From Psychological Persuasion To Abstract Argumentation: A Step Forward

Jean-Baptiste Corrégé[1] and **Emmanuel Hadoux**[2] and **Ariel Rosenfeld**[3]

Abstract. Developing argumentation-based persuasive agents that leverage human argumentative techniques is an open challenge in the computational argumentation field. In this paper, we propose a computational perspective on the psychological techniques people tend to follow during persuasion interactions drawing on psychological evidence. We focus on four well-established psychological techniques, model and investigate them using a recently proposed argumentative computational framework. Our investigation reveals both similarities and gaps between the two which can be either leveraged or addressed in the design of argumentation-based persuasive agents and future theoretical developments.

1 Introduction

A key human skill, used across many domains and activities, is the ability to *persuade*. Politicians strive to persuade their constituents, parents try to persuade their children to eat healthier food, etc. People use many different techniques for persuading others. These *human persuasive techniques* have been thoroughly investigated in the *real world* by psychology researchers. Surprisingly, despite the major advancements of the computational **argumentation theory**, providing grounded techniques and models analysed and tested in theoretical settings, the study of the possible connections between human persuasive techniques and computational models has yet to be properly examined.

In this work we provide a novel investigation of the connections between psychological persuasion literature and argumentation theory. Through this tentative investigation we are able to identify the potential use of psychological persuasive principles in argumentation-based systems and find potential directions for future work in adapting and/or extending current argumentative principles to correctly account for psychological persuasion literature. *Our findings contribute an additional stage in the greater challenge of bridging the gap between argumentation theory and people.*

The paper is structured as follows: In Section 2 we survey related works which tried to bridge the gap between psychology and argumentation, coming from both sides. We also review the *Weighted Attack/Support Argumentation graphs* [14] and the necessary definitions used in this paper. In Section 3, we discuss four well-established psychological persuasive techniques. For each technique, we present the idea underlying the technique along with supportive evidences validating the technique. Finally, in Section 4, we model the psychological persuasive techniques discussed in Section 3 using abstract argumentation and evaluate the resulting model.

[1] LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay
[2] Department of Computer Science, University College London, London, UK
[3] Weizmann Institute of Science, Rehovot, Israel

2 Background

Within the computational argumentation field, a significant effort has been placed on proposing and evaluating models and techniques aimed at allowing an automated agent (*i.e.*, *persuader*) to persuade a person (*i.e.*, *persuadee*). Theoretically, an agent would seek to deploy an *optimal* persuasive policy, mapping each possible state of a dialogue to the *best* argument for the agent to present. This persuasive policy may strive to maximize different objectives:

- likelihood of having a specific set of arguments (*i.e.*, target arguments) accepted at the end of the dialogue, [11, 3].
- persuadee's valuation of a specific point of view (represented as a single target argument) [19].
- belief of the persuadee in the target arguments [12],
- plausibility of the target arguments [14].

However, while different computational argumentative techniques have been proposed and investigated in theoretical settings, *human persuasive techniques* have been thoroughly investigated in the *real world* by psychology researchers. These studies have identified the psychological grounds and characteristics of the different techniques that people actually use. The apparent gap between the notion of persuasion in argumentation theory and human persuasive techniques prevents automated persuasive agents from building upon proven psychological persuasive evidence and thus reduces the potential impact of such agents.

A handful of previous works have examined different facets of the connections between argumentation theory and human behaviour. For example, Rahwan et al. [16] have studied the reinstatement argumentative principle in questionnaire-based experiments, Cerutti et al. [5] examined humans' ability to comprehend formal arguments and Rosenfeld and Kraus [17, 18] have established that the argumentation theory falls short in explaining people's choice of arguments in synthetic and real world argumentative settings. To the best of our knowledge, in this recent line of research, no work has used *psychological evidences* to investigate the computational argumentation theory applicability and its possible adaptation.

