When Will Negotiation Agents Be Able to Represent Us? The Challenges and Opportunities for Autonomous Negotiators

Tim Baarslag and **Michael Kaisers**
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
{T.Baarslag, M.Kaisers}@cwi.nl

Enrico H. Gerding
University of Southampton
Southampton, United Kingdom
eg@soton.ac.uk

Catholijn M. Jonker
Delft University of Technology
Delft, The Netherlands
C.M.Jonker@tudelft.nl

Jonathan Gratch
University of Southern California
Los Angeles, CA, United States
gratch@ict.usc.edu

Abstract

Computers that negotiate on our behalf hold great promise for the future and will even become indispensable in emerging application domains such as the smart grid and the Internet of Things. Much research has thus been expended to create agents that are able to negotiate in an abundance of circumstances. However, up until now, truly autonomous negotiators have rarely been deployed in real-world applications. This paper sizes up current negotiating agents and explores a number of technological, societal and ethical challenges that autonomous negotiation systems have brought about. The questions we address are: in what sense are these systems autonomous, what has been holding back their further proliferation, and is their spread something we should encourage? We relate the automated negotiation research agenda to dimensions of autonomy and distill three major themes that we believe will propel autonomous negotiation forward: accurate representation, long-term perspective, and user trust. We argue these orthogonal research directions need to be aligned and advanced in unison to sustain tangible progress in the field.

1 Introduction

Negotiation, the process of joint decision making, is pervasive in our society. Whenever actors meet and influence each other to forge a mutually beneficial agreement, a form of negotiation is at work [Young, 1991].

Negotiation arises in almost every social and organizational setting, yet many avoid it out of fear or lack of skill and this contributes to income inequality, political gridlock and social injustice [Eisenberg and Lanvers, 2009]. This has led to an increasing focus on the design of autonomous negotiators capable of automatically and independently negotiating with others.

Automated negotiation research is fueled by a number of benefits that computerized negotiation can offer, including better (win-win) deals, and reduction in time, costs, stress and cognitive effort on the part of the user. Moreover, autonomous negotiation will soon become not just desired but *required* in instances where the human scale is simply too slow and expensive. For instance, with the world-wide deployment of the smart electrical grid and the must for renewable energy sources, flexible devices in our household will soon (re-)negotiate complex energy contracts automatically. Another example is the rise of the Internet of Things (IoT), which will introduce countless smart, interconnected devices that autonomously negotiate the usage of sensitive data and make trade-offs between privacy concerns, price, and convenience.

To properly fulfill its representational role in an ever-dynamic environment, a negotiation agent has to balance and adhere to different aspects of autonomous behavior, including self-reliance and the capability and freedom to perform its actions, while at the same time remaining interdependent in its joint activity with the user. While many successes have been achieved in advancing various degrees of autonomy in negotiating agents, it is readily apparent that fully-deployed and truly autonomous negotiators are still a thing of the future. Continued development will be required before agents will be able to forge even mundane agreements such as the personalized renewal of our energy or mobile phone contracts. This begs the obvious question: what is still lacking currently and what is needed for autonomous negotiators to be able to fulfill their promise?

This paper discusses the challenges and upcoming application domains for (almost) entirely autonomous negotiation on people's behalf. We describe the technological challenges associated with these future domains and provide a roadmap towards full autonomy, together with stops along the way, highlighting what we deem important solution concepts for enabling future autonomous negotiation systems. As a basis for our discussion, we provide a unifying view of autonomous negotiation based on three orthogonal dimensions of autonomy that research has focused on so far: being self-sufficient, self-directed, and interdependent. We argue that automated negotiation opportunities of tomorrow are calling for a com-

bined effort in addressing these three pillars of a negotiator's autonomy.

This paper does not aim to survey all research or challenges in the field comprehensively, but rather presents pointers to what we consider important focal points for autonomous negotiation, now and in the future. We pinpoint and elaborate on the following major challenges for autonomous negotiation:

1. Domain knowledge and preference elicitation;
2. Long-term perspective; and
3. User trust and adoption.

Lastly, this paper also pays homage to the 2001 landmark publication by Jennings *et al.* and asks what has happened, 16 years later, with the prospects and challenges of automated negotiation. We examine which main challenges have been addressed, and which stay relevant in a world that offers more opportunities for automated negotiation than ever before.

