**Künstliche Intelligenz manuscript No.**
(will be inserted by the editor)

# Measuring Inconsistency in Multi-Agent Systems

**A. Hunter · S. Parsons · M. Wooldridge**

Received: date / Accepted: date

**Abstract** We introduce and investigate formal quantitative measures of inconsistency between the beliefs of agents in multi-agent systems. We start by recalling a well-known model of belief in multi-agent systems, and then, using this model, present two classes of inconsistency metrics. First, we consider metrics that attempt to characterise the overall degree of inconsistency of a multi-agent system in a single numeric value, where inconsistency is considered to be individuals within the system having contradictory beliefs. While this metric is useful as a high-level indicator of the degree of inconsistency between the beliefs of members of a multi-agent system, it is of limited value for understanding the structure of inconsistency in a system: it gives no indication of the *sources* of inconsistency. We therefore introduce metrics that quantify for a given individual the extent to which that individual is in conflict with other members of the society. These metrics are based on power indices, which were developed within the cooperative game theory community in order to understand the power that individuals wield in cooperative settings.

Anthony Hunter
Department of Computer Science
University College London, UK
E-mail: a.hunter@cs.ucl.ac.uk

Simon Parsons
Department of Computer Science
University of Liverpool, UK
E-mail: s.d.parsons@liv.ac.uk

Michael Wooldridge
Department of Computer Science
University of Oxford, UK
E-mail: mjw@cs.ox.ac.uk

**Keywords** multi-agent systems · inconsistency · Shapley value

## 1 Introduction

In a seminal 1988 paper, Alan Bond and Les Gasser attempted to summarise the key challenges facing the then-nascent multi-agent systems research area [3]. One of the five key challenges they identified was the question of "how to recognise and reconcile disparate viewpoints and conflicting intentions among a collection of agents" [3, p.10]. They advocated the development of principled techniques for understanding, managing, and resolving such inconsistencies. In this paper, we study ways to recognise the source and nature of inconsistencies in a mullti-agent systems as a necessary prescursor to managing them.

Inconsistency in multi-agent systems can manifest itself in the form of inconsistent beliefs (I believe taxes are bad; my spouse believes taxes are good), or in the form of inconsistent preferences (I prefer the family to holiday in California; my spouse prefers the family to holiday in France). Resolving inconsistencies between beliefs is, in multi-agent systems research, primarily the domain of argumentation. Resolving inconsistencies between preferences is primarily the domain of social choice theory and computational social choice.

Our specific aim in the present paper is to develop techniques that enable us to understand both the *scale* and the *structure* of inconsistencies between the *beliefs* of agents in a multi-agent system. There are several reasons why it may be important to obtain an understanding of the scale and

structure of belief inconsistencies in a multi-agent system. Most obviously, when inconsistency occurs within a team of agents, there is typically a need to resolve that inconsistency, which may be time-consuming or costly. If we are to put agents into teams, it therefore makes sense to investigate beforehand the scale and structure of any inconsistency in the team, so that potential inconsistencies can be minimised. Moreover, we might sometimes interpret the fact that one agent is grossly inconsistent with other agents as an indicator of faults, potentially requiring maintenance, or at least meriting further attention.

The model of belief we adopt in the present paper is *sentential* [15]: the belief system of each agent is characterised by a knowledge base of formulae of some logical belief language, together with a set of deduction rules for the belief language. A multi-agent system is given by a set of such deduction structures, one for each agent in the system. Using these multi-agent belief systems as our starting point, we develop two classes of metrics for measuring and analysing belief inconsistency in multi-agent systems. First, we define *societal* measures of inconsistency. These measures give us a single numeric value that quantifies the overall degree of inconsistency in a multi-agent system. We find it necessary to provide more than one such measure because, in order to determine whether the beliefs of agents in a multi-agent system are inconsistent, we must aggregate the beliefs of the agents in the system in some way, and there are many possible ways of aggregating beliefs. Second, we define *individual* measures of inconsistency. These measures attempt to characterise the extent to which the beliefs of individual agents in a system are in conflict with those of other agents in the system. By using these individual measures of inconsistency, we can analyse the structure of inconsistency in a system, by identifying the *sources* of inconsistency. The formulation of our inconsistency measures is derived from the use of power indices in cooperative game theory: in particular, the Banzhaf index and Shapley value [5], building on the use of the Shapley value for measuring inconsistency developed in [12]. Power indices are used in cooperative game theory to evaluate the contribution a particular agent makes in a cooperative setting, and in voting theory, they are used to measure the power that a particular agent has, where power is understood as the ability to influence a particular outcome.

Throughout the paper, we assume some familiarity with logic (e.g., the notion of deduction rules and deductive proof) and computational complexity. We provide a brief summary of the relevant concepts from cooperative game theory, although space limitations prevent any discussion of these concepts – see [5].

## 2 Preliminary Definitions

**Logic Notation:** We assume some prior knowledge of classical logic (e.g., the notion of a rule of inference), but present a brief summary of our key notational conventions, etc. Let $\mathbb{B} = \{\top, \bot\}$ be the set of Boolean truth values, with "$\top$" being truth and "$\bot$" being falsity. We will abuse notation a little by using $\top$ and $\bot$ to denote both the syntactic constants for truth and falsity respectively, as well as their semantic counterparts. Let $\Phi = \{p, q, \ldots\}$ be a denumerable vocabulary of Boolean variables, and let $\mathscr{L}_0$ denote the set of (well-formed) formulae of classical propositional logic over $\Phi$, constructed using the conventional Boolean operators ("$\wedge$", "$\vee$", "$\rightarrow$", "$\leftrightarrow$", and "$\neg$"), as well as the truth constants "$\top$" and "$\bot$". Where $\varphi \in \mathscr{L}_0$, we let $vars(\varphi)$ denote the (possibly empty) set of Boolean variables occurring in formula $\varphi$ (e.g., $vars(p \wedge q) = \{p, q\}$). We write $\models \varphi$ to mean that $\varphi$ is a tautology.

**Agents:** We assume a fixed and finite set $N = \{1, \ldots, n\}$ of *agents*. A *coalition*, $C$, is simply a subset of $N$, $C \subseteq N$. The *grand coalition* is the set of all agents $N$. Notice that in everyday use, the term "coalition" typically implies some common purpose, or commitment to common action. We do not use the term in this sense: a coalition in this paper is nothing more than a set of agents.

**Belief Languages:** We model the beliefs of agents using a *sentential* approach, and more specifically, using the *deduction model of belief* developed by Konolige [15]. With this model, the belief system of an agent $i \in N$ is modelled as a set of formulae of a *logical belief language*, $\mathscr{L}_B$. For example, it could be that $\mathscr{L}_B = \mathscr{L}_0$, i.e., that the language used by agents to represent their beliefs is in fact classical propositional logic. In general, we won't assume much of $\mathscr{L}_B$, except that it is a *logical* language, with a well-defined syntax, semantics, and proof theory, and with notions such as sound and complete inference defined for it.

