A Hundred Schools of Thought Automatically Contending

Peter McBurney and Simon Parsons

Department of Computer Science
University of Liverpool
Liverpool L69 7ZF UK
p.j.mcburney@csc.liv.ac.uk

Department of Computer and Information Science
Brooklyn College
City University of New York
Brooklyn NY 11210 USA
parsons@sci.brooklyn.cuny.edu

We envision a rational cacophony of "numberless infinities" of autonomous software agents engaged in dialogue and argument with each other, debating automatically every topic which humans now discuss and many we do not. This note presents the major elements we believe are necessary for the achievement of this vision, and outlines the progress already made — by ourselves and others — towards its realization. We begin with a discussion as to why we believe argumentation is essential to modeling of autonomous agent interactions.

Much research to date in interaction between autonomous software agents has focused on negotiation, and has used economic auction and game-theoretic mechanisms as the enabling technology (Jennings *et al.* 2001). Perhaps this is understandable given the recent attention paid to electronic commerce and the computational and theoretical challenges involved even in designing simple auction mechanisms. Yet, most everyday human interactions do not involve negotiation; indeed, even for those which do, the participants typically spend more time engaged in relevant dialogue before, during and subsequent to the transaction than in actually exchanging offers with each other. Participants to a negotiation even engage in relevant dialogue with others not party to the potential transaction, a fact which has led marketing theorists to study what they call *word-of-mouth* influences on consumer decision-making, e.g., Czepiel 1974.

It is interesting to ask why this is so: why do human parties to a negotiation exchange more than the simple acceptances and rejections permitted in most auction mechanisms? A key reason is that human participants rarely, if ever, satisfy one of the central tenets of the classical economic models of decision-making (Lindley 1985): they do not usually begin a decision process with pre-determined preferences and utility valuations, but instead these are formed (either partially or completely) in the very process of undertaking the transaction (Searle 2001). In addition, even those participants with predetermined beliefs and preferences are rarely perceived by other participants to hold their views fixedly, thus leading to attempts by those others to persuade them to change their views. These acts of attempted persuasion occur because one party believes the other able and willing to change his or her mind, and that, with such a change, there is a greater chance of successful resolution of the issue being discussed.

We may assume that rational participants to an interaction would not change their minds without receipt of additional information, that is, that they would not change their beliefs, preferences or intentions on the basis merely of whim or malice. With this assumption, the transmission of relevant information between the participants becomes a key determinant of the successful resolution of an interaction, and hence, information-transmission needs to be enabled by the interaction protocol used by the participants. A weakness of economic auction mechanisms as interaction protocols is that they typically permit only limited information to be transmitted between the parties.[1] Negotiation protocols based on argumentation, however, enable the transmission of considerably more information between the participants, and in a manner which is rational and coherent.

These ideas have greater importance if we extend the remit of interaction beyond negotiation, viewing the latter (following Walton and Krabbe 1995) as dialogue concerning the division of some scarce resource. Agents (or humans) engaged in joint deliberation to decide a course of action, for example, may each have different knowledge, none of which is complete but all of which is germane to the problem; achievement of a common joint intention may only be possible by coherent sharing of this distributed knowledge, as is the case for the room-furnishing agents of Parsons *et al.* 1998. Because theories of argumentation apply to a wider class of discourses than negotiation dialogues, they are capable of supporting these other types of interaction; auction mechanisms, in contrast, appear limited to negotiations.

To achieve our vision of a cyber-world of multitudes of automatically arguing agents, a number of distinct components will have to be created and assembled. The first of these will be languages for argument and dialogue, which specify locutions and define syntactical rules for their use. The development of standardised general agent communications languages has been the focus of considerable effort in computer science; see Labrou *et al.* 1999 for a review. We have instead approached the task by defining application-specific protocols for particular types of interactions, protocols which may be seen as examples of electronic *institutions* (Noriega & Sierra 1997). We have done this using formal dialogue games, taken from philosophy, in which two or more players "move" by uttering locutions according to some defined rules. Dating at least from the time of Aristotle, dialogue games have been used by contemporary philosophers to study fallacious modes of reasoning (Hamblin 1970) and as a game-theoretic emantics for intuitionistic and classical logic (Lorenzen & Lorenz 1978), and quantum logic (Mittelstaedt 1979). Within Artificial Intelligence they have been applied to: modeling complex reasoning, for instance, in legal domains (Bench-Capon *et al.* 2000); human-computer interaction (Bench-Capon *et al.* 1991); and for the design of computational protocols for autonomous agent interactions. Examples of dialogue-game agent protocols include systems for: team-formation (Dignum *et al.* 2000); information-seeking dialogues, where one agent seeks the answer to a question from others believed to know it (Hulstijn 2000); persuasion dialogues, where one agent seeks to persuade another to endorse some proposition or plan of action (Dignum *et al.* 2001); negotiation dialogues, where agents discuss how to divide some scarce resource (Amgoud *et al.* 2000); deliberation dialogues, where agents collaborate to decide a course of action in some situation (Hitchcock *et al.* 2001); and inquiry dialogues, where agents jointly seek the answer to some unknown question (McBurney & Parsons 2001b).