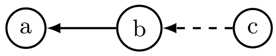

Figure 1: Example of bipolar argument graph where plain arrows mean attacks and dashed arrows mean supports.

In order to perform reasoning in a persuasive context, an argumentation framework needs to be defined (see [4] for a recent review). In its most basic form, an argumentation framework consists of a set of arguments A and an attack relation R over $A \times A$ [7]. In previous investigations of human argumentative behaviour (*e.g.*, [17, 18]), it was noticed that people often use supportive arguments rather than attacking ones, which necessitates the addition of the support relation as suggested in [1]. Furthermore, it is shown that people associate different belief levels in arguments, as suggested in [2], and different strength levels with interactions between arguments, as suggested in [8]. Interestingly, a framework named *Weighted Attacks/Support Argument* [14] embedding all these components has recently been proposed. We review this framework below.

2.1 Weighted Attacks/Support Argument

Weighted Attacks/Support Argument (WASA) graphs [14] are able to model argument graphs with attacks, supports, initial plausibility and strength of interactions between arguments taken into account. This framework merges several concepts: First of all, it is bipolar [1], allowing an additional support relation. Moreover, it uses initial weights as the plausibility for the arguments. In this work, we interpret the plausibility as an initial strength given to an argument.

Definition 1 *A WASA graph is characterized by a triplet* $\mathbb{A} = \langle \mathcal{A}, \mathcal{G}, w \rangle$*, where,*

- $\mathcal{A}$ *is a vector of size* n *ordering a set of arguments,*
- $\mathcal{G}$*, the* transposed adjacency matrix*, a square matrix of order* n*, with* $g_{ij} \in \{-1, 0, 1\}, \forall i, j \in \{1, \dots, n\}$*, where* $g_{ij} = -1$ *(resp. 1) represents an attack (resp. a support) from* j *to* i *and 0 means no relation,*
- w *is a weight vector in* $\mathbb{R}^n$*.*

Example 1 *Example of WASA graph.*

The bipolar argument graph depicted in Figure 1 can be represented as a WASA graph as follows:

$$\left\langle \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} w_0^a \\ w_0^b \\ w_0^c \end{pmatrix} \right\rangle$$

The acceptability of an argument a is calculated using the *Direct Aggregation Semantics* and is called the *acceptability degree* $Deg_{\mathbb{A}}(a)$ in a WASA $\mathbb{A} = \langle \mathcal{A}, \mathcal{G}, w \rangle$.

In order to calculate it, we first need to define a *damping factor* $d \geq 1$, acting as a decreasing effect the further the arguments are from a in the argument graph. Then, we can calculate the propagation matrix

$$Pr^{\mathcal{G},d} = \sum_{i=0}^{\infty} \left(\frac{1}{d} \mathcal{G} \right)^i .$$

Note that the sum is defined at the infinite. However, for most applications, the sum converges to a stable propagation matrix, *i.e.*, a matrix that does not change after further propagation steps as long as $d > m$ with m the maximum indegree in $\mathcal{G}$. When the propagation does not exactly converges, we can stop the process when the matrices before and after the additional propagation step are ϵ-close. In this case the propagation matrix can be approximated by $(I - \frac{1}{d} G)^{-1}$ [14].

Finally, we can calculate the acceptability degree vector

$$Deg_{\mathbb{A},d} = Pr^{\mathcal{G},d} \times w$$

for all the arguments. We denote $Deg_{\mathbb{A},d}(a)$ the acceptability degree of argument a.

In this work, we extend the traditional argument graph depiction as presented in Figure 1 to take into account temporal aspects of the dialogue. Namely, we add additional information to the graph: the *step* at which the argument has been/to be played. This also allows us to represent duplicate arguments that may be played several times in a given dialogue. Specifically, each argument is amended with a subscript, denoting at what step it is presented. A subscript of zero denotes that an argument is not presented at all.