**Deduction Rules and Deduction Structures:** A *deduction rule* for $\mathscr{L}_B$ is a rule of inference that has a fixed and finite number of premises, and that is an effectively computable function of its premises [15, p.21]. Where $\rho$ is a set of deduction rules for $\mathscr{L}_B$, $\varphi \in \mathscr{L}_B$ is an $\mathscr{L}_B$-formula, and $\Delta \subseteq \mathscr{L}_B$ is a set of $\mathscr{L}_B$-formulae, we denote by $\Delta \vdash_\rho \varphi$ the fact that $\varphi$ may be derived from $\Delta$ using the deduction rules $\rho$. A *deduction structure*, $d_i$, for an agent $i \in N$ is a pair:

$d_i = \langle \Delta_i, \rho_i \rangle$, where: $\Delta_i \subseteq \mathscr{L}_B$ is a fixed, finite set of *base beliefs*; and $\rho_i$ is a fixed, finite set of *deduction rules* for $\mathscr{L}_B$. Where $d_i = \langle \Delta_i, \rho_i \rangle$ is a deduction structure, we denote by $bel(d_i)$ the closure of $\Delta_i$ under $\rho_i$, i.e.,

$bel(\langle \Delta, \rho \rangle) = \{\varphi \mid \Delta \vdash_\rho \varphi\}$.

**Measures of Consistency and Inconsistency:** There has recently been interest in techniques for analysing (in)consistency in logical knowledge bases [12, 10]. Though our interest is mainly in inconsistencies that arise across the knowledge bases of multiple agents, we start with a simple and intuitive measure of inconsistency for the knowledge base of a single agent. To do this we define a two-place function $\mathscr{I}$ so that $\mathscr{I}(\Delta, \rho)$ evaluates to 1 if $bel(\langle \Delta, \rho \rangle)$ contains an explicit contradiction, and 0 otherwise:

$$\mathscr{I}(\Delta, \rho) = \begin{cases} 1 \text{ if } \exists \varphi \in \mathscr{L}_B : \{\varphi, \neg\varphi\} \subseteq bel(\langle \Delta, \rho \rangle) \\ 0 \text{ otherwise.} \end{cases}$$

Thus $\mathscr{I}(\Delta, \rho)$ will evaluate to 1 iff $bel(\langle \Delta, \rho \rangle)$ contains an explicit contradiction, i.e., a formula and the negation of that formula.

Observe that if $\mathscr{L}_B = \mathscr{L}_0$, and $\rho$ is a sound and complete set of deduction rules for $\mathscr{L}_0$, then $\mathscr{I}(\cdots)$ will characterise logical consistency for $\mathscr{L}_0$. If $\rho$ is sound but incomplete, then $\mathscr{I}(\cdots)$ will capture a weaker notion of consistency: namely, whether *explicit contradictions* can be detected by applying the rules $\rho$.

**Multi-agent belief systems:** A *multi-agent belief system* is a structure $B = \langle N, d_1, \ldots, d_n \rangle$, where $N = \{1, \ldots, n\}$ is a set of agents, and $d_i = \langle \Delta_i, \rho_i \rangle$ is a deduction structure for agent $i$, capturing the beliefs of agent $i$.

Where $B = \langle N, d_1, \ldots, d_n \rangle$ is a multi-agent belief system, and $C \subseteq N$ ($C \neq \emptyset$) is a coalition, then we define:

$$\Delta_C^{\cup} = \bigcup_{i \in C} \Delta_i \qquad \Delta_C^{\cap} = \bigcap_{i \in C} \Delta_i$$

$$\rho_C^{\cup} = \bigcup_{i \in C} \rho_i \qquad \rho_C^{\cap} = \bigcap_{i \in C} \rho_i$$

$$bel_C^{\cup} = \bigcup_{i \in C} bel(d_i) \; bel_C^{\cap} = \bigcap_{i \in C} bel(d_i)$$

We will say a multi-agent belief system $B$ is *monotonic* if for all $C \subseteq N$ and for all $\Gamma_1 \subseteq \mathscr{L}_B$, if $\Gamma_1 \vdash_{\rho_C^{\cup}} \varphi$ then for all $\Gamma_2 \subseteq \mathscr{L}_B$ we have $\Gamma_1 \cup \Gamma_2 \vdash_{\rho_C^{\cup}} \varphi$.

*Example 1* Consider a multi-agent belief system $B_1$ with agents $N = \{1, 2\}$ such that: $\Delta_1 = \{p, p \to q\}$; $\rho_1 =$ a sound and complete set of deduction rules for $\mathscr{L}_0$; $\Delta_2 = \{q \to \neg p\}$; and $\rho_2 = \emptyset$. We have:

$\mathscr{I}(\Delta_1, \rho_1) = \mathscr{I}(\Delta_2, \rho_2) = \mathscr{I}(\Delta_{\{1,2\}}^{\cap}, \rho_{\{1,2\}}^{\cap}) = 0$
$\mathscr{I}(\Delta_{\{1,2\}}^{\cup}, \rho_{\{1,2\}}^{\cup}) = 1.$

**Cooperative Games and Power Indices:** We use some definitions from the area of cooperative game theory [5]. A *simple cooperative game* is a pair $G = \langle N, v \rangle$, where $N = \{1, \ldots, n\}$ is a set of *players*, and $v : 2^N \to \{0, 1\}$ is the *characteristic function* of the game, which assigns to every set of agents a binary value. If $v(C) = 1$ then we say that $C$ is a *winning coalition*, while if $v(C) = 0$, we say $C$ is a *losing coalition*. We require that $v(\emptyset) = 0$. We say that $G = \langle N, v \rangle$ is *monotone* if $v(C) \geq v(D)$ for every pair of coalitions $C, D \subseteq N$ such that $C \supseteq D$. If $G_1 = \langle N, v_1 \rangle$ and $G_2 = \langle N, v_2 \rangle$ are simple cooperative games with the same player set, we will say they are *equivalent*, and write $G_1 \equiv G_2$, if $v_1(C) = v_2(C)$ for all $C \subseteq N$.