[1] Participants to negotiations may exercise considerable creativity in overcoming the limited capabilities for information transmission of an auction mechanism, as was seen in the possibly collusive behaviour of bidders in the US Federal Communications Commision's PCS spectrum auctions of 1994—1995 (Cramton & Schwarz 2000).

A second requirement for our vision will be mechanisms "in the heads" of participating agents for interpreting locutions uttered in a dialogue and choosing appropriate responses, both single locutions and strategically. For negotiation dialogues, we have proposed mechanisms based on models adopted from the marketing theory of consumer decision-making (McBurney *et al.* 2003). Similar generative mechanisms have been proposed in recent agent negotiation architectures which do not use argumentation, e.g. Faratin 2000. An important question is the extent to which such semantic mechanisms can ever be fully verified, since a sufficiently-clever agent will always be able to simulate insincerely any semantic requirement (Wooldridge 2000). Recent work has proposed the use of a social semantics (Singh 2000), effectively public statements of private mental states such as beliefs and intentions, to ensure all participants assign the same meaning to syntactical statements. Of course, duplicitous agents will still be able to make false declarations of such social commitments.

A third requirement will be for mechanisms to decide to initiate or to enter dialogues of particular types on particular subjects at particular times. For this task, research in computational linguistics linking dialogue locutions with the beliefs and intentions of participants is likely to be relevant. For example, the model of human dialogues of Grosz & Sidner 1986 identifies a purpose for the discourse and purposes for subsidiary segments of the discourse. As the dialogue proceeds, these purposes are jointly recognized by the participants (or not) and then achieved by means of discussion (or not), thus leading to closure or to subsequent dialogue segments and dialogues.

Fourthly, we will require formal rules of encounter and procedure, the argumentation equivalent of auction mechanisms. Examples of such systems of rules include: Robert's rules of order for public meetings (Robert 1986); Alexy's rules for discourses over ethical and moral questions (Alexy 1978/1990); and Hitchcock's principles for rational mutual inquiry (Hitchcock 1991). Because, as we argued above, the value of argumentation in agent interactions lies essentially in the ability and willingness of participants to change their beliefs, desires and intentions, then argumentation-based systems must permit participants to undergo self-transformation (Forrester 1999); i.e. agents must be permitted to change their mental states. But to do this rationally and coherently requires rules for making assertions, for questioning and contesting assertions, and for justifying and retracting prior assertions. Thus, a system of dialogue between rational agents requires an explicit theory of argumentation and rules of encounter. Some work has been undertaken to develop computational versions of such rules, e.g. Prakken & Gordon 1999, but the only dialogue-game agent communications protocol known to us which has been shown formally to comply with explicit rules of encounter is our own protocol for joint reasoning in scientific domains (McBurney & Parsons 2001b).

A fifth requirement for our vision are coherent models of complex dialogues, such as dialogues undertaken in sequence and in parallel, and embeddings of one dialogue inside another. An essential step towards this was the partial typology of six basic types of human dialogues developed by Walton and Krabbe (1995), which categorizes dialogues according to the dialogue purpose, the initial information known to each participant and their respective intentions. Building on this typology, Reed (1998) proposed a computational formalism to represent complex combinations of the basic types. A second computational model for complex dialogues is our dynamic modal formalism (McBurney & Parsons 2002). Before such formalisms can be applied, a better understanding of their properties, such as their computational complexity, will be required.

A final requirement is a formal theory of dialogue protocols within which to compare one protocol with another. Currently, potential users of on-line auction systems are faced with choosing between auctions conducted under different rules and procedures. The same problem will confront intending users of dialogue protocols. When are two dialogue protocols the same? How similar are two different protocols? When is one protocol better than another? In what sense, and for what purpose? These questions will be increasingly important as agent system proliferate and more dialogue protocols are developed. In order to answer such questions, we have begun to develop a mathematical representation – what computer scientists call a *denotational semantics* – for dialogues and dialogue protocols (McBurney & Parsons 2004b).

