SCF2 – an Argumentation Semantics for Rational Human Judgments on Argument Acceptability

Marcos Cramer[1] and Leendert van der Torre[2]

[1] International Center for Computational Logic, TU Dresden, Germany
marcos.cramer@tu-dresden.de
[2] Computer Science and Communications, University of Luxembourg, Esch-sur-Alzette, Luxembourg
leon.vandertorre@uni.lu

Abstract. In abstract argumentation theory, many argumentation semantics have been proposed for evaluating argumentation frameworks. This paper is based on the following research question: Which semantics corresponds well to what humans consider a rational judgment on the acceptability of arguments? There are two systematic ways to approach this research question: A normative perspective is provided by the principle-based approach, in which semantics are evaluated based on their satisfaction of various normatively desirable principles. A descriptive perspective is provided by the empirical approach, in which cognitive studies are conducted to determine which semantics best predicts human judgments about arguments. In this paper, we combine both approaches to motivate a new argumentation semantics called SCF2. For this purpose, we introduce and motivate two new principles and show that no semantics from the literature satisfies both of them. We define SCF2 and prove that it satisfies both new principles. Furthermore, we discuss findings of a recent empirical cognitive study that provide additional support to SCF2.

1 Introduction

The formal study of argumentation is an important field of research within AI [19]. A central focus of this field has been the idea of Dung [15] that under some conditions, the acceptance of arguments depends only on a so-called *attack* relation among the arguments, and not on the internal structure of the arguments. Dung called this approach *abstract* argumentation and called the directed graph that represents the arguments as well as the attack relation between them an *argumentation framework* (AF). Whether an argument is deemed acceptable depends on the decision about other arguments. Therefore the basic concept in abstract argumentation is a *set* of arguments that can be accepted together, called an *extension*. Crucially, there may be several of such extensions, and these extensions may be incompatible. An *extension-based argumentation semantics* takes as input an AF and produces as output a set of extensions.

Two classes of extension-based argumentation semantics have been studied. Dung himself introduced several examples of so-called *admissibility-based* semantics, formalizing the idea that an argument is acceptable in the context of an extension if the extension *defends* the argument, i.e. attacks all the attackers of the argument. In this paper we consider his grounded, complete, preferred, and stable semantics. Moreover, we consider the admissibility-based semantics known as semi-stable semantics [22, 9]. The other kind of extension-based argumentation semantics are *naive-based* semantics, which are based on the idea that acceptable arguments sets are specific maximal conflict-free sets. In this paper we consider the naive, stage, CF2 and stage2 semantics and develop a new naive-based semantics called SCF2.

Abstract argumentation has various potential applications [19], and the choice of the semantics depends on the envisioned application. In this paper, we focus on the following research question: Which semantics corresponds well to what humans consider a rational judgment on the acceptability of arguments?

There are two systematic ways to approach this research question: A normative perspective is provided by the *principle-based approach* [3], in which semantics are evaluated based on their satisfaction of various normatively desirable principles. A descriptive perspective is provided by the *empirical approach* [18], in which cognitive studies are conducted to determine which semantics best predicts human judgments about arguments. In this paper, we combine both approaches.

Two recent empirical cognitive studies on argumentation semantics by Cramer and Guillaume [12, 13] showed CF2 to be better predictors of human argument evaluation than admissibility-based semantics like grounded and preferred. This finding sheds some doubt on principles that are only satisfied by admissibility-based semantics, e.g. Admissibility, Defence and Reinstatement as defined by van der Torre and Vesic [21]. For this reason, in this paper we focus on another existing principle, namely Directionality, and introduce two new ones.

The first new principle we consider is *Irrelevance of Necessarily Rejected Arguments* (*INRA*). Informally, INRA says that if an argument is attacked by every extension of an AF, then deleting this argument should not change the set of extensions. The idea here is that an argument that is attacked by every extension would be rejected by any party in a debate, and hence would never be brought up in a debate. Hence, it should be treated as if it did not even exist.