Agent $i$ is a *swing player* for $C \subseteq N \setminus \{i\}$ if $C$ is not winning but $C \cup \{i\}$ is. We find it useful to define a function $swing(C, i)$ so that this function returns 1 if $i$ is a swing player for $C$, and 0 otherwise, i.e.,

$$swing(C, i) = \begin{cases} 1 \text{ if } v(C) = 0 \text{ and } v(C \cup \{i\}) = 1 \\ 0 \text{ otherwise.} \end{cases}$$

The *Banzhaf score* for agent $i$, denoted $\sigma_i$, is the number of coalitions for which $i$ is a swing player:

$$\sigma_i = \sum_{C \subseteq N \setminus \{i\}} swing(C, i). \tag{1}$$

The *Banzhaf measure*, denoted $\mu_i$, is the probability that $i$ would be a swing player for a coalition chosen at random from $2^{N \setminus \{i\}}$:

$$\mu_i = \frac{\sigma_i}{2^{n-1}} \tag{2}$$

The *Banzhaf index* for player $i \in N$, denoted by $\beta_i$, is the proportion of coalitions for which $i$ is a swing to the total number of swings in the game – thus the Banzhaf index is a measure of relative power, since it takes into account the Banzhaf score of other agents:

$$\beta_i = \frac{\sigma_i}{\sum_{j \in N} \sigma_j} \tag{3}$$

Finally, we define the *Shapley value*; here the *order* in which agents join a coalition plays a role. Let $P(N)$ denote the set of all permutations of $N$, with typical members $\varpi, \varpi'$, etc. If $\varpi \in P(N)$ and $i \in N$, then let $prec(i, \varpi)$ denote the players that precede $i$ in the ordering $\varpi$. (For example, if $\varpi = (a_3, a_1, a_2)$, then $prec(a_2, \varpi) = \{a_1, a_3\}$.) Given this, let $\varsigma_i$ denote the Shapley value of $i$, defined as follows:

$$\varsigma_i = \frac{1}{n!} \sum_{\varpi \in P(A)} swing(prec(i, \varpi), i) \tag{4}$$

A key result in cooperative game theory is that the Shapley value is uniquely characterised by a small set of axioms[1]. Later, when we consider the Shapley value in the context of multi-agent belief systems, we will return to these axioms. To state the axioms, we need some additional terminology. We say player $i \in N$ is a *dummy* if for all coalitions $C \subseteq N$, $v(C \cup \{i\}) = v(C)$. We say players $i,j$ are *symmetric* if for all coalitions $C \subseteq (N \setminus \{i,j\})$ we have $swing(C,i) = swing(C,j)$. It is then well-known that if $i$ is a dummy then $\varsigma_i = 0$, while if $i$ and $j$ are symmetric then $\varsigma_i = \varsigma_j$. Note that the Banzhaf index also satisfies these axioms.

## 3 Societal Measures of Inconsistency

We now move on to the first main concern of this paper: measuring in a principled way the degree of inconsistency present in a multi-agent system. In this section, we will explore *societal* measures of inconsistency: measures of inconsistency that quantify the degree of inconsistency present in society as a whole, irrespective of the properties of individual members of the society. In subsequent sections, we will consider the problem of measuring how inconsistent individuals are with respect to society.

As a starting point, we consider the probability that a non-empty coalition $C \subseteq N$ selected uniformly at random from $2^N \setminus \emptyset$ will have inconsistent beliefs, under the assumption that the beliefs and deduction rules of coalition members are simply pooled together through set theoretic union. We denote this value for a multi-agent belief system $B$ by $\mathscr{S}^{\cup}(B)$:

$$\mathscr{S}^{\cup}(B) = \frac{1}{2^n - 1} \sum_{\substack{C \subseteq N \\ C \neq \emptyset}} \mathscr{I}(\Delta_C^{\cup}, \rho_C^{\cup})$$

In other words, $\mathscr{S}^{\cup}(B)$ is $E[\mathscr{I}(\Delta_C^{\cup}, \rho_C^{\cup})]$, i.e., the expected value of $\mathscr{I}(\Delta_C^{\cup}, \rho_C^{\cup})$ for a coalition $C$ picked uniformly at random. Notice that the value $\mathscr{S}^{\cup}(B)$ captures a "liberal" notion of inconsistency, in the sense that it treats every agent's base beliefs $\Delta$ and deduction rules $\rho$ equally: the base beliefs and deduction rules of every agent are pooled together and conclusions derived. But this is a rather crude way of pooling the beliefs of agents in a system. For example, suppose one agent $i$ is an intuitionistic reasoner, and does not include the law of the excluded middle in his rule set, while other agents are classical reasoners. Then it is possible that some of the conclusions derived from $i$'s base beliefs would not in fact be supported by $i$ (if for example they were derived using the law of the excluded middle).

There are of course many ways of aggregating beliefs, which we will not discuss here (see for example [18]). We present just one alternative – a more *conservative* measure of societal inconsistency, $\mathscr{S}^{\cap}(B)$:

$$\mathscr{S}^{\cap}(B) = \frac{1}{2^n - 1} \sum_{\substack{C \subseteq N \\ C \neq \emptyset}} \mathscr{I}(\Delta_C^{\cap}, \rho_C^{\cap})$$

Thus, the value $\mathscr{S}^{\cap}(B)$ only takes into account base beliefs and deduction rules that are universally accepted. Notice that if $\mathscr{S}^{*}(B) = 1$ for $* \in \{\cup, \cap\}$ then *every* (non-empty) coalition is inconsistent, and in particular, this implies that all the agents within the system have individually inconsistent belief sets.

*Example 2* Referring back to the multi-agent belief system $B_1$ defined in Example 1, we have $\mathscr{S}^{\cap}(B_1) = 0$, and $\mathscr{S}^{\cup}(B_1) = \frac{1}{3}$.

Let us state some properties of these measures.

**Proposition 1**

1. *For all monotonic multi-agent belief systems B, we have:*
   $\mathscr{S}^{\cup}(B) \geq \mathscr{S}^{\cap}(B)$.
2. *There exist monotonic multi-agent belief systems B such that:*
   $\mathscr{S}^{\cup}(B) > \mathscr{S}^{\cap}(B)$.

*Proof* For point (1), suppose $\{\varphi, \neg\varphi\} \subseteq bel(\langle \Delta_C^{\cap}, \rho_C^{\cap} \rangle)$ for some $\varphi \in \mathscr{L}_B$. Then $\{\varphi, \neg\varphi\} \subseteq bel(\langle \Delta_C^{\cup}, \rho_C^{\cup} \rangle)$ from the monotonicity of $B$. It follows that for all $C \subseteq N$, if $\mathscr{I}(\Delta_C^{\cap}, \rho^{\cap}) = 1$ then $\mathscr{I}(\Delta_C^{\cup}, \rho^{\cup}) = 1$, hence

$$\sum_{C \subseteq N : C \neq \emptyset} \mathscr{I}(\Delta_C^{\cup}, \rho_C^{\cup}) \;\geq\; \sum_{C \subseteq N : C \neq \emptyset} \mathscr{I}(\Delta_C^{\cap}, \rho_C^{\cap}),$$

and so $\mathscr{S}^{\cup}(B) \geq \mathscr{S}^{\cap}(B)$. Example 2 serves as a proof of point (2).

## 4 Individuals and Social Consistency

The social (in)consistency metrics we introduced above attempt to quantify the inherent overall (in)consistency of a multi-agent system through a single numeric value. However, returning to the overall aims of this work, giving a single inconsistency value for an entire system gives no information about the *sources* or *structure* of inconsistency,

[1] In the present paper, we will not be concerned with the axiom known as *additivity*.

which will be an important consideration for example if one is to try to resolve or settle the inconsistency. With this consideration in mind, in this section we present measures that characterise *the extent to which individuals influence the consistency of a* system. We start with a motivating example.

