The "Lottery Paradox" Paradox and other Self-Attacking Arguments from the Point of View of Defeasible Argumentation Frameworks

Gustavo A. Bodanza
Logic and Philosophy of Science Research Center (CILF)
Artificial Intelligence Research and Development Laboratory (LIDIA)
Universidad Nacional del Sur, Bahía Blanca
CONICET
Argentina
bodanza@gmail.com

Abstract

The "lottery paradox" paradox –a problem derived from the well-known lottery paradox– arises as the result of chaining defeasible reasons that lead from the statement 'the lottery is fair' to the statement 'the lottery is not fair', conforming in consequence a self-attacking argument. This paper shows two simple semantics that capture the intuitions about what kinds of arguments should become justified in such scenarios. The semantics, called *sustainable* and *lax* respectively, are modeled upon argumentation frameworks and constructed on the common ground of the powerful concept of admissibility introduced by P.M. Dung. Sustainable semantics results to be more "credulous" than preferred semantics, since enables the acceptance of arguments that are attacked only by other arguments that in time are self-attacking. Lax semantics, on the other hand, do the same as sustainable semantics, in addition to the acceptance of arguments that are self-attacking only in an indirect way (viz., those involved in odd-length cycles of attack).
Keywords: Lottery paradox, self-attacking arguments, defeasible argumentation, extension semantics.

1 Introduction

John Pollock (Pollock (1990, 1991, 1994)) studied a particular instance of self-attacking arguments derived from the lottery paradox, a problem that has been introduced by Kyburg (1961) as a problem for reasoning with probabilities. Imagine a fair lottery with 1,000,000 tickets, so that each ticket has one in a million chances of winning. Given a particular ticket, we can tentatively conclude that it will not win, since its probability is very low. But the same can be concluded about every ticket, hence we can conclude that none of them will win. But given that the lottery is fair, one ticket must win. This is called "the lottery paradox". Upon this contradiction Pollock claimed that a further contradiction can be drawn, namely that since none of the tickets will win, the lottery cannot be fair. That is, the argument of the lottery paradox, based upon the premise that the lottery is fair, attacks itself since one of its consequences contradicts its premise. This is what Pollock called the ' "lottery paradox" paradox'.

This work is not about the lottery paradox, but about the "lottery paradox" paradox as a self-attacking argument. The study of the ways an argument can be deemed warranted with regard to its interaction with other arguments is known as *defeasible argumentation*. Roughly, an argument is warranted if it can be successfully defended against the attacks by other arguments. For example, if an argument α attacks an argument β, and β in turn attacks an argument γ, we can say that γ is defended by α; moreover, if α has no attackers, then we can conclude that both α and γ remain warranted (since they can be successfully defended) while β is overruled. Self-attacking arguments are special cases where the attack relation form a cycle of length one.

In the "lottery paradox" paradox, the argument concluding that the lottery is fair (β) is a subargument of the larger argument concluding that the lottery is not fair (α). Hence, argument α attacks its subargument β and so it attacks itself. Since α cannot be successfully defended, it has to be excluded from any extension. This lead us to expect that β be warranted, since it does not receive any sensible attack. After all, the paradox is clearly provoked by a chaining of reasons that goes from β's conclusion ('the lottery is fair') to the whole argument αs conclusion ('the lottery is not fair'). Then, it seems that the attack of α over β should be neglected. But surprisingly, there is no extension semantics as far as we know- that yields this behavior, remaining sound in all other cases.

In this paper we argue that the problem of self-attacking arguments can be approached from an extension semantics view, upon the guiding idea of seeing an argumentation process as a strategic interaction between two agents (though we do not attach ontological weight to this assumption, since the agents may be the two sides that a single reasoner may take while pondering which arguments should be warranted). An agent will select a maximum set of argument that can be defended against a set of argument that any "rational" opponent can select. In the "lottery paradox" paradox, no "rational" agent will choose the sets $\{\alpha\}$ or $\{\alpha, \beta\}$ (because they contain internal attacks), hence the set $\{\beta\}$ is the only maximum set of arguments that can be defended against any "rational" choice

of a hypothetical opponent. Accordingly, an extension semantics should judge $\{\beta\}$ as an extension of that argumentation framework.