2 The Autonomy Diagonal of Negotiation

Autonomous negotiation is more than just *automated* negotiation; it is the freedom to negotiate independently. Rather than being uni-dimensional, autonomy incorporates at least two components [Bradshaw *et al.*, 2003]: *self-sufficiency* (the capability of the actor to take care of itself) and *self-directedness* (the freedom to act within the environment and the means to reach goals). Following [Johnson *et al.*, 2011] we distinguish a third dimension called *support for interdependence* – being able to work with others and influence and be influenced by team members.[1]

We can distinguish three strands of research in automated negotiation that each cluster around one of the three dimensions of autonomy (Fig. 1):

Negotiation support systems.
These systems are designed to assist and train people in negotiation. Some of these systems, such as the Inspire system [Kersten and Lo, 2001], have been widely employed in real-life. However, while negotiation support systems enable interdependence by design, humans predominately supervise and make decisions on the appropriate outcome, which results in low self-sufficiency and self-directedness.

Game theoretical approaches and trading bots.
Game theory's dominant concern is with fully rational players and what each should optimally do. This approach is therefore called *symmetrically prescriptive* [Sebenius, 1992]. The focus is on either equilibrium strategies or protocols that can guarantee a good outcome for both players through mechanism design [Young, 1991]. Agents have a reduced scope for self-directedness in such settings, as they are relatively simple and need to conform to certain strategies (e.g. to bid truthfully in an auction). Similarly,

[1]Note that the notion of autonomy is notoriously difficult to capture (see [Johnson *et al.*, 2011] for an overview). We are concerned here with those aspects especially relevant for negotiation and for their autonomy in relation to their environment; an alternative, more self-contained definition, for example, is an agent's ability to generate its own goals [Luck *et al.*, 2003].

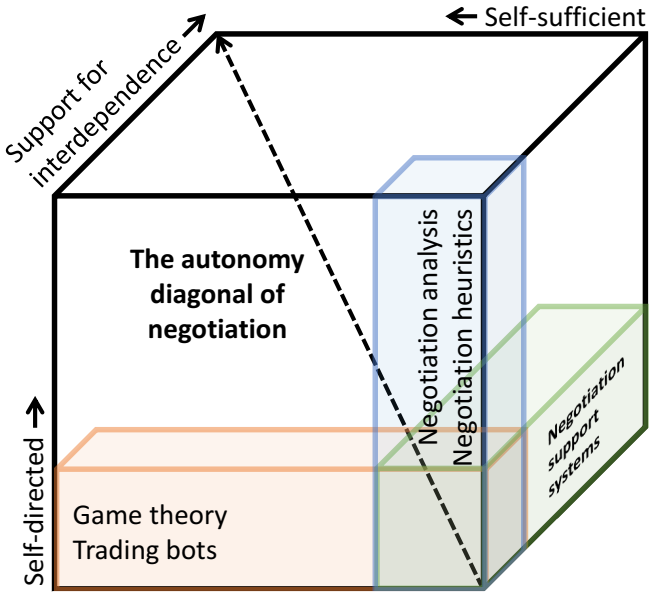

Figure 1: By and large, negotiation research can be clustered around one of the three main orthogonal dimensions of autonomy: *self-sufficiency*, *self-directedness*, and *interdependence*. The efforts of the three need to be integrated to arrive at truly autonomous negotiators that can progress along the *autonomy diagonal*.

real-world trading bots mostly employ simple rule-based functions which have been hard-coded in advance. Examples of this type are among the most advanced autonomous negotiators in terms of self-sufficiency, such as high frequency trading agents for financial exchanges, advertising exchanges, or sniping agents used in eBay [Hu and Bolivar, 2008]. While these approaches are able to function without human intervention and can be highly self-sufficient, they are constrained in terms of freedom to direct the process.

Negotiation analytical approaches.
Negotiation analysis *prescribes* how players should act given a *description* of how others will act. That is, this field is concerned with an *asymmetrical prescriptive/descriptive* view of autonomous negotiation [Sebenius, 1992]. Much research on what are often dubbed simply 'negotiation agents' (or 'heuristics' in game theory literature) falls into this category; e.g. all negotiation agents from the annual automated negotiation competition [Baarslag *et al.*, 2015]. A key feature of this approach is the agent's ability to make judgment calls without intervention (i.e. to construct beliefs based on partial information and act in best response to this belief, typically over opponent types or strategies), while the agent's preferences are often considered externally given. This locates the negotiation analytical approach around the self-directed axis.