*Example 3* Assume $\mathscr{L}_B = \mathscr{L}_0$. Suppose we have a multi-agent belief system $B$ with $N = \{1,2,3,4\}$ and $\Delta_1 = \Delta_2 = \Delta_3 = \{p\}$ while $\Delta_4 = \{\neg p\}$. All agents $i \in N$ have deduction rules $\rho_i$ that are sound and complete for $\mathscr{L}_0$. We have:

$$\mathscr{I}(\Delta_C^{\cup}, \rho_C^{\cup}) = \begin{cases} 1 \text{ if } |C| > 1 \text{ and } 4 \in C \\ 0 \text{ otherwise.} \end{cases}$$

It follows that $\mathscr{I}^{\cup}(B) = \frac{7}{15}$. However, it is intuitively obvious that there is just one agent in this scenario that is the source of inconsistency: agent 4, who believes $\neg p$, while every other agent believes $p$. And yet this agent seems to have quite a dramatic influence on the overall inconsistency of the society, according to the measure $\mathscr{I}^{\cup}(B)$.

This example clearly demonstrates the need for techniques that give a more fine-grained analysis of inconsistency within a multi-agent system, and in particular, techniques that allow us to clearly isolate the sources of inconsistency. The metrics we now present are intended for this purpose.

Where $B = \langle N, d_1, \ldots, d_n \rangle$ is a multi-agent belief system and $* \in \{\cup, \cap\}$ is one of the set theoretic operations of union or intersection, we define a cooperative game $G_B^* = \langle N, v_B^* \rangle$ containing the same set of agents, and with characteristic function $v_B^*$ defined as follows:

$$v_B^*(C) = \mathscr{I}(\Delta_C^*, \rho_C^*).$$

Thus, in the game $G_B^*$, a coalition is "winning" ($v_B^*(C) = 1$) if they are inconsistent (taking $*$ as the aggregation operator for beliefs and rules), and "losing" ($v_B^*(C) = 0$) if they are consistent using the aggregation operator $*$. Note that we do not, of course, mean "winning" in the sense of this being a good thing – we simply follow the terminology of cooperative game theory, and say a coalition are winning if they obtain a value of 1.

We have now established a precise formal relationship between the notion of inconsistency in multi-agent belief systems, and simple cooperative games. With this relationship in place, we will shortly see how power indices from cooperative game theory can be used to analyse inconsistency in multi-agent systems. However, before we do that, let us pause to consider the relationship we have established in a little more detail. We have defined a mapping from multi-agent belief systems to simple cooperative games. It is easy to see that this mapping is many-to-one, in the sense that multiple multi-agent belief systems can map to the same cooperative game. Moreover, the mapping is total, in the sense that every multi-agent belief system induces a simple cooperative game. However, what about the other direction of the mapping? Is it the case that every simple cooperative game is induced by some multi-agent belief system? If we restrict our consideration to monotonic reasoners, the answer is no:

**Proposition 2** *For every monotonic multi-agent belief system $B = \langle N, d_1, \ldots, d_n \rangle$, the corresponding game $G_B^{\cup} = \langle N, v_B^{\cup} \rangle$ is monotone. It follows that there exist simple cooperative games $G = \langle N, v \rangle$ such that for all monotonic multi-agent belief systems $B = \langle N, d_1, \ldots, d_n \rangle$, we have $G \not\equiv G_B^{\cup}$.*

*Proof* We must show that $v_B^{\cup}(C) \geq v_B^{\cup}(D)$ for every pair of coalitions $C, D \subseteq N$ such that $C \supseteq D$. So consider the value $v_B^{\cup}(C)$. There are two possibilities: $v_B^{\cup}(C) = 0$ or $v_B^{\cup}(C) = 1$. Where $v_B^{\cup}(C) = 0$, suppose for sake of contradiction that $v_B^{\cup}(D) = 1$. Then $\exists\{\varphi, \neg\varphi\} \subseteq bel(\langle \Delta_D^{\cup}, \rho_D^{\cup} \rangle)$. Now, since $B$ is monotonic, $bel(\langle \Delta_C^{\cup}, \rho_C^{\cup} \rangle) \supseteq bel(\langle \Delta_D^{\cup}, \rho_D^{\cup} \rangle)$, which implies $\{\varphi, \neg\varphi\} \subseteq bel(\langle \Delta_C^{\cup}, \rho_C^{\cup} \rangle)$ and hence $v_B^{\cup}(C) = 1$; contradiction. Where $v_B^{\cup}(C) = 1$, then $v_B^{\cup}(C) \geq v_B^{\cup}(D)$ follows from the fact that $v_B^{\cup}(D) \in \{0,1\}$.

Now, with the games $G_B^*$ defined, we can directly apply the power indices that were defined earlier. These metrics can be understood as *quantifying the extent to which individual agents affect the consistency of a society*. Formally, where $B$ is a multi-agent belief system and $* \in \{\cup, \cap\}$, we use the following notation:

- $\sigma_i^*(B)$ is the Banzhaf score of player $i$ in the game $G_B^*$, that is, $\sigma_i^*(B)$ is the total number of coalitions that player $i$ is in contradiction with;
- $\mu_i^*(B)$ is the Banzhaf measure of player $i$ in the game $G_B^*$, that is, $\mu_i^*(B)$ is the probability that player $i$ would be in contradiction with a coalition $C$ selected uniformly at random from the set of all possible non-empty coalitions;
- $\beta_i^*(B)$ is the Banzhaf index of player $i$ in game $G_B^*$, that is, $\beta_i^*(B)$ measures the proportion of coalitional inconsistencies in $B$ that $i$ is responsible for;
- $\varsigma_i^*(B)$ is the Shapley value of player $i$ in the game $G^*(B)$, that is, $\varsigma_i^*(B)$ measures the probability that a player $i$ would make the grand coalition inconsistent, taking into account all possible ways in which the grand coalition could form.

| Agent | $\sigma_i^{\cup}(B_2)$ | $\mu_i^{\cup}(B_2)$ | $\beta_i^{\cup}(B_2)$ | $\varsigma_i^{\cup}(B_2)$ |
|---|---|---|---|---|
| 1 | 1 | $\frac{1}{8}$ | $\frac{1}{10}$ | $\frac{1}{12}$ |
| 2 | 1 | $\frac{1}{8}$ | $\frac{1}{10}$ | $\frac{1}{12}$ |
| 3 | 1 | $\frac{1}{8}$ | $\frac{1}{10}$ | $\frac{1}{12}$ |
| 4 | 7 | $\frac{7}{8}$ | $\frac{7}{10}$ | $\frac{9}{12}$ |

**Table 1** Values $\sigma_i^{\cup}(B_2)$, $\mu_i^{\cup}(B_2)$, $\beta_i^{\cup}(B_2)$, and $\varsigma_i^{\cup}(B_2)$ for agents $i \in \{1,2,3,4\}$. See Example 4.