3 Psychosocial Persuasion Principles

We focus on four well-established techniques commonly used by professional in, for instance, sales or marketing, which have been formalized by psychologists. These technique are aimed at persuading other people. Within the field of persuasive technologies, Fogg [9] defined persuasion as "an attempt to shape, reinforce, or change behaviours, feelings, or thoughts about an issue, object, or action".

Each of the four techniques is presented along with the psychological intuition standing behind it and one or two human studies from the literature that corroborated the benefit of the technique. Following psychological terminology, we define the *target request* to be a designated argument in the argumentation framework which represents the persuader's aim or goal – namely, having the persuadee doing or believing something. This target request is equivalent to the goal argument in Rosenfeld and Kraus's framework [19]). Positing only the target argument would probably not suffice to persuade the persuadee in many setting. Therefore, it is necessary to posit additional arguments which interact with the target request or interact with other arguments that may attack the target request. Each technique prescribes a procedure of how and when to posit these additional arguments.

3.1 Foot in the Door

3.1.1 The Premise

The *foot in the door (FITD)* technique has been first described by Freedman and Fraser [10]. This technique consists in asking a small favor before asking for the target behavior (*e.g.*, asking someone for direction before asking for money). Individuals who have been asked a small request before the target one generally tend to answer more favorably compared to individuals who have straightforwadly been asked the target request. This effect is due to the fact that accepting a small, initial request leads individuals to see themselves as being social – "agreeing to requests made by strangers". Consequently, when confronted with a second request, individuals tend to comply with the above perception and accept more willingly a bigger request. Failure to conform to the self-image generated by the first request generates a cognitive dissonance, which can explain compliance.

3.1.2 Studies

In their paper, Freedman and Fraser [10] report two studies. In the first one, the target request was to ask housewives to allow a survey team to come into their homes for two hours to conduct a study about the household products they use. Participants were assigned to one of four experimental groups, depending on the first contact (*i.e.*, the initial request) before asking the target request:

1. They were asked to answer some short questions about the kinds of soaps they use (*FITD*).
2. They were asked if they would be willing to answer different questions but the questions themselves were not asked (*agree only*).
3. They were *merely approached* but not asked anything.
4. There was no initial contact (*control group*).

Results show that the compliance rate is:

1. 52.8% when the *FITD* was used,
2. 33.3% when *agree only*,
3. 27.8% when *merely approached*,
4. 22.2% for the *control group*.

In their second study, the target request was to ask participants to put a very large sign which said "Drive Carefully" in their front yard. The authors designed several types of initial requests (*e.g.*, participants were initially asked to either put a small sign in their garden or sign a petition). A control condition was added, in which participants were not initially approached. In the control condition, only 16.7% of the participants complied with the target request. The highest compliance rate was obtained by asking something that was similar (*i.e.*, put a small sign) and on a similar issue (*i.e.*, safe driving), in which case 76% of the participants agreed to the target request. In the three other configurations, 47% complied with the target request, which remains higher than in the control condition.

Theses results show that making a small initial request before a larger one brings about an increased compliance rate with the target request. This effects holds whether both requests focus on the same behaviour or not and whether both requests target the same issue or not. However, the best compliance rate is achieved when both requests target the same type of behaviour, focused on the same issue.

3.2 Door in the Face

3.2.1 The Premise

The *door in the face, (DITF)* principle has been first theorized by Cialdini et al. [6]. This technique is almost symmetrical to the *FITD* technique discussed above. Using the *DITF* technique, one asks an "unreasonable" request before proposing a smaller one – the target request. The mechanism behind it is that, after the big request have been rejected, proposing a smaller request is perceived as a concession that the persuader has made from her original request. Thus, in order to maintain a certain level of reciprocity in the relation, the persuadee will tend to comply more with the target request than if it was made without the preparatory action.

