Research Statement

Iyad Rahwan
[1]Masdar Institute of Science & Technology, Abu Dhabi, UAE
[3](Visiting Scholar) Massachusetts Institute of Technology, USA
[2](Fellow) School of Informatics, University of Edinburgh, UK

November 2, 2010

Abstract

My work is concerned with engineering interaction patterns that enable groups (of computers or people) to be more effective. My early work focused on engineering protocols for computerized agents to coordinate by means of automated negotiation and argumentation, building on techniques from artificial intelligence, game theory, and argumentation theory. My recent work focuses on using persuasive information technology to influence people's behavior for the better (e.g. enable them to achieve better outcomes in negotiation, or encourage them to reduce their energy consumption).

1 Background

> "*In modern research, studying information and communication really involves three main ingredients. Logic supplies the theory, cognitive reality supplies the phenomena that we are interested in. But there is always a third party involved, viz. the role of computation of some sort.*"
> Johan van Benthem

An intrinsic property of many modern computing paradigms (e.g. service-oriented computing, grid computing, peer-to-peer systems, social networks, electronic markets, and smart environments) is the distribution of information and control among multiple entities (or agents), be they software, human or a mix of both. These systems can rarely operate in isolation, nor can they be controlled centrally. Thus, local interaction is essential for these systems to deliver their individual or collective functions.

As distributed systems grew in sophistication, increasingly richer metaphors were used to study and build them, such as biological, economic and social metaphors (see table below). When individual systems have explicit *cognitive* representations of their internal states, the notion *argumentation* became a natural and promising metaphor.

Argumentation is the natural next step in the evolution of knowledge-based systems from centralized *logical* reasoning to distributed *dialogical* reasoning, a shift exemplified by the shift from Expert Systems towards Multi-Agent Systems and the

Metaphor	Some informatics methodologies inspired by the metaphor
Biological systems	- Cellular automata, artificial life, genetic algorithms - Ant colony optimisation algorithms, swarm robotics
Economic systems	- Market-based resource allocation (e.g. of network bandwidth) - Trading software for Web-based auctions
Social systems	- Rating and reputation systems (e.g. eBay) - Computational social choice, e-voting
Social cognitive systems	- Agent Communication Languages (e.g. FIPA-ACL) - *Computational models of argumentation*

Semantic Web. Recently, computational argumentation emerged as a promising technology in distributed artificial intelligence [27]. Argumentation can be roughly defined as a disciplined social activity of reason, based on the exchange and manipulation of justification. Proposed and deployed applications of automated argumentation range from coordinating Web services, to supporting medical decision-making, to improving Web-based recommender systems.
Against the above background, my high-level *research goal* is:

> *. . . to take advantage of the growing body of formalized knowledge in distributed computer systems, in order to enable more effective and robust conflict-resolution and agreement.*

To achieve this, I focus on two complementary aspects: (1) the *rules* of interaction (e.g. rules of a debate); and (2) the cognitive/computational *processes* that agents use to decide how to communicate (e.g. a particular argumentation or bidding strategy). Thus my work lies at the intersection of Artificial Intelligence (AI) and cognitive science on one hand, and argumentation theory and game theory on the other. I often benefit greatly from collaborating with economists, psychologists and philosophers.

2 Research Achievements

Within the general scope of argumentation and computation, I have made a number of distinct contributions to date, as discussed in the subsections below.

2.1 Automated Negotiation

Negotiation is a process by which agents attempt to reach agreement over the allocation of resources. Mechanisms for automated negotiation in AI include alternating-offer bargaining [7] and auctions [5]. A number of authors advocated the use of argumentation to enable negotiators to influence each others' mental states (e.g. beliefs, goals, preferences) and to exchange information necessary to find agreement more rapidly (e.g. [11]). I made key contributions to automated argument-based negotiation.

Setting the Agenda for Argumentation-Based Negotiation (ABN): In collaboration with Nick Jennings and others, I authored the first comprehensive survey on ABN [26]. This highly-cited article set the research agenda for the field, and influenced subsequent research by major research groups, including groups at IRIT (Toulouse), Liverpool, Utrecht, Imperial College, Bahia Blanca and others.