*Example 4* Consider the multi-agent belief system $B_2$ presented in Example 3. Table 1 summarises the measures $\sigma_i^{\cup}(B_2)$, $\mu_i^{\cup}(B_2)$, $\beta_i^{\cup}(B_2)$, and $\varsigma_i^{\cup}(B_2)$ for agents $i \in \{1,2,3,4\}$. To better understand these values, consider player 1. This player will be a swing player for a coalition $C$ (i.e., will make a coalition $C$ inconsistent) exactly when the coalition $C$ contains agent 4 and no other players. (If other players are in the coalition with 4, it will already be inconsistent, and player 1 will not affect this status.) So $\sigma_1^{\cup}(B_2) = 1$, and similarly for players 2 and 3. Turning to player 4, however, this player will make *any* non-empty coalition $C$ inconsistent, and so $\sigma_1^{\cup}(B_2) = 7$.

At this point, let us return to the axioms for the Shapley value, and try to understand them in terms of our model. To do this, we will need a little extra terminology.

First, we say an *amiable* agent is one that can be added to any consistent set of agents without causing the set to become inconsistent. Thus an amiable agent is not the cause of any conflicts that arise in the set of agents. Formally, $i$ is amiable iff:

$$\forall C \subseteq N : \mathscr{I}(\Delta_C^{\cup}, \rho_C^{\cup}) \geq \mathscr{I}(\Delta_{C\cup\{i\}}^{\cup}, \rho_{C\cup\{i\}}^{\cup}).$$

Then, we say one agent *matches* another if adding either to a set of agents has the same outcome. In other words, two agents match if they are in conflict with the same sets of agents as each other. Obviously this is a symmetrical relationship. Formally, for agents $\{i,j\} \subseteq N$, we say $i$ matches $j$ iff:

$$\forall C \subseteq (N \setminus \{i,j\}) : \mathscr{I}(\Delta_{C\cup\{i\}}^{\cup}, \rho_{C\cup\{i\}}^{\cup}) = \mathscr{I}(\Delta_{C\cup\{j\}}^{\cup}, \rho_{C\cup\{j\}}^{\cup}).$$

**Proposition 3** *For all $B = \langle N, d_1, \ldots, d_n \rangle$:*

1. *If $i$ is amiable in $B$, then $i$ is a dummy player in $G_B^{\cup}$.*
2. *If $i$ matches $j$ in $B$, then $i$ and $j$ are symmetric in $G_B^{\cup}$.*

From standard properties of the Shapley value we obtain:

**Proposition 4** *For all $B = \langle N, d_1, \ldots, d_n \rangle$:*

1. *If $i$ is amiable in $B$, then $\sigma_i^{\cup}(B) = \mu_i^{\cup}(B) = \beta_i^{\cup}(B) = \varsigma_i^{\cup}(B) = 0$.*
2. *If $i$ matches $j$ in $B$, then $\sigma_i^{\cup}(B) = \sigma_j^{\cup}(B)$, $\mu_i^{\cup}(B) = \mu_j^{\cup}(B)$, $\beta_i^{\cup}(B) = \beta_j^{\cup}(B)$, and $\varsigma_i^{\cup}(B) = \varsigma_j^{\cup}(B)$.*

We say that one agent *believes more than another agent* when the closure of the beliefs of the first agent with its proof rules is a superset of the the closure of the beliefs of the second agent with its proof rules. Formally, for agents $i,j \in C$, $i$ believes more than $j$ iff $bel(d_j) \subseteq bel(d_i)$. Then we have, for example:

**Proposition 5** *If $i$ believes more than $j$, then $\beta_i \geq \beta_j$*

## 5 Non-Monotonic Believers

So far in this paper, we have assumed that the reasoning processes employed by agents, as characterised in their belief set $\Delta$ and reasoning rules $\rho_i$, are classical, and in particular, monotonic. In this section, we will explore what happens if we assume that agents use *non-monotonic* reasoning [4]. Although non-monotonic reasoning is a well established research area in the knowledge representation field, it is perhaps less well known in the multi-agent systems area, and so we provide a brief summary of the main ideas.

In classical logic (of which propositional logic and first-order logic are the two key systems), deduction is monotonic, in the sense that the theorems of a theory will expand monotonically with the premises of the theory. More formally, a logical deduction system $\vdash$ is said to be monotonic if $\Delta_1 \vdash \varphi$ implies $\Delta_1 \cup \Delta_2 \vdash \varphi$ for all sets of formulae $\Delta_1$ and $\Delta_2$; this is true of both classical propositional and first-order logic. However, some types of common sense reasoning do not have this property. To use a well-known example, if we are told Tweety is a bird, then we might conclude that Tweety can fly. However, if we are later told that Tweety is a penguin, then we might *retract* this conclusion. The common sense reasoning we are employing here is thus *non-monotonic*, since we are retracting a conclusion (Tweety can fly) after adding more premises (Tweety is a penguin). To formalise such reasoning, many *non-monotonic logics* have been developed, of which *default logic*, *autoepistemic logic* and *circumscription* are perhaps the best known examples (see, e.g., [4] for an introduction and key references).

Now, in our setting, if agents use classical, monotonic forms of reasoning, then this means that *adding* an agent $i \in N$ to a coalition $C \subseteq N$ can only have one of three consequences:

1. the coalition is mutually consistent before $i$ was added, and continue to be consistent after $i$ is added;

2. the coalition is mutually consistent before $i$ was added, but the addition of $i$ makes them inconsistent;
3. the coalition is mutually inconsistent before $i$ was added, and continue to be inconsistent after $i$ is added.

Thus, adding an agent in a setting of monotonic reasoning agents can never *recover* consistency, in the sense that if the members of $C$ are mutually inconsistent, then adding $i$ can only result in continuing inconsistency. However, as we will now see, this need not be true if we allow non-monotonic reasoning agents.

To make the ideas precise, we use a (simplified) autoepistemic logic [16]. Syntactically, the language $\mathscr{L}_{AE}$ of autoepistemic logic extends classical propositional logic with a unary modal modality $K$, where the expression $K\varphi$ should be read "it is known that $\varphi$". Formally, the syntax of $\mathscr{L}_{AE}$ is given by the following grammar:

$$\chi ::= p \mid \neg\chi \mid \chi \vee \chi \mid K\chi$$

where $p$ is a Boolean variable. The remaining classical operators are assumed to be defined as abbreviations in the standard way.

The semantics of $\mathscr{L}_{AE}$ are given with respect to pairs $(\Delta,\pi)$, where $\Delta \subseteq \mathscr{L}_{AE}$ is the belief base, and $\pi : \Phi \to \mathbb{B}$ is a classical valuation for the Boolean variables $\Phi$:

$(\Delta,\pi) \models p$ iff $\pi(p) = \top$ (for $p \in \Phi$);
$(\Delta,\pi) \models K\varphi$ iff $\varphi \in \Delta$;
$(\Delta,\pi) \models \neg\chi$ iff it is not the case that $(\Delta,\pi) \models \chi$;
$(\Delta,\pi) \models \chi_1 \vee \chi_2$ iff either $(\Delta,\pi) \models \chi_1$ or $(\Delta,\pi) \models \chi_2$ or both.