The second principle that we consider is *Strong Completeness Outside Odd Cycles* (*SCOOC*). Informally, SCOOC says that if an argument a and its attackers are not in an odd cycle, then an extension not containing any of a's attackers must contain a. The principle is based on the idea that it is generally desirable that an argument that is not attacked by any argument in a given extension should itself be in that extension. While it is possible to ensure this property in AFs without odd cycles, this is not the case for AFs involving an odd cycle. The idea behind the SCOOC principle is to still satisfy this property as much as possible, i.e. whenever the argument under consideration and its attackers are not in an odd cycle.

We show that of the nine common semantics mentioned above, the only ones that satisfy INRA are grounded, complete and naive semantics, Additionally, we show that a variant of CF2 that we call nsa(CF2) and that consists of first deleting all self-attacking arguments and then applying CF2 semantics also satisfies INRA.

Furthermore, we show that of these ten semantics (the nine mentioned at the beginning as well as nsa(CF2)), the only one that satisfies SCOOC is the stable semantics. But stable semantics satisfies neither Directionality nor INRA. The fact that none of the considered existing semantics satisfies both new principles introduced in this paper raises the question whether these two principles can be satisfied in conjunction. We answer this question positively by defining a novel semantics called *SCF2 semantics* that satisfies both of them.

Finally, we discuss findings of a recent cognitive study by Cramer and Guillaume [13] whose results suggest that SCF2 is more in line with the judgments of participants than any existing semantics. So our hypothesis that SCF2 corresponds well to what humans consider a rational judgment on the acceptability of arguments is motivated not only by theoretical but also by empirical observations. The robustness of these preliminary empirical findings will need to be tested in future studies.

All proofs of theorems in this paper can be fond in a technical report [14].

2 Preliminaries

In this section we define required notions from abstract argumentation theory [15, 2]. Additionally, we define three principles from the literature on principle-based argumentation [3, 21] and present an argument for the case that the Directionality principle is a desirable property for a semantics designed to match what humans would consider a rational judgment on the acceptability of arguments.

Definition 1. *An* argumentation framework (AF) $F = \langle Ar, att\rangle$ *is a finite directed graph in which the set* Ar *of vertices is considered to represent arguments and the set* att *of edges is considered to represent the attack relation between arguments, i.e. the relation between a counterargument and the argument that it counters.*

Definition 2. *An att-path is a sequence* $\langle a_0, \ldots, a_n\rangle$ *of arguments where* $(a_i, a_{i+1}) \in att$ *for* $0 \leq i < n$ *and where* $a_j \neq a_k$ *for* $0 \leq j < k \leq n$ *with either* $j \neq 0$ *or* $k \neq n$*. An* odd *att*-cycle *is an att-path* $\langle a_0, \ldots, a_n\rangle$ *where* $a_0 = a_n$ *and* n *is odd.*

Definition 3. *Let* $F = \langle Ar, att\rangle$ *be an AF, and let* $S \subseteq Ar$*. We write* $F|_S$ *for the restricted AF* $\langle S, att \cap (S \times S)\rangle$*. The set* S *is called* conflict-free *iff there are no arguments* $b, c \in S$ *such that* b *attacks* c *(i.e. such that* $(b, c) \in att$*). Argument* $a \in Ar$ *is* defended *by* S *iff for every* $b \in Ar$ *such that* b *attacks* a *there exists* $c \in S$ *such that* c *attacks* b*. We say that* S attacks a *if there exists* $b \in S$ *such that* b *attacks* a*, and we define* $S^+ = \{a \in Ar \mid S \text{ attacks } a\}$ *and* $S^- = \{a \in Ar \mid a \text{ attacks some } b \in S\}$*.*