3.2.2 Studies

In their original paper [6], the authors report three studies. In one of these studies, the target request was to have students accompany a group of juvenile delinquents on a two-hour trip to the zoo. They grouped the students into three conditions:

1. students with whom they engaged the interaction by asking them first to act as counselors to juvenile delinquents for a period of two years (*big request*),
2. students without any other request except for the two-hour trip (*control group*),
3. students where *both options* were presented and subjects were to choose which (if any) of the two options to take.

The results are as follows:

1. 50% agreed when the door in the face technique was used (first asking the *big request*).
2. 16.7% of the subjects of the *control group* complied with the request.
3. 25% agreed on the small request when *both options* were presented.

3.3 Repetition

3.3.1 The Premise

The *repetition* principle has been developed and tested by Petty and Cacioppo [15]. Simply put, the technique calls for the reformulation of arguments presented multiple times.

This repetition is not endless. Although it has been show that repeating an argument two or three times, under different formulations, may be beneficial, this effect tends to decrease as the number of repetition increases.

3.3.2 Studies

Petty and Cacioppo [15] conducted two studies in which participants heard the same argument (in different formulations) zero (for control), one, three and five times in succession. They were then asked to rate their agreement with the target argument and list the arguments they could recall.

The results show that participants' agreement increases for the first three conditions and decreases when the argument is presented five times.

3.4 Anchoring

3.4.1 Principle

The *anchoring* technique has been described by Tversky and Kahneman [21]. It refers to the tendency of people to generate judgements and estimations based on an initial reference point, an *anchor*. It is thus quite simple to manipulate this anchor by providing it in the argument itself, for instance. In such case, subsequent judgement made by the persuadee are

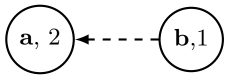

Figure 2: Foot in the door

expected to be biased toward the anchor initially provided. While this technique is more commonly used in numerical settings (*e.g.*, providing an anchor for a value of a product), it can also be applied to any type of arguments that could be ranked. In a sense, anchoring can be seen as a generalization of both the *foot in the door* and the *door in the face* principles. Indeed, in the former case, the persuader uses a smaller request first while in the latter she uses a bigger request first.

3.4.2 Studies

Tversky and Kahneman [21] provided several examples of the use of the anchoring technique. In one study, experimenters asked participants to give a series of estimations (in percentages), such as the percentage of African countries in the United Nations (UN). Before the estimation was provided, a random number between 0 and 100 was presented to participants by spinning a wheel in the participants presence. The results show that different initial number presented on the wheel led participants to generate different estimations: the group that received the number 10 estimated that 25% of African countries were in the UN (on average), whereas the group that received the number 65 estimated it at 45%.

In another study reported, experimenters asked two groups of students to estimate, within five seconds, the product of a numerical expression:

1. $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$, or,
2. $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$, *i.e.*, the exact same sequence but in reverse order.

The median estimate for the first group (with the descending sequence) was 2250, whereas it was 512 for the second group (with the ascending sequence). This result can be simply explained by the fact that subjects based their estimation on the results of the first operations they were able to make, which are obviously higher in the descending sequence.

4 From Psychosocial Persuasion Principles to Argumentation Frameworks

In this work, we will use the following interpretations of the plausibility of an argument and its acceptability degree. We consider that the plausibility w_a of an argument a corresponds to the acceptance ratio of the control conditions of the different psychological experiments. The degree of acceptability $Deg_{\mathbb{A},d}(a) \in [0, 1]$ of an argument a is used as threshold. When the value is bigger (resp. smaller) than 0.5, we expect the acceptance ratio to be bigger (resp. smaller) than 50%. This is consistent with the threshold in the epistemic approach to probabilistic argumentation (see, *e.g.*, [13]).

4.1 Foot in the Door

4.1.1 Argumentation Framework:

Before the application of the *FITD* technique, the persuader wants to have a single argument **a** accepted at the end of the debate (it is the target request). Figure 2 depicts the application of the *FITD* technique on this simple graph. Note that we show the modification in an isolation context. However, in general the arguments that we want to apply the techniques on are usually part of a bigger graph.