The remaining classical connectives are defined in terms of $\vee$ and $\neg$ in the standard way.

We will say a formula of $\mathscr{L}_{AE}$ is *strict* if all Boolean variables appear within the scope of an autoepistemic modality. Thus $Kp$ is strict, while $q \wedge Kp$ is not, because $q$ does not appear within the scope of an autoepistemic modality. Strict formula can be evaluated with respect to belief sets $\Delta$: we do not need to refer to the valuation function $\pi$ when evaluating such formulae. So, for strict formulae $\chi$, we will write $\Delta \models \chi$ to mean that $\chi$ is true when evaluated against belief set $\Delta$.

The notion of logical consequence from an initial set of premises $I$ through $\Delta$ is naturally defined: we write $I \models_\Delta \varphi$. Now, given an initial belief set $\Delta$, an *expansion* is a set $T_\Delta$ that satisfies the following fixpoint equation:

$$T_\Delta = \{\varphi \mid \Delta \models_{T_\Delta} \varphi\}. \quad (5)$$

It is easy to see that such a set $T_\Delta$ satisfies the following properties:

(B1) $T_\Delta$ is closed under propositional consequence;
(B2) $\varphi \in T_\Delta$ implies $K\varphi \in T_\Delta$;
(B3) $\varphi \notin T_\Delta$ implies $\neg K\varphi \in T_\Delta$.

In general, starting from an initial set of beliefs $\Delta$, there may be multiple possible expansions $T_\Delta$, or indeed none[2]. Moreover, determining questions associated with such expansions is computationally hard – typically at the second level of the polynomial hierarchy [8]. To keep things simple, we will therefore consider a restricted autoepistemic reasoning system, which is sufficiently powerful to allow us to express key non-monotonic properties, but which avoids the complexities of "full" autoepistemic logic.

Let us say a *simple autoepistemic rule* is an implication of the form $\chi \to \varphi$, where $\chi$ is a strict autoepistemic formula, and $\varphi$ is a propositional formula. Now, following the terminology of the present paper, we will consider agents that are equipped with a deduction structure $d = \langle\Delta,\rho\rangle$, where $\Delta$ is a finite set of propositional logic formulae representing the base beliefs of the agent, and $\rho$ is a finite set of simple autoepistemic rules.

Given an $\mathscr{L}_{AE}$ deduction structure $d = \langle\Delta,\rho\rangle$, we denote the belief set associated with $\langle\Delta,\rho\rangle$ by $bel(\langle\Delta,\rho\rangle)$, where this set is the smallest set of propositional formulae satisfying the following fixed point equation:

$$bel(\langle\Delta,\rho\rangle) = \Delta \cup \{\varphi \mid \chi \to \varphi \in \rho \text{ and } bel(\langle\Delta,\rho\rangle) \models \chi\} \quad (6)$$

It should be clear that this definition represents a significant simplification of the notion of an expansion as defined in equation (5): most importantly, we are only permitting a highly restricted form of $\mathscr{L}_{AE}$ formulae in $\rho$ (i.e., simple autoepistemic rules), and we are only permitting propositional base beliefs $\Delta$. But the key point about this construction is that, (as we will see shortly), it permits a meaningful type of non-monotonic reasoning, while having the following highly desirable properties:

**Proposition 6** *For all $\mathscr{L}_{AE}$-deduction structures $\langle\Delta,\rho\rangle$, the set $bel(\langle\Delta,\rho\rangle)$ is well-defined, finite, and unique, and can be computed in polynomial time.*

Let us see an example of an $\mathscr{L}_{AE}$ deduction structure, and how it achieves non-monotonic reasoning.

*Example 5* Suppose

$$\Delta = \{bird(tweety)\}$$

[2] Consider the case where $\Delta = \{\neg Kp \to q, \neg Kq \to p\}$. In this case there are two expansions of $\Delta$: one containing $p$ but not $q$, the other containing $q$ but not $p$. The set $\Delta = \{Kp\}$ has no expansions.

and

$$\rho = \{K\,bird(tweety) \wedge \neg K\,penguin(tweety) \rightarrow flies(tweety)\}.$$

Then

$$bel(\langle \Delta, \rho \rangle) = \{bird(tweety), flies(tweety)\}.$$

However, if

$$\Delta = \{bird(tweety), penguin(tweety)\}$$

then

$$bel(\langle \Delta, \rho \rangle) = \{bird(tweety), penguin(tweety)\}.$$

Thus, we do not conclude that Tweety can fly if Tweety is known to be a penguin.

So, let us suppose that we have multi-agent belief systems $B = (N, d_1, \ldots, d_n)$ with $\mathscr{L}_{AE}$ deduction structures. First, let us see a small example.

*Example 6* Suppose we have a multi-agent belief system $B$ with $N = \{1,2,3\}$, $\Delta_1 = \{p\}$, $\Delta_2 = \emptyset$, $\Delta_3 = \{q\}$, $\rho_1 = \emptyset$, $\rho_2 = \{\neg Kq \rightarrow \neg p\}$, and $\rho_3 = \emptyset$. Now, $\Delta^{\cup}_{\{1,2\}} = \{p\}$, $\rho^{\cup}_{\{1,2\}} = \{\neg Kq \rightarrow \neg p\}$, and so $bel(\langle \Delta^{\cup}_{\{1,2\}}, \rho^{\cup}_{\{1,2\}} \rangle) = \{p, \neg p\}$. and so $\nu^{\cup}_B(\{1,2\}) = 1$: players 1 and 2 are inconsistent. However, if we add player 3, we have: $\Delta^{\cup}_{\{1,2,3\}} = \{p,q\}$, $\rho^{\cup}_{\{1,2,3\}} = \{\neg Kq \rightarrow \neg p\}$, hence $bel(\langle \Delta^{\cup}_{\{1,2,3\}}, \rho^{\cup}_{\{1,2,3\}} \rangle) = \{p,q\}$, and so $\nu^{\cup}_B(\{1,2,3\}) = 0$. Thus, adding player 3 transforms the coalition from inconsistency to consistency, intuitively by giving them an additional piece of information that allows them to avoid applying the single autoepistemic rule $\neg Kq \rightarrow \neg p$.