- *S is a* complete extension *of F iff it is conflict-free, it defends all its arguments and it contains all the arguments it defends.*
- *S is a* stable extension *of F iff it is conflict-free and it attacks all the arguments of $Ar \setminus S$.*
- *S is the* grounded extension *of F iff it is a minimal with respect to set inclusion complete extension of F.*
- *S is a* preferred extension *of F iff it is a maximal with respect to set inclusion complete extension of F.*
- *S is a semi-stable extension of F iff it is a complete extension and there exists no complete extension S_1 such that $S \cup S^+ \subset S_1 \cup S_1^+$.*
- *S is a stage extension of F iff S is a conflict-free set and there exists no conflict-free set S_1 such that $S \cup S^+ \subset S_1 \cup S_1^+$.*
- *S is a naive extension of F iff S is a maximal conflict-free set.*

CF2 semantics was first introduced by Baroni *et al.* [4]. The idea behind it is that we partition the AF into *strongly connected components* and recursively evaluate it component by component by choosing maximal conflict-free sets in each component and removing arguments attacked by chosen arguments. We formally define it following the notation of Dvořák and Gaggl [16]. For this we first need some auxiliary notions:

Definition 4. *Let $F = \langle Ar, att \rangle$ be an AF, and let $a, b \in Ar$. We define $a \sim b$ iff either $a = b$ or there is an att-path from a to b and there is an att-path from b to a. The equivalence classes under the equivalence relation $\sim$ are called* strongly connected components *(SCCs) of F. We denote the set of SCCs of F by $SCCs(F)$. Given $S \subseteq Ar$, we define $D_F(S) := \{b \in Ar \mid \exists a \in S : (a, b) \in att \wedge a \not\sim b\}$.*

The simplified SCC-recursive scheme used for defining CF2 and stage2 is a function that maps a semantics σ to another semantics $\text{scc}(\sigma)$:

Definition 5. *Let σ be an argumentation semantics. The argumentation semantics $scc(\sigma)$ is defined as follows. Let $F = \langle Ar, att \rangle$ be an AF, and let $S \subseteq Ar$. Then S is an $scc(\sigma)$-extension of F iff either*

- *$|SCCs(F)| = 1$ and S is a σ-extension of F, or*
- *$|SCCs(F)| > 1$ and for each $C \in SCCs(F)$, $S \cap C$ is an $scc(\sigma)$-extension of $F|_{C \setminus D_F(S)}$.*

CF2 semantics is defined to be $\text{scc}(\textit{naive})$, and *stage2* semantics is defined to be $\text{scc}(\textit{stage})$.

Apart from the function scc, we introduce a further function – called nsa – that also maps a semantics to another semantics. Informally, the idea behind $\text{nsa}(\sigma)$ is that we first delete all self-attacking arguments and then apply σ. For defining nsa formally, we first need an auxiliary definition:

Definition 6. *Let $F = \langle Ar, att \rangle$ be an AF. We define the* non-self-attacking restriction *of F, denoted by $NSA(F)$, to be the AF $F|_{Ar'}$, where $Ar' := \{a \in Ar \mid (a, a) \notin att\}$.*

Definition 7. *Let σ be an argumentation semantics. The argumentation semantics $nsa(\sigma)$ is defined as follows. Let $F = \langle Ar, att \rangle$ be an AF, and let $S \subseteq Ar$. We say that E is an $nsa(\sigma)$-extension of F iff E is a σ-extension of $NSA(F)$.*

We now define the Directionality principle introduced by Baroni and Giacomin [3]. For this, we first need an auxiliary notion:

Definition 8. *Let $F = \langle Ar, att \rangle$ be an AF. A set $U \subseteq Ar$ is* unattacked *iff there exists no $a \in Ar \setminus U$ such that a attacks some $b \in U$.*

Definition 9. *A semantics σ satisfies the* Directionality *principle iff for every AF F and every unattacked set U, it holds that $\sigma(F|_U) = \{E \cap U \mid E \in \sigma(F)\}$.*

The Directionality principle corresponds to an important feature of the human practice of argumentation, namely that if a person has formed an opinion on some arguments and is confronted with new arguments, they will only feel compelled to reconsider their judgment on the prior arguments if one of the new arguments attacks one of the prior arguments. Apart from our own intuition, we can also refer to the results of an empirical cognitive study on argumentation that shows that humans are able to systematically judge the directionality of attacks between arguments [11]. Thus we consider the Directionality principle crucial for the goal that we focus on in this paper.