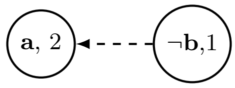

Figure 3: Door in the face.

Using the *FITD* principle, the persuader starts by playing a small argument **b** in order to have it accepted by the persuadee. She then plays argument **a** as in the original graph. However, this time, the argument **a** is supported by argument **b**. Therefore, the acceptance of **b** benefits the potential acceptance of **a**.

Example 2 *Figure 2 2 can be instantiated as follows:*

a *We come to your house to ask you questions.*
b *You answer some questions over the phone.*

The strategy of the persuader is to increase the chance of acceptance of **a** *by triggering the acceptance of* **b** *before.*

4.1.2 Analysis

The WASA associated to the single argument before the application of the *FITD* principle is trivial. Therefore we show below the procedure directly on the modified graph of Figure 2. The WASA is defined as follows:

$$\mathbb{A}' = \left\langle \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} w_a \\ w_b \end{pmatrix} \right\rangle$$

The propagation matrix after convergence, with $d = 2$ is:

$$\begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{pmatrix}$$

When calculating the degree of acceptability for both **a** and **b**, we have $Deg_{\mathbb{A},d}(a) = w_a - \frac{w_b}{2}$ and $Deg_{\mathbb{A},d}(b) = w_b$.

This means that argument **a** is accepted iff $w_a - \frac{w_b}{2} > 0.5$. Referring to the study presented in Section 3.1.2, we see that the acceptance ratio in the control condition is 22.2%. We consider $w_a = 0.222$ and that argument **a** is thus accepted iff $w_b > 0.556$. According to [10], about two third of the participants agreed with the smaller request. Therefore argument **a** should be accepted and indeed, 52.8% of the participants agreed with the target request when the *FITD* was used.

4.2 Door in the Face

4.2.1 Argumentation framework

In the same idea than previously, the initial graph is a single target argument. However, for the transformation, we consider in this case that the support of the target request is the negation of the second argument. Note that this is an abuse of notation where we really mean that it is the non-acceptance of argument **b** that reinforces argument **a**.

Example 3 *Let us instantiate the arguments of Figure 3, in a context of the zoo trip experiment, as follows:*

a *Look after juvenile delinquents for a two-hour trip.*
b *Look after them for two years.*

4.2.2 Analysis

The WASA associated with Figure 3 is:

$$\mathbb{A}' = \left\langle \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} w_a \\ 1 - w_b \end{pmatrix} \right\rangle$$

The WASA is the same as previously, except for the initial plausibility part for argument **b**.

In this case, argument **a** is accepted iff $w_a + \frac{1-w_b}{2} > 0.5$.

Following the study presented in Section 3.2.2, $w_a = 0.167$. Therefore, we need to have $1 - w_b > 0.666$ in order to have the target request accepted. In [6], the authors state that no participants accepted the bigger request prior to be presented with the second, smaller one. Following the same intuition, we consider $w_b = 0$. We can then conclude that the target request should be accepted. The psychological study suggests the same conclusion.

4.3 Repetition

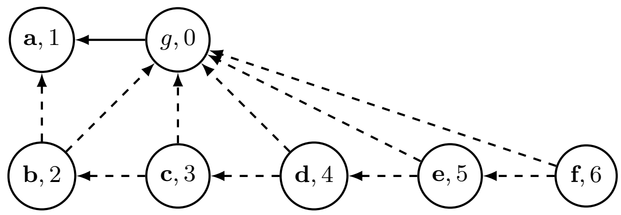

Figure 4: Repetition

Figure 4 depicts how the repetition technique can be represented. Argument a is the target request while arguments b to f are the **same** argument (in different formulations) repeated a certain number of times. As stated before, this differ from traditional abstract argumentation where arguments are represented in the graph irrespectively of the way they are used. In this case, by representing the timestep at which the argument has been played allows us to represent that an argument has been played several times in the dialogue. Argument g is a fictitious argument reinforced each time a repetition is made.