Interest-Based Negotiation: In my PhD work, I introduced a number of formal models of interest-based negotiation (IBN), a form of ABN in which agents exchange arguments about their underlying goals [28, 17, 19]. After completing my PhD, I obtained a highly-competitive *Australian Research Council* (ARC) Discovery Grant (2004-2007), in collaboration with Liz Sonenberg (Melbourne) and Frank Dignum (Utrecht). Together with our post-doctoral fellow Philippe Pasquier, we presented the first systematic analysis of when certain classes of IBN protocols increase the likelihood of agreement [24, 25]. We also presented the first empirical study of IBN strategies in an article that won the "*Best Technical Paper Award*" at ICEC [15].

Engineering Methodologies for Automated Negotiation: In collaboration with Nick Jennings, Peter McBurney and Liz Sonenberg, I developed STRATUM, the first agent-oriented software engineering methodology for designing heuristic negotiation strategies (e.g. for the Trading Agent Competition) [29]. I also worked (jointly with Leon Sterling and Thomas Juan) on facilitating requirements analysis that enables selecting the appropriate negotiation protocols for the task at hand [20].

The Behavioral Economics of IBN: It was an open question whether IBN protocols could be beneficial in computer-mediated communication among people. While goal revelation can help people discover cooperative deals, revealing private information can make them vulnerable to exploitation. In collaboration with researchers from MIT/Harvard and Simon Fraser University, we presented the first human study of a computer-mediated IBN protocol using a simulated task negotiation and real monetary rewards [8]. Results revealed that goal revelation has a positive effect on the aggregate performance of negotiators, and on the likelihood of agreement. Analysis also showed that goal revelation was used by independent players to assist dependent players without incurring a loss themselves, thus leading to socially (Pareto) beneficial outcomes.

ABN in a Society: Argumentation rarely happens in isolation of the surrounding social context. Together with collaborators at the University Southampton and Liverpool, we developed a framework for capturing various types of *social arguments* that can be used in negotiation [10]. The simulations we conducted represent the first empirical investigation of ABN beyond simple one-on-one negotiation.

2.2 Supporting Argumentation on the Semantic Web

The World Wide Web is an ideal platform for argumentative expression and communication, due to its ubiquity and openness. Much argumentation takes place on personal blogs, discussion forums, news commentary sites, etc. However, these methods do not capture the structure of argumentative viewpoints explicitly, making the task of searching, evaluating, comparing and relating arguments difficult.

Recently, an increasing number of Web applications provide specific support for large-scale argumentation.[1] When compared with traditional methods of Web discourse, these tools enable better visualization, navigation and analysis of the 'state

[1] E.g. see `www.truthmapping.com`, `www.debategraph.com` and `cohere.open.ac.uk`

of the debate' by participants and, potentially, by automated tools. I am interested in supporting such applications by providing tools for rich argument representation.

The Argument Interchange Format: My interest in the computational representation of argument led to my involvement in the current effort to produce a standard Argument Interchange Format (AIF). The AIF aims to provide a common, extensible ontology of argument-related concepts. I co-organized an AgentLink technical forum which initiated the AIF project and lead to the first AIF specification [3]. Subsequently, I presented the first implementation of the AIF using Semantic Web ontology languages (namely RDF Schema and OWL), jointly with Chris Reed and my MSc student Fouad Zablith. The result was the first Semantic Web-based prototype system for the annotation, querying, and manipulation of complex argument structures [18, 30].

2.3 Game Theory and Argumentation

Argumentation is intrinsically a strategic affair. Different parties often have conflicting goals, which can range from convincing an audience of a particular opinion, to avoiding discussion of particular issues, to simply making the opponent look bad or inconsistent. Parties also have different argumentation strategies that they may use in order to achieve their goals. Such strategies dictate the kind of information they may choose to share, as well as the way in which they stir the dialog.

Despite strategies being very relevant to argumentation, very little work exists on understanding the strategic incentives in argumentation dialogues. This is possibly due to argumentation being a far more complex affair than, say, an auction or a voting protocol. In particular, there is no well-established mathematical foundation for argumentation in game theory [14]. This is a significant obstacle to applying argumentation in open systems (e.g. agents exchanging arguments on the Semantic Web).

Argumentation Mechanism Design: Recently, Kate Larson and I introduced a new perspective on the study of argumentation protocols by defining it as a mechanism design problem [21, 23]. In particular, we presented Argumentation Mechanism Design (ArgMD), in which argument evaluation criteria can be studied in terms of the games they induce on the arguing agents. This contrasts significantly with traditional work, which studies argument evaluation criteria from the perspective of one omniscient agent [6]. The ArgMD approach enabled us to prove the first analytical results pertaining to the strategic properties of abstract argumentation mechanisms. In particular, we identified conditions under which the Dung's grounded semantics [6] induces certain types of agents to reveal their private arguments truthfully.