The solution takes up from a wide consensus among researchers in the argument systems' community, who seek to solve the problem of self-attacking arguments by rejecting any justification for arguments in any odd-length cycles of attack. We propose *sustainable semantics* (previously introduced in [2]) as the way of capturing this lead. The platform on which sustainable semantics is developed is the argumentation frameworks model of Dung. The main idea introduced there is the admissibility of sets of arguments. This notion defines the way a set of arguments can be defended from external attacks. We build upon this idea, stating that the defense of a set of arguments should not be defined in absolute terms but relative to other possible challenging sets of arguments. In this way we obtain a notion of *cogency* on which we base our sustainable semantics.

On the other hand, self-attacking arguments can arise by means of indirect attacks. The most simple case is given by a 3-length cycle of attacks: α attacks β attacks γ attacks α. Since each one of these arguments attacks its defender, each one attacks itself in an indirect way. This is also a common hypothesis in the argument systems community. Nevertheless, one can imagine some scenarios for that framework in which the self-attacking character of the involved arguments is not clear. We will introduce here *lax extensions*, a semantics that enables the acceptance of such arguments, while avoids the undesirable interference of (direct) self-attacking arguments in a similar way as sustainable semantics.

The paper is organized as follows. The main features of Dung's system are presented in section 2. In section 3 we present the guiding intuitions of our proposal. Section 4 shows the formal concept of *cogency*, the common ground on which both sustainable and lax semantics are developed. Sections 5 and 6 discuss the treatment of self-attacking arguments by sustainable and lax semantics, respectively. Section 7 shows the relationships between the semantics. Finally, section 8 summarizes our conclusions.

2 Dung's argumentation frameworks

The formal platform on which this work is based is Dung's argumentation frameworks (Dung (1995)). An argumentation framework is a pair $AF = \langle AR, \mathit{attacks}\rangle$, where AR is a set of abstract entities called 'arguments' and $\mathit{attacks}$ is a binary relation $\mathit{attacks} \subseteq AR \times AR$ intended to denote attacks among arguments. The fundamental question about argument frameworks is which arguments remain "justified" under the attacks among them. A subset of "justified" arguments is called an *extension*, and alternative criteria yield different kinds of extensions. Two basic concepts allow to obtain all the extensions defined by Dung: 'acceptability' of arguments, and 'admissibility' of a set of arguments. The formal definitions of these notions are:

DEFINITION **2.1** *(Dung ([3])). In any argumentation framework AF an argument σ is said* acceptable *w.r.t. a subset S of arguments of AR, in case that*

for every argument τ such that τ attacks σ, there exists some argument $\rho \in S$ such that ρ attacks τ[1]*. A set of arguments S is said* admissible *if each $\sigma \in S$ is acceptable w.r.t. S, and is conflict-free, i.e., the attack relation does not hold for any pair of arguments belonging to S. A* preferred extension *is any maximally admissible set of arguments of AF. A* complete extension *of AF is any conflict-free subset of arguments which is a fixed point of $F(\cdot)$, where $F(S) = \{\sigma : \sigma$ is acceptable w.r.t. $S\}$, while the* grounded extension *is the least (w.r.t. $\subseteq$) complete extension.*

Extensions are said to constitute the "semantics" of an argumentation framework, because they capture the notion (weaker than truthfulness) of *justifiability* of arguments. Different kinds of extensions give way to alternative semantics, that is, different intuitions about which arguments can be justified. The usual interpretation is that semantics justifying a single extension correspond to some kind of skeptical arguer. On the other hand, in a semantics that yields multiple extensions, to select their intersection also characterizes a kind of skeptical attitude, while the arbitrary choice of one of the extensions represents a credulous decision. For example, in Dung's system, preferred, complete and stable extensions usually yield several extensions, so they can be used to model alternative forms of credulity, while different conceptions of skepticism can be modeled by the grounded extension or the intersections (each for each notion of credulity) of the credulous extensions.

EXAMPLE **2.1** Assume $AF = \langle\{\alpha, \beta\}, \{(\alpha, \alpha), (\alpha, \beta)\}\rangle$. Then $\emptyset$ is the grounded extension and the only preferred and complete extension. There is no stable extension in this framework.

3 A strategical approach to argumentation

Dung's view of argument justification does not assign any role whatsoever to the utterer/s of the arguments. His framework can be interpreted as the decision-making setting of a single agent, who ponders the arguments in the framework. Another interpretation is to see the arguments as laid out by an agent to defend herself against the arguments presented by another agent. This distinction motivated some researchers to view defeasible argumentation from a dialogical point of view, modeling dialogues between two players (pro and con) who develop an argumentation strategy by advancing arguments one by one in alternation (e.g. Vreeswijk and Prakken (2000)).