Example 6 serves as a proof of the following (contrast with Proposition 2):

**Proposition 7** *There exist $\mathscr{L}_{AE}$ multi-agent belief systems $B$ for which the corresponding game $G^{\cup}_B = \langle N, \nu^{\cup}_B \rangle$ is not monotone.*

In fact, the autoepistemic reasoning framework we have defined is rich enough to capture all simple cooperative games (contrast with Proposition 2):

**Proposition 8** *For every simple cooperative game $G = \langle N, \nu \rangle$ there exists an $\mathscr{L}_{AE}$ multi-agent belief system $B$ such that $G \equiv G^{\cup}_B$.*

*Proof* Given $G$, we construct an $\mathscr{L}_{AE}$ multi-agent belief system $B$ and show that $G \equiv G^{\cup}_B$. Let $W_G$ denote the set of winning coalitions in $G$:

$$W_G = \{C \subseteq N \mid \nu(C) = 1\}.$$

For each player $i \in N$, define a Boolean variable $x_i$, and in addition define one further variable $z$. Given a coalition $C \subseteq N$, define an $\mathscr{L}_{AE}$ formula $\psi_C$ as follows:

$$\psi_C = \left( \bigwedge_{i \in C} Kx_i \right) \wedge \left( \bigwedge_{j \in (N \setminus C)} \neg Kx_j \right).$$

Now define an $\mathscr{L}_{AE}$ formula $\Psi_{W_G}$ characterising the set of all winning coalitions of $G$:

$$\Psi_{W_G} = \bigvee_{C \in W_G} \psi_C.$$

For each agent $i \in N$, define $\Delta_i = \{x_i, z\}$. Now define a single autoepistemic rule $r$ as follows:

$$r = \Psi_{W_G} \rightarrow \neg z$$

and set $\rho_i = \{r\}$ for all $i \in N$. We now claim that:

$$\forall C \subseteq N : \nu(C) = 1 \text{ iff } \nu^{\cup}_B(C) = 1.$$

To see this, observe that for all $C \subseteq N (C \neq \emptyset)$, we have $\langle \Delta^{\cup}_C, \rho^{\cup}_C \rangle \models \Psi_{W_G}$ iff $C \in W_G$. Now, since $z \in \Delta^{\cup}_C$, for all $C \subseteq N (C \neq \emptyset)$, we have $\{z, \neg z\} \subseteq bel(\langle \Delta^{\cup}_C, \rho^{\cup}_C \rangle)$ iff $C \in W_G$. Thus $\mathscr{I}(\Delta^{\cup}_C, \rho^{\cup}_C) = 1$ iff $C \in W_G$.

Recall that in Section 4 we asked whether it is the case that every simple cooperative game is induced by some multi-agent belief system composed of monotonic reasoners, and found that the answer was no. The above result shows that if we allow agents to be non-montonic reasoners then it is possible to induce every simple cooperative game from some mullti-agent belief system.

Next, recall that, when discussing monotonic reasoning agents, we showed that taking the union of a set of beliefs would always give rise to at least as much inconsistency as taking the intersection (Proposition 1). We now show that this does not hold for non-monotonic believers.

**Proposition 9** *There exist $\mathscr{L}_{AE}$ multi-agent belief systems $B$ such that $\mathscr{I}^{\cap}(B) > \mathscr{I}^{\cup}(B)$.*

Let us now consider the complexity of computing the various inconsistency measures with respect to $\mathscr{L}_{AE}$ belief systems. We have the following result, which can be understood as saying that computing power indices such as $\beta^{\cup}_i(B)$ is in the same complexity class as computing these indices in many other cooperative game settings [5]:

**Proposition 10** *Given an $\mathscr{L}_{AE}$ multi-agent belief system $B = \langle N, d_1, \ldots, d_n \rangle$ and an agent $i \in N$, the problem of computing $\sigma^{\cup}_i(B)$ is #P-complete. It follows that computing $\mu^{\cup}_i(B)$ and $\beta^{\cup}_i(B)$ is #P-complete with respect to Turing reductions.*

*Proof* For membership, consider a non-deterministic Turing machine that first guesses a coalition $C \subseteq N$ and then accepts iff the following condition is satisfied:

$$\exists\{\varphi, \neg\varphi\} \in bel(\langle \Delta_C^{\cup}, \rho_C^{\cup} \rangle).$$

Computing $bel(\langle \Delta_C^{\cup}, \rho_C^{\cup} \rangle)$ can be done in polynomial time, and so the condition can be evaluated in polynomial time. The number of accepting computations of the Turing machine is exactly the number of coalitions $C \subseteq N$ such that $\mathscr{I}(\langle \Delta_C^{\cup}, \rho_C^{\cup} \rangle) = 1$, and so computing $\sigma_i$ is in #P.

For hardness we reduce #SAT: the problem of computing the number of satisying assignment for a given propositional formula $\varphi$. Without loss of generality, we can assume that the #SAT instance $\varphi$ is in CNF; assume $vars(\varphi) = \{x_1, \ldots, x_k\}$. Now, denote by $\varphi^*$ the strict $\mathscr{L}_{AE}$ formula obtained from $\varphi$ by systematically replacing each positive literal $x_i$ that occurs in $\varphi$ by $Kx_i$ and each negative literal $\neg x_i$ by $\neg Kx_i$. Observe that since $\varphi$ is assumed to be in CNF, the formula $\varphi^*$ that we obtain through this transformation is indeed a strict $\mathscr{L}_{AE}$ formula. We now define a multi-agent belief system $B_\varphi$ as follows. For each variable $x_i \in vars(\varphi)$ we create an agent $a_i$, and also create one additional agent $a_{k+1}$. In addition to the variables $\{x_1, \ldots, x_k\}$ we create a new variable $z$. For all $1 \leq i \leq k$ we define $\Delta_i = \{x_i, z\}$, and $\rho_i = \emptyset$. Finally we define $\Delta_{k+1} = \emptyset$ and $\rho_{k+1} = \{\varphi^* \rightarrow \neg z\}$. We now claim that $\sigma_{k+1}^{\cup}(B_\varphi)$ is exactly the number of satisfying assignments for the #SAT instance $\varphi$. To see this, first observe that for all $C \subseteq \{a_1, \ldots, a_k\}$, $\mathscr{I}(\Delta_C^{\cup}, \rho_C^{\cup}) = 0$ since $\Delta_C^{\cup}$ is a set of positive literals and $\rho_C^{\cup} = \emptyset$. Now, for all $C \subseteq \{a_1, \ldots, a_k\}$, $swing(C, a_{k+1}) = 1$ iff the propositional assignment

$$\pi : \{x_1, \ldots, x_k\} \rightarrow \{\top, \bot\}$$

defined by:

$$\pi(x_i) = \begin{cases} \top \text{ if } a_i \in C \\ \bot \text{ otherwise} \end{cases}$$

is such that $\pi \models \varphi$. Thus $\sigma_{k+1}^{\cup}(B_\varphi)$ is exactly the number of satisfying assignments for the #SAT instance $\varphi$.

## 6 Related Work & Conclusions

In contemporary multi-agent systems research, *argumentation* is perhaps the key approach to handling inconsistency [2, 19]. Argumentation can be understood as being concerned with developing techniques for deriving justifiable, rational conclusions from knowledge bases that contain inconsistencies. Argumentation is not, however, primarily concerned with understanding the structure or source of inconsistency, which is the aim of the present paper. In argumentation research, the aim instead is largely to understand what counts as a *rational position* in the presence of inconsistency.

