for assessing the abstract argumentation semantics of reinstatement, Cogn. Sci. 34 (2010) 1483–1502. URL: https://doi.org/10.1111/j.1551-6709.2010.01123.x. doi:10.1111/J.1551-6709.2010.01123.X.