While we take this view of argumentation as arising from the interaction of two agents, the novelty of our approach consists in that we focus on argumentation strategies instead of considering the properties of a sequential dialogue. An argumentation strategy can be viewed as the set of arguments that an intentional agent may lay out in response to the arguments of the other agent in a dialogue.

[1] In order to simplify the notation we will say that σ *attacks* A (conversely, A *attacks* σ) iff there exists some $\tau \in A$ such that σ *attacks* τ (τ *attacks* σ, respectively).

As this notion involves all possible combinations of arguments, we consider that any subset of AR in an argumentation framework $AF = \langle AR, attacks \rangle$ can be seen as an argumentation strategy (this includes, of course, strategies that are not cleaver at all). In this context, our main working hypothesis is that the rationality of choosing an argumentation strategy is relative to the argumentation strategies that the rival may select. What matters here is not whether an argumentation strategy (as a subset of arguments) is "admissible" in the whole framework, but if it is so relative to the sub-frameworks formed by the strategy itself together with the possible strategies of the other agent.

Under this view we can rethink the main problems of argumentation. While usual questions about argumentation ask how arguments can be compared and justified, we want to know now how argumentation strategies can be compared and justified. Once that question is settled, we extend the justification to individual arguments. The main methodological steps in the characterization of justification are based on the following assumptions:

- argumentation strategies can be compared on the basis of the attack relation among the arguments that they include;
- argumentation strategies can be justified on the basis of their comparison; and
- the justification of an argument obtains from the justification of the argumentation strategies to which it belongs.

We will see that, in this way, the undesirable interferences of self-attacking arguments will disappear without any necessity of considerations ad hoc.

4 The notion of cogency

The first idea is to establish a comparison among argumentation strategies (just 'strategies' henceforth) using the notion of admissibility. We will consider a strategy A at least as cogent as a strategy B if and only if B is conflict-free and all its arguments can be defended against the arguments of B. This amounts to saying that A is an admissible set within the framework formed just by the arguments in $A \cup B$. In order to introduce this definition, we first need an auxiliary notion: a restriction of an argumentation framework AF by a subset of arguments $S \subseteq AR$, is the framework $AF|_S = \langle S, attacks|_S \rangle$, where $attacks|_S$ is the restriction of $attacks$ to S.

Remark. From now on we will refer to an arbitrary but fixed argumentation framework AF, so all definitions and results will be relative to that framework.

DEFINITION **4.1** For any pair of strategies A, B, we say that A is *at least as cogent as* B, in symbols '$A \succeq_{Cog} B$' iff A is admissible in $AF|_{A \cup B}$.

It immediately follows that:

PROPOSITION **4.1** A strategy A is admissible iff for every strategy B, $A \succeq_{Cog} B$.

Note that according to this result, it is easy to recover the preferred semantics by requiring a strategy A to be a maximal strategy (for set-theoretic inclusion) such that for every $B \subseteq AR$, $A \succeq_{Cog} B$. Admissibility is sufficient for cogency, but cogency is weaker than admissibility: while we understand that A is admissible if it can be defended against any other strategy, A is just cogent if it can be defended against any other strategy which can be defended against A. Formally:

DEFINITION **4.2** A strategy A is *relatively cogent* iff for every strategy B, if $B \succeq_{Cog} A$ then $A \succeq_{Cog} B$.

This simple concept of relatively cogent strategies will enable us to define two different kinds of extension semantics for dealing with the problem of self-attacking arguments. The first one, which we will call 'sustainable', will behave exactly the same as preferred semantics unless self-attacking arguments are present, while the second one, which we will call 'lax', will be more credulous than both preferred and sustainable semantics since it will invoke a weaker criterion of defense. The following example will help for understanding them.

EXAMPLE **4.1** *(Continuation of example 2.1)* In the argumentation framework of example 2.1 (which represents the "lottery paradox" paradox), the set of all possible strategies is $2^{\{\alpha,\beta\}} = \{\emptyset, \{\alpha\}, \{\beta\}, \{\alpha,\beta\}\}$. Then, $\{\beta\}$ and $\emptyset$ are the only relatively cogent strategies. It is clear that $\{\alpha\}$ is not relatively cogent since $\emptyset \succeq_{Cog} \{\alpha\}$ but $\{\alpha\} \not\succeq_{Cog} \emptyset$. Note also that $\{\alpha\} \not\succeq_{Cog} \{\beta\}$.