4.3.1 Analysis

The WASA associated is as follows for 5 repetitions. The WASA for the other steps can be easily deduced from this one.

$$\mathbb{A} = \left\langle \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} w_a \\ w_b \\ w_c \\ w_d \\ w_e \\ w_f \\ w_g \end{pmatrix} \right\rangle$$

If we take a damping factor of exactly one plus the maximum indegree in the graph induced by each step (*i.e.*, 2 for 0 repetition, 3 for 1 repetition, 4 for 4 repetitions and 6 for 5 repetitions), we obtain the following equations for the acceptance of argument **a** in each condition.

0 repetition $w_a - \frac{wg}{2}$
1 repetition $w_a + \frac{2*w_b}{9} - \frac{w_g}{3}$
3 repetitions $w_a + \frac{3*w_b}{16} - \frac{w_c}{64} - \frac{17*w_d}{256} - \frac{w_g}{4}$
5 repetitions $w_a + \frac{5*w_b}{36} - \frac{w_c}{216} - \frac{37*w_d}{1296} - \frac{253*w_e}{7776} - \frac{1549*w_f}{46656} - \frac{w_g}{6}$

In this case, multiple solutions are possible. According to [15] $w_a = 0.372$. We decide to associate the following values with the different arguments: $w_b = 0.9, w_c = 0.5, w_d = 0.2, w_e = 0.05, w_f = 0.01$ and $w_g = 0$. We assume that the argument that is being repeated is an argument with a high plausibility value at the beginning but that its strength decreases as it keeps being used. As argument **g** is a fictitious one, we give it an initial value of 0.

With these values, the acceptability degree for each condition is: 0.372 for 0 repetition, 0.572 for 1 repetition, 0.520 for 3 repetitions and 0.487 for 5 repetitions. Therefore, argument **a** is accepted in the conditions 1 and 3 and rejected for 0 and 5. These results agree with the original study.

4.4 Anchoring

As a generalization of the foot in the door and the door in the face principle, the argument graph for the anchoring principle is also a generalization. However, the initial plausibilities in the WASA are no tied to the actual arguments this time but rather to their position in the ranking and the objective in the persuasion problem.

For instance, if the objective is to sell a car at the highest price possible, the first argument should be a price above the actual price and then, in a second time, the actual price. On the other hand, if the objective is to buy the very same car, it is better to give a very low price first and then converge towards the price we were willing to pay from the beginning. Therefore, in the former case, the initial value should be low for the extremely high price, increasing the more it closes the gap with the price, and then decreasing again as is goes further down, past the price and vice-versa in the latter case.

5 Discussion

In this paper we have made a modest step towards bridging the gap between abstract argumentation and psychological evidence for persuasion. This can be viewed as part of a larger effort to investigate what drives human decision-making in the argumentative context [20].

We have explained how to design argument graphs modelling four different psychological techniques, commonly used by people, and we have shown how they can be used to theoretically explain the observed results in human studies.

In future works we plan to include additional psychological techniques and a deeper analysis of the WASA and other abstract argumentation frameworks. A comparison with the traditional semantics in bipolar argumentation frameworks is another interesting direction. Finally, performing user studies based on the new representation is crucial to validate this new hybrid formalization.

Acknowledgements

This research is part funded by EPSRC Project EP/N008294/1 (Framework for Computational Persuasion).

REFERENCES

[1] Leila Amgoud, Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Pierre Livet, 'On bipolarity in argumentation frameworks', *International Journal of Intelligent Systems*, **23**(10), 1062–1093, (2008).

[2] Pietro Baroni, Marco Romano, Francesca Toni, Marco Aurisicchio, and Giorgio Bertanza, 'Automatic evaluation of design alternatives with quantitative argumentation', *Argument & Computation*, **6**(1), 24–49, (2015).