As can be gleaned from the fields indicated above, autonomous negotiation has garnered attention from different research directions and has managed to advance in key aspects of autonomous behavior. As a result, we now have negotiators that exist independently in the real world, delegated

Major challenge	Main autonomy dimensions	Building blocks	Example opportunities	Solutions roadmap
Domain knowledge and preference elicitation (Section 3.1)	Self-sufficiency & Interdependence	Preference elicitation on-the-fly	Privacy and IoT	Value of information indicators, robust performance estimates
		Domain modeling	Smart grids	Separate user/agent domain models, expert mappings
Long-term perspective (Section 3.2)	Self-sufficiency & Self-directedness	Repeated interactions	Communities, smart homes, autonomous driving	Temporally integrative negotiations, reputation metrics
		Non-stationary preferences	B2B, entertainment booking	Cost-efficient tracking, context-dependent models, preference dynamics
User trust and adoption (Section 3.3)	Self-directedness & Interdependence	Acceptability and participation	Conflict resolution, customer retainment	Co-creation, adjustable autonomy, transfer of control
		Transparent consequences	Sharing economy, decentralized marketplaces	Transparency and openness, worst-case bounds, risk measures

Table 1: Overview of major challenges in autonomous negotiation and the main dimensions of autonomy to which they relate. Each challenge is subdivided in building blocks along with example opportunities and a solution roadmap.

with a gamut of available strategies to freely choose among, and that can engage in supportive interdependence; *just not all at the same time*.

This may explain why it has proven difficult to extend the progress made in this field to truly representative negotiating agents. Of course we acknowledge that to a lesser degree, combined work on all dimensions has been performed (as depicted by the three-colored cube in Figure 1); we simply argue that the main automated negotiation research lines have developed in parallel to one of the three autonomy directions. Research-wise, it is unquestionably a sound strategy to first explore the autonomy axes in separation. As Figure 1 suggests, we can make substantive progress in autonomous negotiation by continuing to advance along *the autonomy diagonal*, which has inspired the focal points of the challenges we present in the next section.

3 Major Challenges

The various aspects of autonomy drive three major open challenges for autonomous negotiation, of which the overall theme can be summarized as *trusted and sustained representation*. We describe the challenges and their building blocks below, together with a number of explicit opportunities in each case (see Table 1 for an overview).

Just like autonomy itself, each challenge outlined here is *multi-dimensional*; i.e., each challenge pertains to at least two dimensions of autonomy, thereby providing the impetus to further advance along the autonomy diagonal. Note that many of these challenges intersect and cannot be entirely untangled; for example, adequate user preference extraction will not only increase the user model accuracy, but may also boost user trust.

3.1 Domain Knowledge and Preference Elicitation

Co-dependence between user and agent requires that they synchronize their world model. This requirement relates mainly to the agent's self-sufficiency and interdependence, which can be enhanced by imparting the agent with accurate and timely user preferences about the negotiation process and co-constructing the real-world intricacies of the domain.

Preference elicitation on-the-fly

In order to faithfully represent the user, an autonomous negotiator needs to engage with the user to make sure it constructs an accurate preference model (see e.g. [Hunter, 2015]). However, users are often unwilling or unable to engage with a negotiation system, and hence prudence needs to be exercised when interacting with the user to avoid elicitation fatigue. This is especially important in domains where people are notably reluctant to engage with the system at length, for instance in privacy negotiations.

As a consequence, automated negotiators of the future are required to not only strike deals with limited available user information, but also to assess which additional information should be elicited from the user, while minimizing user bother [Baarslag and Kaisers, 2017]. This challenge is still as relevant (and for the most part still unaddressed) as when it was raised in [Jennings *et al.*, 2001]. However, as a way forward, we believe future research should particularly emphasize *preference elicitation on-the-fly*: that is, active preference extraction *during* negotiation(s). Potential benefits include a significantly reduced initial preference elicitation phase (as featured in many negotiation support systems) and the ability to select the most informative query to pose to the user at the most relevant time.

To facilitate this, new performance-based metrics are required that can assess how supplementary preference information influences negotiation performance. Adaptive utility elicitation models provide a good starting point for representing probabilistic utility-based preferences that allow for incremental updating over time (e.g. using Bayesian reasoning), in the vein of [Chajewska *et al.*, 2000]. The viability of a negotiation query can for instance be measured in terms of the expected value of information [Boutilier, 2002] in order to assess the marginal utility of altering belief states.