3 Two New Principles

The first new principle we consider is *Irrelevance of Necessarily Rejected Arguments* (*INRA*). Informally, INRA says that if an argument is attacked by every extension of an AF, then deleting this argument should not change the set of extensions. The idea here is that an argument that is attacked by every extension would not be held by any party, and hence would never be brought forwards in a debate. Hence, it should be treated as if it did not even exist.

In order to formally define the INRA principle, we first need to define a notation for an AF with one argument deleted:

Definition 10. *Let $F = \langle Ar, att \rangle$ be an AF and let $a \in Ar$ be an argument. Then F_{-a} denotes the restricted AF $F|_{Ar \setminus \{a\}}$.*

Definition 11. *Let σ be an argumentation semantics. We say that σ satisfies* Irrelevance of Necessarily Rejected Arguments *(*INRA*) iff for every AF $F = \langle Ar, att \rangle$ and every argument $a \in Ar$, if every $E \in \sigma(F)$ attacks a, then $\sigma(F) = \sigma(F_{-a})$.*

The second principle that we consider is *Strong Completeness Outside Odd Cycles* (*SCOOC*). Informally, SCOOC says that if an argument a and its attackers are not in an odd cycle, then an extension not containing any of a's attackers must contain a.

In order to formally define the Strong Completeness Outside Odd Cycles principle, we first need to define the auxiliary notion of a set of arguments being *strongly complete outside odd cycles*.

Definition 12. *Let $F = \langle Ar, att \rangle$ be an AF, and let $A \subseteq Ar$. We say that A is* strongly complete outside odd cycles *iff for every argument $a \in Ar$, if no argument in $\{a\} \cup \{a\}^-$ is in an odd att-cycle and $A \cap \{a\}^- = \emptyset$, then $a \in A$.*

Definition 13. *Let σ be an argumentation semantics. We say that σ satisfies* Strong Completeness Outside Odd Cycles (SCOOC) *iff for any AF F, every σ-extension of F is strongly complete outside odd cycles.*

The SCOOC principle is related to the property of *strong completeness*: An extension E is *strongly complete* iff every argument not attacked by E is in E. We call this property *strong completeness* as it is a strengthening of completeness, which states that every argument defended by E is in E.

The stable semantics is the only widely studied argumentation semantics that satisfies strong completeness. More precisely, the stable semantics can be characterized by the conjunction of conflict-freeness and strong completeness. In other words, one can say that the stable semantics is motivated by the idea that a violation of strong completeness constitutes a paradox and should therefore be avoided.

The stable semantics satisfies strong completeness at the price of allowing for situations in which there are no extensions and hence no judgment can be made on any argument whatsoever. Such cases are always due to odd *att*-cycles. So we can say that odd *att*-cycles – unless resolved through arguments attacking the odd cycle – cause paradoxical situations. The idea of most semantics other than stable semantics is to somehow contain these paradoxes so that they do not affect our ability to make judgments about completely or sufficiently unrelated arguments.

The idea of the SCOOC principle is that while in odd cycles we may not be able to avoid a paradoxical judgments about the arguments, i.e. a judgment in which an argument is not accepted even though none of its attackers is accepted, such paradoxical judgments should be completely avoided outside of odd cycles.

How does that differ from the containment of paradoxical situations provided by existing semantics? Admissibility-based semantics do not allow for any judgment about an argument in an unattacked odd cycle; however this undecided status is not limited to odd cycles, but carries forward to arguments that are not in an odd cycle but that are *att*-reachable from an odd cycle.