[3] Elizabeth Black, Amanda J Coles, and Christopher Hampson, 'Planning for persuasion', in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 933–942. International Foundation for Autonomous Agents and Multiagent Systems, (2017).

[4] Gerhard Brewka, Sylwia Polberg, and Stefan Woltran, 'Generalizations of dung frameworks and their role in formal argumentation', *IEEE Intelligent Systems*, **29**(1), 30–38, (2014).

[5] Federico Cerutti, Nava Tintarev, and Nir Oren, 'Formal arguments, preferences, and natural language interfaces to humans: an empirical evaluation', in *Proceedings of the Twenty-first European Conference on Artificial Intelligence*, pp. 207–212. IOS Press, (2014).

[6] Robert B Cialdini, Joyce E Vincent, Stephen K Lewis, Jose Catalan, Diane Wheeler, and Betty Lee Darby, 'Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique.', *Journal of personality and Social Psychology*, **31**(2), 206, (1975).

[7] 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), 321–357, (1995).

[8] Paul E Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge, 'Weighted argument systems: Basic definitions, algorithms, and complexity results', *Artificial Intelligence*, **175**(2), 457–486, (2011).

[9] Brian J Fogg, 'Persuasive computers: perspectives and research directions', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 225–232. ACM Press/Addison-Wesley Publishing Co., (1998).

[10] Jonathan L Freedman and Scott C Fraser, 'Compliance without pressure: the foot-in-the-door technique.', *Journal of personality and social psychology*, **4**(2), 195, (1966).

[11] Emmanuel Hadoux, Aurélie Beynier, Nicolas Maudet, Paul Weng, and Anthony Hunter, 'Optimization of probabilistic argumentation with markov decision models', in *International Joint Conference on Artificial Intelligence (IJCAI)*, (2015).

[12] Emmanuel Hadoux and Anthony Hunter, 'Strategic sequences of arguments for persuasion using decision trees', in *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, (2017).

[13] A. Hunter and M. Thimm, 'Probabilistic reasoning with abstract argumentation frameworks', *Journal of Artificial Intelligence Research*, (2017). (in press).

[14] Till Mossakowski and Fabian Neuhaus, 'Bipolar weighted argumentation graphs', *arXiv preprint arXiv:1611.08572*, (2016).

[15] Richard E Petty and John T Cacioppo, 'Effects of message repetition and position on cognitive response, recall, and persuasion', *Journal of personality and Social Psychology*, **37**(1), 97–109, (1979).

[16] Iyad Rahwan, Mohammed I Madakkatel, Jean-François Bonnefon, Ruqiyabi N Awan, and Sherief Abdallah, 'Behavioral experiments for assessing the abstract argumentation semantics of reinstatement', *Cognitive Science*, **34**(8), 1483–1502, (2010).

[17] Ariel Rosenfeld and Sarit Kraus, 'Argumentation theory in the field: An empirical study of fundamental notions', in *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing, Forlì-Cesena, Italy, July 21-25, 2014.*, (2014).

[18] Ariel Rosenfeld and Sarit Kraus, 'Providing arguments in discussions on the basis of the prediction of human argumentative behavior', *ACM Transactions on Interactive Intelligent Systems*, **6**(4), 30:1–30:33, (December 2016).

[19] Ariel Rosenfeld and Sarit Kraus, 'Strategical argumentative agent for human persuasion', in *22nd European Conference on Artificial Intelligence (ECAI)*, volume 285, p. 320. IOS Press, (2016).

[20] Ariel Rosenfeld and Sarit Kraus, *Predicting Human Decision Making: From Prediction to Intelligent Agent Design*, Morgan & Claypool, 2018.

[21] Amos Tversky and Daniel Kahneman, 'Judgment under uncertainty: Heuristics and biases', in *Utility, probability, and human decision making*, 141–162, Springer, (1975).