Another challenge is for a negotiation strategy to determine its actions effectively in light of its imprecise information state. Techniques for decision making under uncertainty

could assist in this and could thereby give rise to novel negotiation strategy concepts, for instance by incorporating the notion of expected *expected* utility [Boutilier, 2003] to express the expected negotiation payoff over all possible instantiations of the user model.

The above discussion largely follows the standard assumptions of rational choice theory: i.e. that people's preferences can be accurately elicited. Unfortunately, several idiosyncrasies of human psychology complicate these assumptions. Not only do people often have difficulty explicitly expressing their preferences, a person's willingness to accept an agreement is also only partially determined by how they feel about the final agreement. For example, Curhan's Subjective Value Inventory [Curhan *et al.*, 2006] identifies four factors that predict which agreements people will accept. Besides feelings about the material outcome (e.g., "the extent to which the terms of the agreement benefit you"), agreements are shaped by feelings about the self (e.g., "did you lose face"), feelings about the process (e.g., "did the counterpart listen to your concerns") and feelings about the relationship (e.g., "did the negotiation build a good foundation for a future relationship").

Research also illustrates that elicited utility functions are highly sensitive to subtle contextual factors. For example, *framing effects* emphasize that preferences between outcomes can reverse depending on whether they are seen as losses or gains with respect to some reference point. In a negotiation, the reference point is often the perceived value that the other party receives, even though this knowledge doesn't change the individual's objective outcome. As a result, outcomes can be readily manipulated simply by changing the form and nature of information conveyed [Gratch *et al.*, 2016]. More broadly, valuations in a negotiation are shaped by emotion, including emotions that arise from the process, but also beliefs about what other parties feel (see, e.g., [Barry *et al.*, 2004]). Given the highly context-sensitive nature of on-the-fly preference elicitation, such considerations will have to be taken into account in its design and implementation.

Domain modeling

The quality of the negotiation outcome depends not only on the faithfulness of the preference model of an autonomous negotiator, but also on the accuracy of the domain model. The old 'garbage in, garbage out' truism applies here, as the quality of the offered solution depends so heavily on a correct domain description.

However, domain modeling, and certainly formal modeling, is an expertise that cannot be expected from an arbitrary user. Therefore, users require either expert guidance, or explicit domain modeling support. Modeling in close cooperation with a domain expert runs the risk of perpetuating people's uncertainty about the model, thereby limiting their ability to make necessary adjustments. When modeling support is provided by the system, the knowledge representation language used will be inherently simple as it has to be understood by arbitrary negotiators. This is especially important in domains where users employ automated negotiation without any expertise, such as in the smart grid, which can result in the wrong evaluation of bids. Highly accurate models, on the other hand, also have their disadvantages: they can display complex non-linearities [Lopez-Carmona *et al.*, 2012], in which case even assessing the utility of a proposal can prove NP-hard [de Jonge *et al.*, 2015].

This inspires the following open research question: what is the impact of simplifying the domain and preference models to keep the layman user on board? An answer might come from using two models, as suggested in [Hindriks *et al.*, 2008]: an accurate, but complex one that serves as a reference model for the agent, and a more comprehensive one for interaction with the user. Proper clarification and explanation could then be elicited from a process of co-creation [Sanders and Stappers, 2008] or participatory design [Simonsen and Robertson, 2012] between modeling experts and domain experts. Ideally, a reflecting phase should be included during and after negotiations, in which the human (and perhaps eventually the agent) can provide feedback to allow for long-term co-evolution.

The above points also apply to the appropriateness and understandability of the protocol governing the negotiation. Typically, a pre-negotiation phase provides an opportunity for the negotiation parties to engage in a debate about what protocol to employ. A corresponding challenge is to construct a best practice repository for negotiation techniques, as mentioned in [Jennings *et al.*, 2001]. This has been tackled at least partially through recent efforts in creating a negotiation handbook for negotiation protocols [Marsa-Maestre *et al.*, 2013].

Whatever approach is chosen, experts in formal modeling will be needed to instantiate a domain model that sufficiently captures all salient features. Those experts are pivotal to the negotiation agent business model and will be responsible for mapping user-understandable interests to the negotiation issues within complex domains. These are likely to become future jobs; i.e., real estate agents informing procurement agents of the future. Relevant research areas, and courses for training these experts, will be on collaborative and supportive modeling.