Naive-based semantics like CF2, stage and stage2 allow for judgments about arguments in an unattacked odd cycle, but also at the cost of affecting the way arguments that are not in odd cycles are interpreted. For example, CF2 allows for a six-cycle to be interpreted in a doubly paradoxical way despite the fact that it is an even cycle that can be interpreted in a non-paradoxical manner. This behavior of CF2 was also considered problematic by Dvořák and Gaggl [16], who used this example to motivate their stage2 semantics, but as established by Theorem 4 below, stage2 also fails to avoid paradoxical judgments about arguments that are not themselves involved in an odd cycle.

The SCOOC principle was designed to systematically identify whether a semantics suffers from this problem. As it turns out, all the standard semantics other than stable do suffer from the problem, i.e. do not satisfy SCOOC.

We will now look at which semantics satisfy or do not satisfy each of the two principles that we have defined.

Theorem 1. *The grounded, complete, naive and nsa(CF2) semantics satisfy INRA.*

Theorem 2. *Stable, preferred, semi-stable, stage, stage2 and CF2 semantics violate INRA.*

Theorem 3. *Stable semantics satisfies SCOOC.*

Theorem 4. *Complete, grounded, preferred, semi-stable, naive, stage, CF2, stage2 and nsa(CF2) semantics violate SCOOC.*

4 SCF2 Semantics

In this section, we define and study the new semantics SCF2, which satisfies both of the new principles introduced in the previous section as well as the Directionality principle defined in the preliminaries. Furthermore, we will motivate the design choices in the definition of SCF2 by looking at how semantics defined in a similar way as SCF2 fail to satisfy at least one of Directionality, INRA or SCOOC.

We have seen in the previous section that nsa(CF2) satisfies INRA but does not satisfy SCOOC. The idea behind the definition of SCF2 is that we modify the definition of nsa(CF2) by already enforcing SCOOC at the level of the single SCCs considered in the SCC-recursive definition of nsa(CF2). For this, we define a variant of naive semantics called *SCOOC-naive semantics*.

Definition 14. *Let $F = \langle Ar, att \rangle$ be an AF, and let $A \subseteq Ar$. We say that A is an* SCOOC-naive extension *of F if A is subset-maximal among the conflict-free subsets of Ar that are strongly complete outside odd cycles.*

Recall that CF2 is defined to be scc(naive), i.e. nsa(CF2) = nsa(scc(naive)). For defining SCF2, we just replace naive semantics by SCOOC-naive semantics in this definition.

Definition 15. *SCF2 semantics is defined to be nsa(scc(SCOOC-naive)).*

In other words, SCF2 works by first deleting all self-attacking arguments and then applying the SCC-recursive scheme that is also used in the definition of CF2, but applying SCOOC-naive semantics instead of naive semantics to each single SCC. SCF2 satisfies Directionality, INRA and SCOOC, which we have argued to be desirable principles when evaluating a semantics designed to correspond well to what humans would consider a rational judgment on the acceptability of arguments. The somewhat complex definition of SCF2 raises the question whether a simpler definition could also be enough to satisfy these three principles.

To approach this question systematically, we would like to point out that the definition of SCF2 contains three features that distinguishes it from naive semantics: It starts by deleting all self-attacking arguments (the function nsa), it proceeds by applying the SCC-recursive scheme (the function scc), and within each SCC, it applies SCOOC-naive rather than naive semantics. If we consider each of these three features a switch that we can switch on or off, we have eight definitions of semantics, namely naive, nsa(naive), SCOOC-naive, nsa(SCOOC-naive), scc(naive), nsa(scc(naive)), scc(SCOOC-naive) and nsa(scc(SCOOC-naive)). One can easily see that naive = nsa(naive), so these eight definitions define only seven different semantics, whose properties we now study in order to show that only SCF2 satisfies all three principles Directionality, INRA and SCOOC.

Table 1 shows which of these seven semantics satisfies which of these three principles (we use the standard name CF2 for scc(naive) and use the short name SCF2 to refer to nsa(scc(SCOOC-naive))). Note that SCF2 satisfies all three principles, while no other of these seven semantics satisfies all three principles.