Recent work on measuring inconsistency in logical knowledge bases is closely related to the present paper [13, 14, 9, 12]. However, this work is focused on measuring the inconsistency of a logical theory, rather than multi-agent inconsistency. Measuring inconsistency has proven to be a useful tool in analysing various information types [11]. We should also note recent work by Ågotnes *et al.* on using power indices from cooperative game theory to analyse how *knowledgeable* individual agents are with respect to a particular formula of epistemic logic [1]. They use the setting of possible worlds semantics to analyse knowledge, but the ideas are similar to ours. The basic idea is to use cooperative solution concepts to try to quantify the extent to which an agent "contributes" to knowledge of a particular fact. The key difference is that we focus on quantifying inconsistency, essentially by pooling the knowledge of agents in the system. While there are several well-known models of mutual knowledge used in the literature of possible worlds semantics [7], these models do not admit the possibility of inconsistency: if an agent is considered to know something, then that thing must be true. Our approach is somewhat similar in that we use a model of *belief* but the relationship is superficial. Our work uses a *sentential* model of belief, based on that proposed by [15].

We began this paper by citing Bond and Gasser [3] and their argument that a key challenge in multi-agent systems is to be able to recognise and reconcile disparate viewpoints. The work we have presented so far provides a mechanism for *recognising* inconsistency. A natural question to ask at this point is how this helps us in *reconciling* different viewpoints. In this section we give one answer to this question, showing how the inconsistency measures can potentially reduce the computational effort in reconciling the beliefs of a set of agents. The approach we consider for reconciling beliefs is argumentation, in particular, the argumentation-based persuasion dialogue studied in [17]. This dialogue is a process by which two agents with inconsistent sets of base beliefs $\Delta_i$ can establish a consistent set of beliefs[3]. Parsons [17] proves that the number of messages exchanged by the agents in this process is proportional to the size of the agents' sets of base beliefs, but looking at the proof in detail reveals

[3] Exactly how they achieve this isn't relevant here, but in essence they recursively construct a grounded extension [6] so that when the dialogue terminates both agents agree on the acceptability of a common set of beliefs.

that this is a loose upper limit — the number of moves is bounded by the size of the belief base because in the worst case the agents will disagree on *every* formula in the belief base and have to work through the resolution process for each one in turn. In fact, the number of messages exchanged is determined by the number of inconsistencies — the agents will have to go through one round of persuasion for each pair of formulae that are inconsistent.

Now, consider that a set of agents is trying to identify a coalition that will engage in some task. The choice of possible coalitions will depend on what abilities different agents can bring to the task. However, if we make the reasonable assumption that the agents will need to reach consensus about their beliefs in order to complete their task, then since the amount of effort this will require is dependent on the number of inconsistencies, using the Banzhaf index $\beta_i$ (which we recall identifies the proportion of inconsistencies that an agent is responsible for) can be used to help select coalition members, and hence can reduce the work that coalitions then have to do.

Of course, computing $\beta_i$ is not cheap, but it is a computation cost, not a communication cost (unlike the cost of resolving the inconsistency). In domains in which communication is expensive, it may well be worth selecting coalitions to minimise inconsistencies, rather than attempting to resolve inconsistencies at run-time.

**Acknowledgements** Wooldridge gratefully acknowledges the support of the ERC under Advanced Investigator Grant 291528 ("RACE").

**Anthony Hunter** is a professor of artificial intelligence, and head of the Intelligent Systems Research Group, in the UCL Department of Computer Science. His research interests are in: computational models of argument for decision-making and sense-making; modelling and strategies of participants in persuasion; measuring and analysing inconsistency; and systems for aggregating knowledge.

**Simon Parsons** is Professor in Autonomous Systems and co-director of the smARTlab in the Department of Computer Science at the University of Liverpool. His current research interests are: computational argumentation, especially models of argument for reasoning about trust and its impact on decision making, market-based systems, and coordination mechanisms for multi-robot teams.

**Michael Wooldridge** is a Professor in the Department of Computer Science at the University of Oxford, and Senior Research Fellow at Hertford College Oxford. His current research interests are focused around the formalisation and computational analysis of rational action in multi-agent systems.

## References

1. Ågotnes, T., van der Hoek, W., Wooldridge, M.: Scientia potentia est. In: Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2011). Taipei, Taiwan (2011)
2. Besnard, P., Hunter, A.: Elements of Argumentation. The MIT Press: Cambridge, MA (2008)
3. Bond, A.H., Gasser, L. (eds.): Readings in Distributed Artificial Intelligence. Morgan Kaufmann Publishers: San Mateo, CA (1988)
4. Brewka, G., Dix, J., Konolige, K. (eds.): Nonmonotonic Reasoning: An Overview. Center for the Study of Language and Information (1997)
5. Chalkiadakis, G., Elkind, E., Wooldridge, M.: Computational Aspects of Cooperative Game Theory. Morgan-Claypool (2011)
6. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. Artificial Intelligence **77**, 321–357 (1995)
7. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning About Knowledge. The MIT Press: Cambridge, MA (1995)
8. Gottlob, G.: Complexity results for nonmonotonic logics. Journal of Logic and Computation **2**, 397–425 (1992)
9. Grant, J., Hunter, A.: Measuring inconsistency in knowledgebases. Journal of Intelligent Information Systems **27**, 159–184 (2006)
10. Grant, J., Hunter, A.: Measuring consistency gain and information loss in stepwise inconsistency resolution. In: Proceedings of European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (LNCS 6717), pp. 362–373. Springer-Verlag: Berlin, Germany (2011)
11. Hunter, A.: How to act on inconsistent news: Ignore, resolve, or reject. Data and Knowledge Engineering **57**, 221–239 (2006)
12. Hunter, A., Konieczny, S.: On the measure of conflicts: Shapley inconsistency values. Artificial Intelligence **174**, 1007–1026 (2011)
13. Knight, K.M.: Measuring inconsistency. Journal of Philosophical Logic **31**, 77–98 (2002)
14. Konieczny, S., Lang, J., Marquis, P.: Quantifying information and contradiction in propositional logic through epistemic tests. In: Proceedings of the 18th International Joint Conference on Artificial Intellignce (IJCAI'03), pp. 106–111 (2003)
15. Konolige, K.: A Deduction Model of Belief. Pitman Publishing: London and Morgan Kaufmann: San Mateo, CA (1986)
16. Marek, W., Truszczynski, M.: Autoepistemic logic. Journal of the ACM **38**(3), 588–619 (1991)
17. Parsons, S., Wooldridge, M., Amgoud, L.: Properties and complexity of some formal inter-agent dialogues. Journal of Logic and Computation **13**(3), 347–376 (2003)

18. Pigozzi, G.: Belief merging and the discursive dilemma: An argument-based account to paradoxes of judgment aggregation. Synthese **152**(2), 285–298 (2006)
19. Rahwan, I., Simari, G.R. (eds.): Argumentation in Artificial Intelligence. Springer-Verlag: Berlin, Germany (2009)