Argumentation and Social Welfare: In related work, Kate and I initiated a new research program into the *social desirability* of different argument evaluation criteria [22]. In particular, we conducted an extensive analysis of the Pareto optimality of different approaches to evaluating a given set of conflicting arguments. We show that, depending on the preferences of agents involved, certain semantics can sometimes be provably more socially desirable than others.

2.4 Miscellaneous Other Work

Within the broader context of multi-agent systems and their applications, I made a number of other contributions that do not make use of argumentation-based techniques.

Dynamic Taxi Dispatch Policies: In collaboration with Sherief Abdallah and our MSc student Aamena Alshamsi, we presented a system that uses multi-agent self-organisation to improve taxi dispatch [1]. We simulated an existing deployed taxi dispatch system that is operated in the Middle East and primed it with real-world data. This simulator was then used to study the effectiveness of autonomously self-organizing the network topology used by taxi zones in routing customer requests. The results were very promising, showing that autonomous self-organization can significantly outperform the existing deployed taxi dispatcher.

Intentional Learning: Together with Liz Sonenberg and our PhD student Budhitama Subagdja, we presented an architecture for *intentional learning* in Belief-Desire-Intention (BDI) agents [32]. The architecture is an extension of the BDI architecture in which the learning process is explicitly described as plans. Learning plans are meta-level plans which allow the agent to introspectively monitor its mental states and update other plans at run time. This work contributes to the state of the art in agent architectures in two major ways: (1) it bridges the gap between the BDI agent model and the need for a learning capability; (2) it provides a new perspective on how to build an agent that is capable of learning. This work combines the strengths of learning and BDI agent frameworks in a rich language for describing deliberation processes. The approach enables domain experts to specify learning processes and strategies explicitly, while still benefiting from procedural domain knowledge expressed in plan recipes.

Iterative Task Allocation: In collaboration with Christian Guttmann and Michael Georgeff, we defined the 'collective iterative allocation' problem, in which a group of agents endeavors to find the best allocations (of agents to tasks) through refinements of these allocations over time [9]. The paper presented a number of theoretical and empirical results on when different voting rules can converge to optimal allocations.

3 Future Research Agenda

Within the general context of *argument & computation*, my current research agenda is to work towards a general theory of strategic argumentation, and its applications. This agenda focuses on two main streams: (1) designing effective argumentation in open systems; and (2) engineering argumentation that takes into account bounded-rationality. I discuss these two streams in more detail in the following sub-sections.

3.1 Argumentation in Open Systems

An open system (e.g. the Web) allows computers unrestricted access without guarantees that they conform to centralized control. Such computational entities may have *incentive* to act *strategically* (i.e. in a self-interested manner, taking into account the actions of other agents). For example, they may deviate from the protocol or provide false information to manipulate the outcome to their own advantage. This brought foreword a new major challenge in designing computer systems, namely ensuring good overall behavior despite the fact that individual agents are strategic.

Indeed, many areas in distributed computing have benefited greatly from the study of incentives and strategies, typically using the formal tools of game theory and mechanism design[2] [14]. For example, game-theoretic techniques have been used to analyze 'selfish routing' over Internet protocols [14, Chapter 18], understand strategic bidding in Web-based auctions [14, Chapter 11], and design manipulation-resistant online reputation and feedback mechanisms [14, Chapter 27].

However, most research to-date focused on relatively simple protocols, in which strategies amount to relatively simple choices, such as: Which link to route a packet to? What price to bid in an auction, and when? What rating (from 1 to 5) to give a seller on eBay? In richer settings (e.g. a Semantic Web environment [2]), a strategy amounts to exchanging high-level knowledge that interacts in a more complex manner with knowledge presented by other agents. These settings call for an understanding of strategic behavior in the context of knowledge exchange, or argumentation.

Existing frameworks for automated or computer-mediated argumentation are severely underdeveloped when it comes to the analysis of incentives and strategies. This is mainly due to the lack of a theoretical foundation of argumentation in game theory. For argumentation to be widely used in distributed open systems, the protocols need to be robust against strategic manipulation by engineering the right incentives.