5 Sustainable semantics

The first semantics just arises by asking for a strategy to be a maximal relatively cogent strategy. We call it 'sustainable' suggesting a strategy that holds against any articulated criticism.

DEFINITION **5.1** A strategy A is a *sustainable extension* of AF iff A is a maximal (w.r.t. $\subseteq$) relatively cogent strategy in AF.

Sustainable extensions avoid the undesirable interference of self-attacking arguments, as shown in the following example:

EXAMPLE **5.1** *(Continuation of example 2.1)* In the argumentation framework of example 2.1 $\{\beta\}$ is the only sustainable extension since, besides itself, $\emptyset$ is the only strategy as cogent as $\{\beta\}$, but $\{\beta\}$ is as cogent as $\emptyset$ and is larger.

It is easy to see that sustainable extensions yield a more credulous semantics than preferred extensions. The framework of the above example is a case in which the only preferred extension, $\emptyset$, is strictly included in the only sustainable extension, so the latter is clearly more credulous than the preferred extension. We can generalize this fact:

THEOREM **5.1** If A is a preferred extension then there exists some sustainable extension B such that $A \subseteq B$.

Given a preferred extension, the sustainable extension to which it belongs may be the preferred extension itself.

EXAMPLE **5.2** *("The Nixon diamond")* In the famous benchmark problem known as "the Nixon diamond" there are two arguments: α: 'Nixon is pacifist, since he is quaker and quakers tend to be pacifist'; and β: 'Nixon is not pacifist, since he is republican and republicans tend not to be pacifist'. The argumentation framework representing the problem is $AF = \langle\{\alpha, \beta\}, \{(\alpha, \beta), (\beta, \alpha)\}\rangle$, where $\{\alpha\}$ and $\{\beta\}$ are both the preferred and sustainable extensions (the grounded extension is $\emptyset$).

EXAMPLE **5.3** Assume $AF = \langle\{\alpha, \beta, \gamma\}, \{(\alpha, \beta), (\beta, \gamma), (\gamma, \alpha)\}\rangle$. Then $\emptyset$ is the only extension at all, which is grounded, preferred and sustainable at the same time.

6 Lax semantics

The second semantics enables the acceptance of more arguments than sustainable semantics. It justifies any strategy A such that each one of the strategies that "defeat" A are in time "defeated" by some other strategy. This idea, which does not work when applied to isolated arguments, do work well when applied to sets of arguments (provided some concessions, as we shall see). To see that it does not work for isolated arguments, let us consider an argumentation framework where we just have $attacks = \{(\alpha, \beta), (\beta, \gamma), (\gamma, \delta)\}$. If we expect to accept any argument which defeaters are all defeated by some arguments, then we would have to accept δ, since its only defeater, γ, is in time defeated by β; but this is counterintuitive since β is defeated by α, which is undefeated, and so γ, the defeater of δ, is ultimately defended, hence δ should remain unjustified. Let us consider strategies now. First, the notion of "defeat" among strategies can be understood in terms of cogency: a strategy A "defeats" a strategy B if A is at least as cogent as B but B is not at least as cogent as A (that is, A can be defended against B but B cannot be defended against A). So, if we expect to accept any strategy which defeaters are all defeated by some strategy in the above setting, we will see that the strategy $\{\alpha, \gamma\}$ is to be accepted, since there is no strategy that defeats it. On the other hand, the strategy $\{\delta\}$ is not to be accepted, since is defeated by $\{\alpha, \gamma\}$, which is not defeated by any strategy. This outcome is in accordance with all the known extension semantics for argumentation frameworks.

The following definition gives us a formalization of the "defeat" relation among strategies:

DEFINITION **6.1** A strategy A is *strictly more cogent* than a strategy B, in symbols $A \succ_{Cog} B$, iff $A \succeq_{Cog} B$ and $B \not\succeq_{Cog} A$.

Now, we take those strategies which "defeaters" are all "defeated" in time by some other strategies.

DEFINITION **6.2** A strategy A is *weakly cogent* iff for every strategy B, if $B \succ_{Cog} A$ then there exists some strategy C such that $C \succ_{Cog} B$.