Given a formal theory of dialogue-game agent interaction protocols, we should be able to use this to articulate a set of evaluation criteria for protocols, providing guidance to both designers and potential users. We have presented first lists of evaluation criteria (McBurney, Parsons & Wooldridge 2002, Rehg *et al.* 2004), drawing on our protocol design experience, and informed by criteria proposed for the design and assessment of automated auction mechanisms. However, the economic problems for which auction mechanisms have been designed can usually assume the existence of a common unit of inter-personal utility comparison, namely money. Consequently, the criteria proposed for assessment of auction mechanisms have emphasized the assessment of outcomes rather than the assessment of procedures used to reach those outcomes. In contrast, many decisions, especially those in public policy domains, have possible outcomes which are not comparable, as comparison requires inter-personal utility assessments. For example, to assess whether banning cigarette smoking would be a desirable policy would require comparison of the disutilities to one set of individuals, current and future smokers, with the utilities to others, non-smokers; to add to the comparison difficulty, some of these individuals are as yet unborn. That many decisions result in incommensurable outcomes leads us to believe that we also require formal theories of *decision-process rationality* to complement the *decision-outcome rationality* defined by maximum-expected-utility in classical decision theory (Lindley 1985). Recent research in argumentation (e.g., Wohlrapp 1998), in political science (e.g., Renn *et al.* 1995) and in economics (e.g., Sen 2002), may lead to a theory of rational decision-processes, and may guide the design of dialogue protocols for public policy domains, as in McBurney & Parsons 2001a.

As is clear from this list of requirements, the realization of our vision will draw on contributions from argumentation, computer science, decision theory, economics, linguistics, marketing theory, philosophy, political science, pure mathematics, rhetorical theory and sociology. While achievement of our vision presents significant theoretical and practical challenges, the fact that so many areas of human thought are involved is why we find this research area so exciting.

References

R. Alexy (1978/1990). A theory of practical discourse. In S. Benhabib and F. Dallmayr, editors, *The Communicative Ethics Controversy*, pages 151–190. MIT Press, Cambridge, MA, USA. (Published in German 1978).

L. Amgoud, N. Maudet, and S. Parsons (2000). Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings 4th International Conference on Multi-Agent Systems (ICMAS-2000)*, pages 31–38, Boston, MA, USA, 2000. IEEE.

T. J. M. Bench-Capon, P. E. Dunne, and P. H. Leng (1991). Interacting with knowledge-based systems through dialogue games. In *Proceedings 11th International Conference on Expert Systems and Applications*, pages 123–140, Avignon, France.

T. J. M. Bench-Capon, T. Geldard, and P. H. Leng (2000). A method for the computational modelling of dialectical argument with dialogue games. *Artificial Intelligence and Law*, 8:233–254.

P. Cramton and J. A. Schwarz (2000). Collusive bidding: Lessons from the FCC Spectrum Auctions. *Journal of Regulatory Economics*, 17:229–252.

J. A. Czepiel (1974). Word of mouth processes in the diffusion of a major technological innovation. *Journal of Marketing Research*, 11:172–180.

F. Dignum, B. Dunin-Keplicz, and R. Verbrugge (2000). Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lesperance, editors, *Proceedings 7th International Workshop on Agent Theories, Architectures, and Languages (ATAL-2000)*, pages 141–156, Boston.

F. Dignum, B. Dunin-Keplicz, and R. Verbrugge (2001). Creating collective intention through dialogue. *Logic Journal of the IGPL*, 9(2):305–319.

P. Faratin (2000). *Automated Service Negotiation Between Autonomous Computational Agents*. PhD thesis, Queen Mary and Westfield College, University of London, UK.

J. Forester (1999). *The Deliberative Practitioner: Encouraging Participatory Planning Processes*. MIT Press, Cambridge, MA, USA.

B. J. Grosz and C. L. Sidner (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3): 175–204.

C. L. Hamblin (1970). *Fallacies*. Methuen, London, UK.

D. Hitchcock (1991). Some principles of rational mutual inquiry. In F. van Eemeren *et al.*, editor, *Proceedings 2nd International Conference on Argumentation (ISSA-91)*, pages 236–243, Amsterdam. SICSAT.