	Directionality	INRA	SCOOC
naive = nsa(naive)	×	✓	×
SCOOC-naive	×	×	✓
nsa(SCOOC-naive)	×	×	✓
CF2	✓	×	×
nsa(CF2)	✓	✓	×
scc(SCOOC-naive)	✓	×	✓
SCF2	✓	✓	✓

Table 1. Properties of SCF2 and six semantics that are related to it with respect the three principles considered in this paper

The results displayed in Table 1 are proven in a technical report [14]. Additionally, we prove there that every AF has an SCF2 extension.

5 Empirical cognitive studies

Rahwan et al. [19] argue that Artificial Intelligence research will benefit from the interplay between logic and cognition and that therefore "logicians and computer scientists ought to give serious attention to cognitive plausibility when assessing formal models of reasoning, argumentation, and decision making". Based on the observation that in the previous literature on formal argumentation theory, an example-based approach and a principle-based approach were used to motivate and validate argumentation semantics, they propose to complement these approaches by an *experiment-based approach* that takes into account empirical cognitive studies on how humans interpret and evaluate arguments. They made a first contribution to this new approach by presenting and discussing the results of two such studies that they conducted in order to test the cognitive plausibility of simple and floating reinstatement [19].

While the argumentation frameworks used in Rahwan et al.'s studies could not distinguish between preferred semantics and naive-based semantics like CF2, two more recent studies by Cramer and Guillaume [12, 13] address this issue. Both of these studies made use of a group discussion methodology that is known to stimulate more rational thinking. According to the results of the first study [12], CF2, SCF2, stage and stage2 semantics are significantly better predictors for human judgments on the acceptability of arguments than admissibility-based semantics like grounded, preferred, complete or semi-stable (binomial tests, all p-values < 0.001). However, this study did not involve argumentation frameworks that allow to distinguish between CF2, SCF2, stage and stage2 semantics.

According to the results of Cramer and Guillaume's second study [13], SCF2, CF2 and grounded semantics are better predictors for human judgments on the acceptability of arguments than stage, stage2, preferred or semi-stable semantics (binomial tests, all p-values < 0.001). Additionally, the results suggest that SCF2 is a better predictor than CF2 and grounded semantics, but the results for that are not significant. We will now explain these results in more depth.

As explained in Section 3, Dvořák and Gaggl [16] critique a feature of CF2 semantics, namely that in the case of a six-cycle, CF2 allows two opposite arguments to be accepted together. The second study by Cramer and Guillaume [13] confirms that this criticism is in line with human judgments of argument acceptability. We briefly summarize the data on which this judgment is made (a more detailed explanation can by found in [13]): Based on the overall responses of the participants in the study, Cramer and Guillaume point out that 12 of the 61 participants of their study have a high frequency of incoherent responses, so that they disconsider them from the further analysis. Among the remaining 49 participants, 22 follow a simple cognitive strategy of marking arguments as *Undecided* whenever there is a reason for doubt (in line with the grounded semantics), while 27 participants do not follow this strategy. Cramer and Guillaume call these 27 participants the *coherent non-grounded participants*.

In the case of 11 out of the 12 argumentation frameworks considered in the study, the majority of these 27 coherent non-grounded participants make judgements that are in line with CF2 semantics. The only exception to this is an argumentation framework involving a six-cycle, in which only 33% of the coherent non-grounded participants make a judgement in line with CF2 semantics, while 60% make a judgements that is in line with SCF2, stage2, preferred and semi-stable semantics.

Dvořák and Gaggl [16] themselves had used this criticism against CF2 to motivate their stage2 semantics, but in the study by Cramer and Guillaume [13], stage2 performed significantly worse than SCF2, since all five arguments in the only AF on which stage2 and SCF2 differed were evaluated by most participants (including most coherent non-grounded participants) in line with SCF2 rather than with stage2.

In combination with the principle-based argument for SCF2 presented in the previous two sections, these preliminary findings provide additional support for

our hypothesis that SCF2 corresponds well to what humans consider a rational judgment on the acceptability of arguments.