Against the above background, in this line of work, I have the following broad aim: *to build a game-theoretic foundation for argumentation-based communication in distributed computer systems*. Specifically, I aim to achieve two objectives: **[Objective 1:]** to characterize the possible incentives in agent argumentation, and to understand how these incentives influence agents' argumentation strategies; **[Objective 2:]** to build on the understanding of incentives in order to engineer improved protocols that enforce desirable outcomes (e.g. more effective or more truth-enforcing). This work will build on the recent promising results I achieved with my collaborators (see Section 2.3).

I plan to apply the knowledge gained from this work by developing a self-enforcing Web-based system for social argumentation. This will consolidate my work on strategies with my work on argument representation on the Semantic Web (see Section 2.2).

3.2 Arguing with Bounded-Rational Agents

I am interested in the cognitive/computational aspects of argumentative communication (cf. pragmatics). In particular, I am interested in the interplay between *normative* behavior (e.g. optimal game-theoretic strategies) and *empirical* observation of the way humans and software agents may argue with one another using heuristics.

In the context of automated negotiation, it has recently been demonstrated that theoretically optimal software agents do not necessarily perform well when negotiating when bounded-rational agents, such as humans [12]. As it turns out, well-designed heuristics, which take into account the specific biases of human judgment, can significantly outperform theoretically-optimal strategies. I am interested in bringing these

[2]Game theory is the mathematical study of strategic behaviour. Mechanism design is a sub-discipline which aims at designing interaction rules that enforce good overall outcomes even when agents are strategic.

kinds of insights into the study of automated argumentation, and in applying them to build systems that argue with, or facilitate argumentation among humans.

3.3 Mass-Persuasion through Social Incentives

Recently, I have come to the realization that much of our opinion and behavior is driven by subtle social processes, such as social influence and peer pressure, as opposed to rational deliberative reasoning. Much recent research shows that social processes affect everything from our health [4], to the music we like [31].

With this realization, it becomes important to take into account social processes when designing persuasive technologies. Indeed, this idea has been the driving force behind phenomena such as viral marketing [33]. Moreover, social incentives have shown great success in mobilizing society towards a common goal, as was demonstrated in the DARPA Network Challenge winning strategy [16].

I am currently working with Alex (Sandy) Pentland's research group at the MIT Media Lab on designing social incentive mechanisms for influencing the collective behavior of groups. Preliminary theoretical analysis shows that, under certain conditions, we can leverage the power of existing social ties in order to reduce free-riding in public good games [13]. Experimental work is underway to verify this hypothesis.

References

[1] A. Alshamsi, S. Abdallah, and I. Rahwan. Multiagent self-organization for a taxi dispatch system. In *AAMAS '09: Proceedings of the 8th international joint conference on Autonomous agents and multiagent systems*, 2009.

[2] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, pages 29–37, May 2001.

[3] C. I. Chesñevar, J. McGinnis, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, and S. Willmott. Towards an argument interchange format. *The Knowledge Engineering Review*, 21(4):293–316, 2006.

[4] N. Christakis and J. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370, 2007.

[5] P. Cramton, Y. Shoham, and R. Steinberg, editors. *Combinatorial Auctions*. MIT Press, 2006.

[6] 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):321–358, 1995.

[7] P. Faratin, C. Sierra, and N. R. Jennings. Using similarity criteria to make tradeoffs in automated negotiations. *Artificial Intelligence*, 142(2):205–237, 2002.

[8] Y. Gal, S. D'souza, P. Pasquier, I. Rahwan, and S. Abdallah. The effects of goal revelation on computer-mediated negotiation. In *Proceedings of the Annual meeting of the Cognitive Science Society*, Amsterdam, The Netherlands, 2009.

[9] C. Guttmann, M. Georgeff, and I. Rahwan. Collective iterative allocation: Enabling fast and optimal group decision making. *Web Intelligence and Agent Systems*, 8(1):(to appear, accepted Dec 2008), 2010.

[10] N. C. Karunatillake, N. R. Jennings, I. Rahwan, and P. McBurney. Dialogue games that agents play within a society. *Artificial Intelligence*, 173(9-10):935–981, 2009.

[11] S. Kraus, K. Sycara, and A. Evenchik. Reaching agreements through argumentation: A logical model and implementation. *Artificial Intelligence*, 104(1–2):1–69, 1998.