D. Hitchcock, P. McBurney, and S. Parsons (2001). A framework for deliberation dialogues. In H. V. Hansen, C. W. Tindale, J. A. Blair, and R. H. Johnson, editors, *Proceedings 4th Biennial Conference of the Ontario Society for the Study of Argumentation (OSSA-2001)*, Windsor, Ontario, Canada.

J. Hulstijn (2000). *Dialogue Models for Inquiry and Transaction*. PhD thesis, Universiteit Twente, Enschede, The Netherlands.

N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, M. Wooldridge, and C. Sierra (2001). Automated negotiation: prospects, methods and challenges. *Group Decision and Negotiation*, 10(2): 199–215

Y. Labrou, T. Finin and Y. Peng (1999): Agent communication languages: the current landscape. *IEEE Intelligent Systems*, 14 (2): 45—52.

D. V. Lindley (1985). *Making Decisions*. John Wiley and Sons, London, UK, second edition.

P. Lorenzen and K. Lorenz (1978). *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt, Germany.

P. McBurney, R. M. van Eijk, S. Parsons, and L. Amgoud (2003). A dialogue-game protocol for agent purchase negotiations. *Journal of Autonomous Agents and Multi-Agent Systems*, 7 (3): 235—273.

P. McBurney and S. Parsons (2001a). Intelligent systems to support deliberative democracy in environmental regulation. *Information and Communications Technology Law*, 10(1): 33–43.

P. McBurney and S. Parsons (2001b). Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and Artificial Intelligence*, 32(1—4): 125—169.

P. McBurney and S. Parsons (2002). Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11 (3): 315—334.

P. McBurney and S. Parsons [2004a]: Dialogue games in multi-agent systems. *Informal Logic*. *(In press)*

P. McBurney and S. Parsons (2004b). A denotational semantics for deliberation dialogues. In: N. R. Jennings, C. Sierra, E. Sonenberg and M. Tambe (Editors): *Proceedings Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2004)*. New York City, NY, USA. *(In press)*

P. McBurney, S. Parsons and M. Wooldridge (2002). Desiderata for agent argumentation protocols. In: C. Castelfranchi and W. L. Johnson (Editors): *Proceedings First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, Bologna, Italy, pages 402—409. New York City, NY, USA: ACM Press.

P. Mittelstaedt (1979). *Quantum Logic*. D. Reidel, Dordrecht, The Netherlands.

P. Noriega and C. Sierra (1997). Towards layered dialogical agents. In J. P. Muller, M. J. Wooldridge, and N. R. Jennings, editors, *Intelligent Agents III: Proceedings 3rd Workshop on Agent Theories, Architectures, and Languages*, LNAI 1193, pages 173–188, Berlin, Germany. Springer.

S. Parsons, C. Sierra, and N. R. Jennings (1998). Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3): 261–292.

H. Prakken and T. F. Gordon (1999). Rules of order for electronic group decision making — a formalization methodology. In J. A. Padget (Editor), *Collaboration between Human and Artificial Societies: Coordination and Agent-Based Distributed Computing*, LNAI 1624, Berlin, Germany. Springer.

C. Reed (1998). Dialogue frames in agent communications. In Y. Demazeau, editor, *Proceedings 3rd International Conference on Multi-Agent Systems (ICMAS-98)*, pages 246–253. IEEE Press.

W. Rehg, P. McBurney, and S. Parsons (2004). Computer decision-support systems for public argumentation: criteria for assessment. *AI and Society* (in press).

O. Renn, T. Webler, and P. Wiedemann (Editors) (1995). *Fairness and Competence in Citizen Participation: Evaluating Models for Environmental Discourse*. Kluwer, Dordrecht, The Netherlands.

H. M. Robert (1986). *Robert's Rules of Order. The Standard Guide to Parliamentary Procedure*. Bantam Books, New York, NY, USA.

J. Searle (2001): *Rationality in Action.* Cambridge, MA, USA: MIT Press.

A. Sen (2002): *Rationality and Freedom*. Cambridge, MA, USA: Harvard University Press.

M. P. Singh (2000). A social semantics for agent communications languages. In F. Dignum, B. Chaib-draa, and H. Weigand (Editors), *Workshop on Agent Communication Languages, IJCAI-99*, Berlin, Germany. Springer.

D. N. Walton and E. C. W. Krabbe (1995). *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Albany, NY, USA. State University of New York Press.

H. Wohlrapp (1998). A new light on non-deductive argumentation schemes. *Argumentation*, 12: 341–350.

M. Wooldridge (2000). Semantic issues in the verification of agent communication languages. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(1): 9–31.