6 Related work

In the previous section we have already considered related empirical work. In this section we focus on work related to the principle-based approach to abstract argumentation that we have employed in this paper.

The principle-based analysis of argumentation semantics was initiated by Baroni and Giacomin [3] to choose among the many extension-based argumentation semantics that have been proposed in the formal argumentation literature. The handbook chapter of van der Torre and Vesic [21] gives a classification of fifteen alternatives for argumentation semantics using twenty-seven principles discussed in the literature on abstract argumentation. Dvořák and Gaggl [16] introduce stage2 semantics by showing how it satisfies various desirable properties, similarly to how we motivate SCF2 semantics in this paper.

Moreover, additional extension-based argumentation semantics and principles have been proposed by various authors. For example, Besnard *et al.* [6] introduce a system for specifying semantics in abstract argumentation called SESAME. Moreover, many principles have been proposed for alternative semantics of argumentation frameworks such as ranking semantics [1], and for extended argumentation frameworks, for example for abstract dialectical frameworks [8].

The principle of Irrelevance of Necessarily Rejected Arguments is closely related to the well-studied area of dynamics of argumentation, in which also various principles have been proposed which are closely related to INRA. Cayrol *et al.* [10] were maybe the first to study revision of frameworks using a principle-based analysis, and they have been related to notions of equivalence [5, 17]. Boella *et al.* [7] define principles for abstracting (i.e., removing) an argument, and Rienstra *et al.* [20] define a variety of persistence and monotony properties for argumentation semantics. Our INRA principle is inspired by and closely related to the *skeptical IO monotony principle* that they define. The difference is that their principle considers adding an attack rather than removing an argument.

7 Conclusion and Future Work

Motivated by empirical cognitive studies on argumentation semantics, we have introduced a new naive-based argumentation semantics called SCF2. A principle-based analysis shows that it has two distinguishing features:

1. If an argument is attacked by all extensions, then it can never be used in a dialogue and therefore it has no effect on the acceptance of other arguments. We call it *Irrelevance of Necessarily Rejected Arguments*.
2. Within each extension, if none of the attackers of an argument is accepted and the argument is not involved in a paradoxical relation, then the argument is accepted. We define paradoxicality as being part of an odd cycle, and we call this principle *Strong Completeness Outside Odd Cycles*.

We have argued that these features together with the findings from empirical cognitive studies make SCF2 a good candidate for an argumentation semantics that corresponds well to what humans consider a rational judgment on the acceptability of arguments.

The empirical approach to abstract argumentation theory is still a relatively new approach that needs to be developed further by modifying and improving the methodology of existing studies in the design of future studies. The current paper provides a well-motivated hypothesis that can be tested more rigorously in future empirical studies, namely the hypothesis that SCF2 predicts human judgments on the acceptability of arguments better than other abstract argumentation semantics.

On the theoretical side, more work is required to determine which other principles studied in the literature are satisfied by SCF2. Moreover, dialogue-based decision procedures must be defined, and the complexity of the various decision problems must be established. Finally, an extension towards structured argumentation should be investigated.