Finally, we choose the maximal weakly cogent strategies.

DEFINITION **6.3** A is a *lax extension* iff A is maximally (w.r.t. $\subseteq$) weakly cogent.

EXAMPLE **6.1** *(Continuation of example 2.1)* In the argumentation framework corresponding to the "lottery paradox" paradox, the only weakly cogent strategies are $\{\beta\}$ and $\emptyset$ (no strategy is strictly more cogent than them). Then, $\{\beta\}$ is the only lax extension.

In this example, it is clear that no strategy A is such that $A \succ_{Cog} \{\beta\}$, hence the condition for $\{\beta\}$ to be weakly cogent is vacuously accomplished. In more intuitive terms, we can say that no strategy –$\{\alpha\}$ included– is a "defeater" of $\{\beta\}$.

The very difference between lax and sustainable extensions arises in presence of attack cycles of length odd and greater than 1.

EXAMPLE **6.2** Assume $AF = \langle\{\alpha, \beta, \gamma\}, \{(\alpha, \beta), (\beta, \gamma), (\gamma, \alpha)\}\rangle$. Then $\emptyset$ is the only sustainable extension while $\{\alpha\}$, $\{\beta\}$ and $\{\gamma\}$ are all lax strategies.

In order to recognize the rationale behind lax extensions in cases of this sort, let us consider an interpretation of the above example.

EXAMPLE **6.3** *(Interpretation of example 6.2)* Assume $AR = \{\alpha, \beta, \gamma\}$ where the arguments are:
α : "Symptoms x and y suggest the presence of disease d_1, so I should apply therapy t_1";
β : "Symptoms x and z suggest the presence of disease d_2, so I should apply therapy t_2";
γ : "Symptoms x and w suggest the presence of disease d_3, so I should apply therapy t_3".

Assume these are the main arguments considered by an agent $\mathcal{H}$ (say, a M.D.) having to make a decision on which therapy should be applied to some patient who has symptoms x, y, z and w. Then $\mathcal{H}$ establishes the attack relation as follows:

- α attacks β, since $\mathcal{H}$ thinks that therapy t_1 is less invasive than t_2;
- β attacks γ, since $\mathcal{H}$ thinks that symptom z is more clearly detected than symptom w;

- γ attacks α, since $\mathcal{H}$ thinks that the consequences of disease d_3 would be more serious than those of disease d_1.

This example shows that, whatever the decision of $\mathcal{H}$, it is clearly more prudent to select any of the arguments than to refuse all of them. Hence, the semantics given by lax extensions is more in accordance with this rationale than that of sustainable (and preferred) extensions.

Bodanza and Tohmé ([2]) introduced *tolerant* extensions, a semantics that is akin to lax semantics with respect to the treatment of odd-length cycles of attack.

DEFINITION **6.4** *(Bodanza and Tohmé ([2]))* A strategy $A \subseteq AR$ is *cyclically cogent* iff for every strategy C, if $C \succ_{Cog} A$ then there exists a sequence $D_1, \ldots, D_m$ of strategies such that $D_1 = C$, $D_m = A$ and $D_{k+1} \succ_{Cog} D_k$, for $1 \leq k < m$. $A \subseteq AR$ is a *tolerant extension* iff A is a maximal (w.r.t. $\subseteq$) cyclically cogent strategy.

The behavior of both tolerant and lax semantics is at odds with Dung's conception of indirect attack. An argument α *indirectly attacks* an argument β iff there exists a sequence of arguments $\alpha_0, \ldots, \alpha_{2n+1}$ such that $\alpha = \alpha_0$, $\beta = \alpha_{2n+1}$ and α_{i+1} attacks α_i, for $0 \leq i \leq 2n$. Regarding the arguments belonging to an odd-length cycle of attacks, all of them are indirectly self-attacking according to this definition. Nevertheless, our example 6.3 seems to challenge a general application of this view, inasmuch it is at least not clear at all that every argument of the framework posses an attack on itself, even though it were indirect.

On the other hand, the example shows the convenience of picking any one of the arguments in the cycle in such practical decision making scenarios. From this point of view the problem is not one of epistemic justification, since the argumentation tends not to justify what to believe but what to do. So both lax and tolerant semantics seem to fit well to this view.