3.2 Long-term Perspective

Given the effort involved in domain modeling and preference elicitation, the opportunities for automated negotiation are even clearer in domains where an agent frequently faces similar negotiation situations. Most research on negotiation agents, however, has focused on single encounters. The different challenges and opportunities for such long-term negotiations hinge on the volatility of both the opponent pool and the user's preferences.

Repeated encounters

There are many propitious opportunities for applying negotiation in repeated encounters. For example, in community energy exchange [Alam *et al.*, 2015], agents can trade energy from storage and local sources between neighboring homes and businesses. Another example is the smart home, where different occupants will have different needs and preferences and have to reach mutual agreements, e.g. about the temperature of the house and the use of devices. Other settings, in which the agent faces many different opponents, include self-driving vehicles, where vehicle-to-vehicle and vehicle-

to-infrastructure negotiation can play an important role (e.g., negotiating priority at intersections).

Negotiation opportunities for isolated encounters can be very limited, since often a resource (e.g. electricity or giving way) is needed without necessarily offering anything immediately in return (except possibly money or virtual currencies). However, explicitly considering the *temporal dimension* allows agents to receive or concede something now in return for conceding or receiving the same resource later. In other words, single-issue, distributive negotiations can be turned into richer, multi-issue, integrative negotiations, with more scope to achieve win-win solutions [Mell *et al.*, 2015].

A significant challenge for long-term reciprocal encounters is that future needs are often uncertain, and so it is difficult to commit to giving up or requesting specific future resources. Possible solutions involve money or virtual currencies which can be redeemed at a later stage and can undergo temporal discounting if necessary, but they do not address the distributive nature of multi-issue negotiation. They also introduce additional challenges: using actual money requires an exchange rate with the resources involved, while it may not be desirable to introduce money in certain settings; e.g. when they rely, to some degree, on unincentivized cooperation and altruistic behavior. Virtual currencies (including distributed ledger approaches) can be traded bilaterally in a "like for like" manner, addressing the exchange problem, but then other issues arise, e.g. how much does each agent receive to begin with, what happens if an agent runs out, and to what extent do they provide a real incentive if agents can go into debt without any consequences?

Another possible solution is to rely on altruism and using trust ratings and reputation systems to provide the desired incentives (e.g. using favors and ledgers [Mell *et al.*, 2015]). In such cases, 'altruism' can be a self-interested strategy if this is reciprocated at a later state, possibly involving a different opponent. While reputation mechanisms are well known to incentivize cooperation in the prisoner's dilemma, more research on this is needed in the context of (repeated) automated negotiation.

Unfortunately, negotiation methods that seek to identify efficient and fair (envy-free) agreements face, in addition to the above, a number of *psychological* challenges. People adopt a variety of interpretations as to what is fair and negotiations often involve disputes over which principle to apply [Welsh, 2003]. For example, in the context of organ donation, the *equity principle* would allocate resources on the basis of ability, effort or merit, the *equality rule* would treat individuals the same, whereas the *principle of need* is usually achieved by allocating according to individuals medical condition, socio-economical status or other relevant needs. Other complications involve moral constraints on certain exchanges. For example, it is considered morally repugnant to exchange money for bodily organs, so an agreement that combines material interests with sacred values may be seen as substantially worse than an independent evaluation of these elements would suggest [Dehghani *et al.*, 2010].

Although these challenges might seem insurmountable, there are several ways to incorporate these biases into conventional computational methods. One approach is to incorporate psychological factors into the utility function. Indeed, Fehr and Schmidt have shown how this can be done without violating the basic tenets of utility theory [Fehr and Schmidt, 2006]. Some of the challenges with fairness can be addressed by making the process more transparent (Section 3.3). Another approach is to incorporate modest psychological extensions to rational methods. For example, framing effects can be handled through the use of prospect theory (e.g., [Yang *et al.*, 2011]).

Non-stationary preferences

While short-lived instantiations of representational agents may assume that there are some true and stationary preferences to be elicited from the user, in long-term negotiations, these very preferences may evolve over the course of weeks or months according to certain *preference dynamics*. If an autonomous negotiator acts on elicited information for an extended period of time without accounting for existing drift in preferences, it will erroneously fulfill outdated design objectives. Even if the drop in performance is noticed by the user, this leads to a plunge in user trust and adoption, or a de-facto shortened time of deployment. This is a typical example of opacity that can result from an excess of unchecked autonomy [Norman, 1990]. As a result, long-term negotiation requires an increase in co-dependence, at the cost of throttled-down self-directedness; e.g., by repeated assessment of the preference representation quality, with intermittent elicitation actions whenever their anticipated benefits exceed their costs.