[12] R. Lin, S. Kraus, J. Wilkenfeld, and J. Barry. Negotiating with bounded rational agents in environments with incomplete information using an automated agent. *Artificial Intelligence*, 172(6-7):823–851, 2008.

[13] A. Mani, I. Rahwan, and A. Pentland. Localizing externalities in social networks: Inducing peer pressure to enforce socially efficient outcomes. In *Proceedings of the Workshop on Information in Networks (WIN), New York University*. 2010.

[14] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, New York, USA, 2007.

[15] P. Pasquier, R. Hollands, F. Dignum, I. Rahwan, and L. Sonenberg. An empirical study of interest-based negotiation. In M. L. Gini, R. J. Kauffman, D. Sarppo, C. Dellarocas, and F. Dignum, editors, *Proceedings of the Ninth International Conference on Electronic Commerce (ICEC)*, pages 339–348, New York NY, USA, 2007. ACM Press.

[16] G. Pickard, I. Rahwan, W. Pan, M. Cebrian, R. Crane, A. Madan, and A. Pentland. Time Critical Social Mobilization: The DARPA Network Challenge Winning Strategy. *Arxiv preprint arXiv:1008.3172*, 2010.

[17] I. Rahwan. *Interest-based Negotiation in Multi-Agent Systems*. PhD thesis, Department of Information Systems, University of Melbourne, Australia, 2004.

[18] I. Rahwan. Mass argumentation and the Semantic Web. *Journal of Web Semantics*, 6(1):29–37, 2008.

[19] I. Rahwan and L. Amgoud. An Argumentation-based Approach for Practical Reasoning . In G. Weiss and P. Stone, editors, *5th International Joint Conference on Autonomous Agents & Multi Agent Systems, AAMAS'2006, Hakodate, Japan*, pages 347–354, New York, USA, 2006. ACM Press.

[20] I. Rahwan, T. Juan, and L. Sterling. Integrating social modelling and agent interaction through goal-oriented analysis. *Computer Systems Science & Engineering*, 21(2), 2006.

[21] I. Rahwan and K. Larson. Mechanism design for abstract argumentation. In L. Padgham, D. Parkes, J. Mueller, and S. Parsons, editors, *7th International Joint Conference on Autonomous Agents & Multi Agent Systems, AAMAS'2008, Estoril, Portugal*, pages 1031–1038, 2008.

[22] I. Rahwan and K. Larson. Pareto optimality in abstract argumentation. In D. Fox and C. Gomes, editors, *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI-2008)*, Menlo Park CA, USA, 2008.

[23] I. Rahwan, K. Larson, and F. Tohmé. A characterisation of strategy-proofness for grounded argumentation semantics. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 251–256, 2009.

[24] I. Rahwan, P. Pasquier, L. Sonenberg, and F. Dignum. On the benefits of exploiting underlying goals in argument-based negotiation. In R. C. Holte and A. Howe, editors, *Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI-2007)*, pages 116–121, Menlo Park CA, USA, 2007.

[25] I. Rahwan, P. Pasquier, L. Sonenberg, and F. Dignum. A formal analysis of interest-based negotiation. *Annals of Mathematics and Artificial Intelligence*, page (accepted), 2009.

[26] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation based negotiation. *Knowledge Engineering Review*, 18(4):343–375, 2003.

[27] I. Rahwan and G. R. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, 2009.

[28] I. Rahwan, L. Sonenberg, and F. Dignum. Towards interest-based negotiation. In J. Rosenschein, T. Sandholm, M. J. Wooldridge, and M. Yokoo, editors, *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2003)*, pages 773–780. ACM Press, 2003.

[29] I. Rahwan, L. Sonenberg, N. R. Jennings, and P. McBurney. STRATUM: A methodology for designing heuristic agent negotiation strategies. *Applied Artificial Intelligence*, 21(6):489–527, 2007.

[30] I. Rahwan, F. Zablith, and C. Reed. Laying the foundations for a world wide argument web. *Artificial Intelligence*, 171(10–15):897–921, 2007.

[31] M. Salganik, P. Dodds, and D. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854, 2006.

[32] B. Subagdja, I. Rahwan, and L. Sonenberg. Intentional learning agent architecture. *Autonomous Agents and Multiagent Systems*, 18(3):417–0470, 2009.

[33] D. J. Watts and J. Peretti. Viral marketing in the real world. *Harvard Business Review*, May 2007.