As said above, the behavior of tolerant and lax semantics is akin, but it is not exactly the same. While the fact that tolerant extensions are always lax extensions is obvious, consider the following counterexample for the converse:

EXAMPLE **6.4** Assume $AF = \langle \{\alpha, \beta, \gamma\}, \{(\alpha, \beta), (\beta, \alpha), (\beta, \gamma), (\gamma, \alpha)\} \rangle$. Then $\{\alpha\}$ and $\{\beta\}$ are lax extensions but only $\{\beta\}$ is tolerant.

While tolerant semantics behaves as credulously as preferred semantics in this example, the behavior of lax semantics is clearly more credulous. The acceptance of $\{\alpha\}$ by lax semantics is due to the simple fact that its only "defeater", $\{\gamma\}$, is in time "defeated" by $\{\beta\}$ (this is the kind of concessions mentioned at the beginning of this section). This is, indeed, the reason why we call such extensions "lax": they are just strategies which potential "defeaters" are not outright warranted.

In sum, sustainable semantics on one hand and lax and tolerant semantics on the other offer two alternative sorts of tools for dealing with the problem of self-attacking arguments, the first one in accordance with Dung's point of view of indirect attacks, and the later in disagreement with that view but arguably more in accordance with practical decision making.

7 Relationships among different extension semantics

Sustainable and lax strategies are related among them and other semantics as follows:

THEOREM **7.1**

1. Every preferred extension is contained in some sustainable extension.
2. Every sustainable extension is contained in some tolerant extension.
3. Every tolerant extension is a lax extension, but not vice versa.

Proof.

1. Immediate from 4.1.
2. It is clear that every sustainable extension is cyclically cogent, hence it must be contained in some maximal cyclically cogent strategy, i.e., in a tolerant extension.
3. Obvious. Example 6.4 shows that lax extensions are not always tolerant.

□

PROPOSITION **7.1** In every argumentation framework which is free of self-attacking arguments the class of the preferred extensions and the class of the sustainable extensions coincide (but the class of lax extensions does not necessarily coincide with them).

Proof. Assume that A is a sustainable extension which is not preferred, and let us show that there exists at least one self-attacking argument. Note that A is not admissible, so there exists some argument α that attacks A but it is not attacked by A. Hence we have to cases: (i) $\{\alpha\} \succ_{Cog} A$, or (ii) α is self-attacking and so $\{\alpha\} \not\succeq_{Cog} A$. But (i) contradicts the hypothesis that A is sustainable, hence (ii) is the case. □

8 Conclusion

In this work we have analyzed the "lottery paradox" paradox from the point of view of extension semantics for argumentation frameworks, and showed two approaches to avoid the interference of a self-attacking argument in the acceptance of some of the subarguments that compound it. Sustainable semantics enables that outcome while its behavior is tantamount to that of preferred semantics in argumentation frameworks free of self-attacking arguments. On the other hand, we have shown some intuitions at odds with the conception of indirect attack, arguing that arguments involved in odd-length cycles of attacks are not necessarily self-attacking, and so they can be accepted (especially when argumentation is about practical decisions). We have offered here lax extensions as a semantics capturing that intuition. The behavior is akin, but not equivalent, to other semantics in the literature like tolerant semantics ([2]) and CF2 semantics ([1]).

Acknowledgments

This work was supported by ANPCyT and Universidad Nacional del Sur, Argentina, under projects PICTO 731, PICT 693 and PGI 24/ZI29.

References

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

[2] Bodanza, G. and F. Tohmé: Two approaches to the problems of self-attacking arguments and general odd-Length cycles of attack. *Journal of Applied Logic*, to appear.

[3] Dung, P. M.: On the Acceptability of Arguments and its Fundamental Role in Non-Monotonic Reasoning, Logic Programming, and n-Person Games. *Artificial Intelligence*, 77: 321–357, 1995.

[4] Pollock, J.: Self-defeating arguments. *Minds and Machines*, 1: 367–392, 1991.

[5] Pollock, J.: Justification and defeat. *Artificial Intelligence* 67: 377–407, 1994.

[6] Pollock, J.: Defeasible reasoning with variable degrees of justification. *Artificial Intelligence* 133: 233–282, 2001.

[7] Vreeswijk, G., and Prakken, H.: Credulous and Sceptical Argument Games for Preferred Semantics. In *Proc. JELIA 2000*, 239–253, 2000. LNAI 1919.