This reframes the challenge posed in Section 3.1 of preference elicitation to *cost-efficient tracking of non-stationary preferences* in long-term negotiation, with possible applications ranging from leisure bookings to business-to-business (B2B) negotiations. Inspiration for tackling this challenge may come from the area of news recommender systems, which has embraced context-dependent models [Adomavicius and Tuzhilin, 2015] and preference dynamics [Li *et al.*, 2014] in response to the inherent need to capture fast-paced preference evolution. Such models have promising merit for being transfered to negotiation strategies that balance the preciseness of preference representation with relevant and timely but costly elicitation, extending preliminary work in that area [Baarslag and Kaisers, 2017].

3.3 User Trust and Adoption

While the agent depends on the user for knowledge and guidance (as described in Section 3.1), the user relies on a self-directed agent for a good outcome. To alleviate unwillingness to relinquish control and to guarantee user satisfaction with and adherence to the final outcome, the user needs to trust the system through co-participation, transparency, and proper representation.

User participation

Lessons learned from collaborative human-robot teams indicate that it is important to be able to escalate to the meta-level (i.e. have humans participate) when necessary [Feltovich *et al.*, 2012]. The need for escalating to a higher authority applies whenever a negotiator represents a group or a company (e.g., a union, or stakeholder organizations in general). In

such cases, the negotiator can only make deals that fall within certain margins.

The idea of collaborative control, or mixed-initiative control (see e.g. [Fong *et al.*, 2001; Feltovich *et al.*, 2012]), might become essential to obtain the most out of complex negotiations. In this envisioned line of research, each negotiation party consists of at least one human and one negotiation agent. The agent should do the brunt of the negotiation work to find possible agreements with the other negotiation parties and which can presented to their human partners for feedback and new input. The research challenge is to determine when, how, and how often to switch the initiative from human to agent and vice versa.

Transparent consequences

There is an inherent tension between increased self-directedness and trust, which dampens the adoption of increasingly autonomous negotiators: on the one hand, an autonomous negotiatior's relevance is directly proportional to its ability to impact the user independently in meaningful ways (e.g. fiscal, well-being, reputation, and so on); but, in turn, the user's trust and willingness to relinquish control is conditional on understanding the agent's reasoning and consequences of its actions. The two can be reconciled by making the outcome space more *transparent* to the user, and by enabling the user to specify the permissible means in the form of *principles*. The challenge is that the negotiation agent's reasoning abilities may very well exceed the domain insights of a nonspecialist user, thus requiring a translation from stochastic performance models of self-directed expert reasoning into laymen terms that adequately convey expectations and risks.

Note that we suggest transparency as the key concept here, which subsumes Jennings' notion of predictability [Jennings *et al.*, 2001]. Predictability is essential towards the user to instill trust, but can be disastrous towards the opponent because of the potential for exploitability. We argue unpredictable behavior is in fact desirable as a negotiation tactic as a confusing and randomization device, as long as the consequences are transparently explained to the user.

The uncertainty inherent in negotiation can be captured in performance models and risk metrics, where the complexity should be scaled to the criticality of the consequences for the user. If the performance intervals are sub-critical, then simple guarantees on the range of possible outcomes may suffice (such as price bounds provided by Uber for individual rides), leaving it up to the user to build and judge the average performance model; otherwise, measures of risk are required, such as Conditional Value at Risk (CVar) [Shafie-khah *et al.*, 2016]

In the end, the potency of autonomous negotiators is as much contingent on the acceptance by their users as by their counter-parties. The most promising incubators of autonomous negotiators are ecosystems in which autonomous agents provide a unique source of societal value that is distributed over all stakeholders, as in the application of demand response for smart grids. Open platforms for value distribution have recently seen increased attention in flagship applications such as the cryptocurrency *bitcoin* and the decentralized world wide web *Blockstack* [Ali *et al.*, 2016]. The digital API of these systems offers fertile grounds for a level playing field for competition and may soon provide a common interface for automated negotiators.

4 Concluding Observations

Autonomous systems that are capable of negotiating on our behalf are among society's key technological challenges for the near future, and their uptake is important for many critical economical application areas. In this paper, we present a roadmap to arrive at such representative and trusted negotiators that are endowed with a long-term perspective. By continuing along this trajectory, negotiation research can address perhaps the biggest challenge of all: a co-active approach that simultaneously advances the autonomy of a negotiation agent in all its aspects.