References

1. L. Amgoud and J. Ben-Naim. Ranking-Based Semantics for Argumentation Frameworks. In W. Liu, V. S. Subrahmanian, and J. Wijsen, editors, *Scalable Uncertainty Management*, pages 134–147, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
2. P. Baroni, M. Caminada, and M. Giacomin. Abstract argumentation frameworks and their semantics. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, pages 159–236. College Publications, 2018.
3. P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10):675–700, 2007. Argumentation in Artificial Intelligence.
4. P. Baroni, M. Giacomin, and G. Guida. SCC-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168(1):162–210, 2005.
5. R. Baumann. Normal and strong expansion equivalence for argumentation frameworks. *Artif. Intell.*, 193:18–44, 2012.
6. P. Besnard, S. Doutre, V. H. Ho, and D. Longin. SESAME - A System for Specifying Semantics in Abstract Argumentation. In M. Thimm, F. Cerutti, H. Strass, and M. Vallati, editors, *Proceedings of the First International Workshop on Systems and Algorithms for Formal Argumentation (SAFA) co-located with the 6th International Conference on Computational Models of Argument (COMMA 2016), Potsdam, Germany, September 13, 2016.*, volume 1672 of *CEUR Workshop Proceedings*, pages 40–51. CEUR-WS.org, 2016.
7. G. Boella, S. Kaci, and L. W. N. van der Torre. Dynamics in Argumentation with Single Extensions: Abstraction Principles and the Grounded Extension. In C. Sossai and G. Chemello, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 10th European Conference, ECSQARU 2009, Verona, Italy, July 1-3, 2009. Proceedings*, volume 5590 of *Lecture Notes in Computer Science*, pages 107–118. Springer, 2009.
8. G. Brewka, S. Ellmauthaler, H. Strass, J. Wallner, and S. Woltran. *Abstract dialectical frameworks*. College Publications, International, 2018.

9. M. W. A. Caminada, W. A. Carnielli, and P. E. Dunne. Semi-stable semantics. *J. Log. Comput.*, 22(5):1207–1254, 2012.
10. C. Cayrol, F. D. de Saint-Cyr, and M. Lagasquie-Schiex. Revision of an Argumentation System. In G. Brewka and J. Lang, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*, pages 124–134. AAAI Press, 2008.
11. M. Cramer and M. Guillaume. Directionality of attacks in natural language argumentation. In C. Schon, editor, *Proceedings of the Workshop on Bridging the Gap between Human and Automated Reasoning*, volume 2261, pages 40–46. RWTH Aachen University, CEUR-WS.org, 2018. http://ceur-ws.org/Vol-2261/.
12. M. Cramer and M. Guillaume. Empirical Cognitive Study on Abstract Argumentation Semantics. *Frontiers in Artificial Intelligence and Applications*, pages 413–424, 2018.
13. M. Cramer and M. Guillaume. Empirical Study on Human Evaluation of Complex Argumentation Frameworks. In *Proceedings of JELIA 2019*, 2019. Full paper available at http://icr.uni.lu/mcramer/downloads/2019_JELIA.pdf.
14. M. Cramer and L. van der Torre. SCF2 – an Argumentation Semantics for Rational Human Judgments on Argument Acceptability: Technical Report. *arXiv e-prints*, Aug 2019.
15. P. M. 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.
16. W. Dvořák and S. A. Gaggl. Stage semantics and the SCC-recursive schema for argumentation semantics. *Journal of Logic and Computation*, 26(4):1149–1202, Aug 2016.
17. E. Oikarinen and S. Woltran. Characterizing strong equivalence for argumentation frameworks. *Artificial Intelligence*, 175(14-15):1985–2009, 2011.
18. I. Rahwan, M. I. Madakkatel, J.-F. Bonnefon, R. N. Awan, and S. Abdallah. Behavioral Experiments for Assessing the Abstract Argumentation Semantics of Reinstatement. *Cognitive Science*, 34(8):1483–1502, 2010.
19. I. Rahwan and G. R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 1st edition, 2009.
20. T. Rienstra, C. Sakama, and L. W. N. van der Torre. Persistence and Monotony Properties of Argumentation Semantics. In E. Black, S. Modgil, and N. Oren, editors, *Theory and Applications of Formal Argumentation – Revised Selected Papers*, volume 9524 of *Lecture Notes in Computer Science*, pages 211–225. Springer, 2015.
21. L. van der Torre and S. Vesic. The principle-based approach to abstract argumentation semantics. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*. College Publications, 2018.
22. B. Verheij. Two Approaches to Dialectical Argumentation: Admissible Sets and Argumentation Stages. In *In Proceedings of the biannual International Conference on Formal and Applied Practical Reasoning (FAPR) workshop*, pages 357–368. Universiteit, 1996.