Finally, looking even further forward, it is worth noting that people negotiate differently through intermediaries than they would face-to-face. The literature on *representation effects* suggests that people may show less regard for fairness and ethical behavior when negotiating through a third (human) party [Chugh *et al.*, 2005]. Indeed, human lawyers are ethically permitted and, to some extent, expected to lie on behalf of their clients [Gratch *et al.*, 2016]. This raises the question as to whether agents should similarly lie on behalf of a user, e.g. by using argumentation and persuasion technology [Dimopoulos and Moraitis, 2014]. Analogous to recent research on ethical dilemmas in self-driving cars, people may claim that negotiation agents should be ethical, but sacrifice these ideals if it maximizes their profits. The natural dichotomy between recognizing the agent's autonomy and taking responsibility for its actions is best resolved by acknowledging user responsibility for the agent's design objectives (what should be achieved) and principles (how it should be achieved). This also illustrates an additional impetus for having humans understand the agent: feeling responsibility for the agent's actions implies an understanding what the agent is doing. Fortunately, some recent research on agent negotiators suggests that people may act more ethically when negotiating via computer agents [de Melo *et al.*, 2016], but far more research is needed to understand how *artificial representation effects* arise.

Acknowledgments

This research has received funding through the ERA-Net Smart Grids Plus project Grid-Friends, with support from the European Union's Horizon 2020 research and innovation programme.

References

[Adomavicius and Tuzhilin, 2015] G Adomavicius and A Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 191–226. Springer, 2015.

[Alam *et al.*, 2015] M Alam, E H Gerding, et al. A scalable inter-dependent multi-issue negotiation protocol for energy exchange. In *IJCAI*, pages 1098–1104, 2015.

[Ali *et al.*, 2016] M Ali, J Nelson, et al. Blockstack: A global naming and storage system secured by blockchains. In *USENIX ATC*, pages 181–194, 2016.

[Baarslag and Kaisers, 2017] T Baarslag and M Kaisers. The value of information in automated negotiation: A decision model for eliciting user preferences. AAMAS 2017, 2017.

[Baarslag *et al.*, 2015] T Baarslag, R Aydoğan, et al. The automated negotiating agents competition, 2010-2015. *AI Magazine*, 36(4):115–118, 2015.

[Barry *et al.*, 2004] B Barry, I S Fulmer, et al. I laughed, I cried, I settled: The role of emotion in negotiation. *The handbook of negotiation and culture*, pages 71–94, 2004.

[Boutilier, 2002] C Boutilier. A POMDP formulation of preference elicitation problems. In *AAAI*, pages 239–246, 2002.

[Boutilier, 2003] C Boutilier. On the foundations of expected expected utility. IJCAI'03, pages 285–290, 2003.

[Bradshaw *et al.*, 2003] J M Bradshaw, P J Feltovich, et al. Dimensions of adjustable autonomy and mixed-initiative interaction. In *International Workshop on Computational Autonomy*, pages 17–39. Springer, 2003.

[Chajewska *et al.*, 2000] U Chajewska, D Koller, et al. Making rational decisions using adaptive utility elicitation. In *AAAI*, pages 363–369, 2000.

[Chugh *et al.*, 2005] D Chugh, M H Bazerman, et al. Bounded ethicality as a psychological barrier to recognizing conflicts of interest. *Conflicts of interest: Challenges and solutions in business, law, medicine, and public policy*, pages 74–95, 2005.

[Curhan *et al.*, 2006] J R Curhan, H A Elfenbein, et al. What do people value when they negotiate? Mapping the domain of subjective value in negotiation. *Journal of personality and social psychology*, 91(3):493, 2006.

[de Jonge *et al.*, 2015] D de Jonge, C Sierra, et al. *Negotiations over large agreement spaces*. PhD thesis, 2015.

[de Melo *et al.*, 2016] C M de Melo, S Marsella, et al. Do as I say, not as I do: Challenges in delegating decisions to automated agents. In *AAMAS*, pages 949–956, 2016.

[Dehghani *et al.*, 2010] M Dehghani, S Atran, et al. Sacred values and conflict over Iran's nuclear program. *Judgment and Decision Making*, 5(7):540, 2010.

[Dimopoulos and Moraitis, 2014] Yannis Dimopoulos and Pavlos Moraitis. Advances in argumentation based negotiation. *Negotiation and Argumentation in Multi-agent Systems: Fundamentals, Theories, Systems and Applications*, pages 82–125, 2014.

[Eisenberg and Lanvers, 2009] T Eisenberg and C Lanvers. What is the settlement rate and why should we care? *Journal of Empirical Legal Studies*, 6(1):111–146, 2009.

[Fehr and Schmidt, 2006] E Fehr and K M Schmidt. The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1:615–691, 2006.

[Feltovich *et al.*, 2012] P J Feltovich, B van Riemsdijk, et al. Autonomy and interdependence in human-agent-robot teams. 2012.

[Fong *et al.*, 2001] T Fong, C Thorpe, and C Baur. *Collaborative control: A robot-centric model for vehicle teleoperation*, volume 1. Carnegie Mellon University, The Robotics Institute, 2001.

[Gratch *et al.*, 2016] J Gratch, Z Nazari, et al. The misrepresentation game: How to win at negotiation while seeming like a nice guy. In *AAMAS*, pages 728–737, 2016.

[Hindriks *et al.*, 2008] K Hindriks, C M Jonker, et al. Avoiding approximation errors in multi-issue negotiation with issue dependencies. In *ACAN*, 2008.

[Hu and Bolivar, 2008] W Hu and A Bolivar. Online auctions efficiency: a survey of ebay auctions. In *WWW*, pages 925–934, 2008.

[Hunter, 2015] A Hunter. Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 3055–3061. AAAI Press, 2015.

[Jennings *et al.*, 2001] N R Jennings, P Faratin, et al. Automated negotiation: Prospects, methods and challenges. *GDN*, 10(2):199–215, 2001.

[Johnson *et al.*, 2011] M Johnson, J M Bradshaw, et al. *The Fundamental Principle of Coactive Design: Interdependence Must Shape Autonomy*, pages 172–191. Springer Berlin Heidelberg, 2011.

[Kersten and Lo, 2001] G E Kersten and G Lo. Negotiation support systems and software agents in e-business negotiations. In *ICEB*, pages 19–21, 2001.

[Li *et al.*, 2014] L Li, L Zheng, et al. Modeling and broadening temporal user interest in personalized news recommendation. *Expert Systems with Applications*, 41(7):3168–3177, 2014.

[Lopez-Carmona *et al.*, 2012] M A Lopez-Carmona, I Marsa-Maestre, et al. Addressing stability issues in mediated complex contract negotiations for constraint-based, non-monotonic utility spaces. *Autonomous Agents and Multi-Agent Systems*, 24(3):485–535, 2012.

[Luck *et al.*, 2003] M Luck, M D'Inverno, et al. *Autonomy: Variable and Generative*, pages 11–28. Springer US, Boston, MA, 2003.

[Marsa-Maestre *et al.*, 2013] I Marsa-Maestre, M Klein, et al. From problems to protocols: Towards a negotiation handbook. *Decision Support Systems*, 2013.

[Mell *et al.*, 2015] J Mell, G Lucas, et al. An effective conversation tactic for creating value over repeated negotiations. In *AAMAS '15*, pages 1567–1576, 2015.

[Norman, 1990] D A Norman. The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 327(1241):585–593, 1990.

[Sanders and Stappers, 2008] E B-N Sanders and P J Stappers. Co-creation and the new landscapes of design. *Co-design*, 4(1):5–18, 2008.

[Sebenius, 1992] J K Sebenius. Negotiation analysis: A characterization and review. *Management Science*, 38(1):18–38, 1992.

[Shafie-khah *et al.*, 2016] M Shafie-khah, D Z Fitiwi, et al. Simultaneous participation of demand response aggregators in ancillary services and demand response exchange markets. In *2016 IEEE/PES T&D*, pages 1–5, May 2016.

[Simonsen and Robertson, 2012] J Simonsen and T Robertson. *Routledge international handbook of participatory design*. Routledge, 2012.

[Welsh, 2003] N A Welsh. Perceptions of fairness in negotiation. *Marq. L. Rev.*, 87:753, 2003.

[Yang *et al.*, 2011] R Yang, C Kiekintveld, et al. Improving resource allocation strategy against human adversaries in security games. In *IJCAI*, volume 22, page 458. Citeseer, 2011.

[Young, 1991] H P Young. *Negotiation analysis*. University of Michigan Press, 1991.