Introduction to the Special Issue on Human Interaction with Artificial Advice Givers

NAVA TINTAREV, Bournemouth University
JOHN O'DONOVAN, University of California Santa Barbara
ALEXANDER FELFERNIG, Graz University of Technology

Many interactive systems in today's world can be viewed as providing advice to their users. Commercial examples include recommender systems, satellite navigation systems, intelligent personal assistants on smartphones, and automated checkout systems in supermarkets. We will call these systems that support people in making choices and decisions *artificial advice givers (AAGs)*: They propose and evaluate options while involving their human users in the decision-making process. This special issue addresses the challenge of improving the interaction between artificial and human agents. It answers the question of how an agent of each type (human and artificial) can influence and understand the reasoning, working models, and conclusions of the other agent by means of novel forms of interaction. To address this challenge, the articles in the special issue are organized around three themes: (a) human factors to consider when designing interactions with AAGs (e.g., over- and under-reliance, overestimation of the system's capabilities), (b) methods for supporting interaction with AAGs (e.g., natural language, visualization, and argumentation), and (c) considerations for evaluating AAGs (both criteria and methodology for applying them).

CCS Concepts: • **Information systems → Decision support systems**; • **Human-centered computing → Human computer interaction (HCI);**

Additional Key Words and Phrases: Agent-based interaction, anthropomorphism, argumentation, advising agents, human decision making, emotions, facial actions, feedforward and feedback, gestures, human-agent interaction, human argumentation, human-like computing, interaction paradigms, recommendation, reliance on automation, use image, vague language, visualization

ACM Reference Format:
Nava Tintarev, John O'Donovan, and Alexander Felfernig. 2016. Introduction to the special issue on human interaction with artificial advice givers. ACM Trans. Interact. Intell. Syst. 6, 4, Article 26 (December 2016), 12 pages.
DOI: http://dx.doi.org/10.1145/3014432

1. AIMS AND SCOPE OF THE SPECIAL ISSUE

Some systems that support people in making choices and decisions can be viewed as *artificial advice givers* (AAGs): They propose and evaluate options while involving their human user in the decision-making process. The five articles in this special issue consider two types of AAGs: interactive decision support systems and recommender systems. These systems differ in terms of their degree of autonomy and the extent to which users can influence reasoning processes and conclusions. For these sorts of systems, there are benefits and challenges to keeping the human decision makers in the loop (as opposed to, for example, the case of a fully autonomous car). AAGs enable users not only to understand the system's advice and reasoning but also to call it

Authors' addresses: N. Tintarev, Poole House, Talbot Campus, Fern Barrow, Poole, BH12 5BB; email: nava.tintarev@abdn.ac.uk; J. O'Donovan, Department of Computer Science, University of California, Santa Barbara, CA 93106-5110; email: jod@cs.ucsb.edu; A. Felfernig, Technische Universität Graz, Inffeldgasse 16b, 8010 Graz; email: afelfern@ist.tugraz.at.

DOI: http://dx.doi.org/10.1145/3014432

into question and to influence the system's reasoning. Over time, such a collaboration can support the evolution of the decision makers' understanding and requirements concerning the domain in question as well as the evolution of the advice-giving system. This special issue considers how an agent of each type (human and artificial) can influence and understand the reasoning, working models, and conclusions of the other agent by means of novel forms of interaction.

The articles in this special issue discuss (a) human factors to consider when designing interactions with AAGs (e.g., over- and under-reliance, overestimation of the system's capabilities), (b) methods for supporting interaction with AAGs (e.g., natural language, visualization, and argumentation), and (c) considerations for evaluating AAGs (both criteria and methodology for applying them). By designing and testing improved forms of support for interactive collaboration between human decision makers and artificial advice givers, we can enable decision-making processes that better leverage the strengths of both types of collaborators.

2. HUMAN INTERACTION WITH ARTIFICIAL ADVICE GIVERS: WHAT AND WHY?

This section begins by summarizing the background of artificial advice giving, and it makes a case for decision making that is supported by artificial advice givers. It then describes human factors that need to be considered when designing interactions with AAGs. It goes on to give an overview of current approaches to decision making with AAGs, such as natural language dialog, argumentation, and visualization. The section concludes with a discussion of criteria for evaluating AAGs.

2.1. Background

Advice-giving systems have been around for a few decades in the form of intelligent tutoring systems [Sleeman and Brown 1985], expert systems [Carroll and McKendree 1987], and recommender systems [Schafer et al. 1999]. More recently, there has been a renewed interest and investment in both artificial intelligence, as well as new interaction paradigms [Dix 2016]. This renewed interest is partly due to successes in artificial intelligence such as Watson playing Jeopardy on the level of a human expert [Ferrucci et al. 2010; Ferrucci 2012], deep learning enabling an intelligent system to play (and win) Go [Silver et al. 2016], and commercial semi-autonomous cars. Recent advances also include commercial advice-giving systems such as Apple's Siri, Microsoft's Cortana, and Google Now.

These developments have enabled better artificial advice giving that supports and even augments human capabilities. As these advice-giving systems increase in complexity, their designers have also come to realize that a standard graphical user interface (GUI) is often not sufficient to harness their power. These systems build on the Computers as Social Actors paradigm and the Media Equation, which posit that people treat computers much as they treat other people [Nass et al. 1994; Reeves and Nass 1996]. To support these more complex interactions, AAG systems increasingly support agent-based interaction (an interaction paradigm that has been a topic of human-computer interaction research for several decades [Behrend and Thompson 2011; Qiu and Benbasat 2009; Hess et al. 2005; Cowell and Stanney 2005; Bickmore and Cassell 2001; Walker et al. 1994; Quintanar et al. 1987; Nickerson 1976]). In these systems, the user interacts with an autonomous virtual entity (i.e., an agent) using, for example, natural language as a means of interaction.

2.2. A Case for Decision Making That Is Supported by Artificial Advice Givers

It is in this context of large-scale advancement in artificial intelligence that the National Science Foundation's division on Information and Intelligent Systems[1] in the

[1]http://www.nsf.gov/div/index.jsp?div=IIS, retrieved July 2016.

United States has started to promote research on the interrelated roles of humans, systems, and information. A similar investment is being made in a research program on *Explainable Artificial Intelligence* by the Defense Advanced Research Projects Agency (DARPA).[2] Similarly, one of the main UK research funding bodies, Engineering and Physical Sciences Research Council, has been exploring possibilities for improved, humanlike Artificial Intelligence (AI): *"offering the prospect of computation which is akin to that of humans, where learning and making sense of information about the world around us can match our human performance"* [EPSRC 2016]. This investment also reflects a renewed interest in natural and humanlike interaction paradigms.

In tandem, there have been some concerns in the public and the media that AI systems will soon overtake humans in terms of capacity, threatening our jobs or even our safety [Hawking et al. 2014]. But the above systems are focused on specific tasks and require a great deal of training and tuning. These limitations are in some domains mitigated by the availability of large amounts of data, driven by research in the field of Big Data. However, even the domains for which data are sufficiently rich have their own limitations with regard to the quality of the data and the biases introduced through data filtering/cleaning or parametrization of algorithms [Amatriain et al. 2011; Witten and Frank 2005]. Many tasks, such as visual pattern finding and dealing with novel situations, are still performed better by people.

In response to the recent developments in AI and to these concerns, we take the approach that AI is moving toward a collaboration that augments human cognitive capacity, rather than toward the development of "slave" tools that are growing into digital "masters." It is our view that AAGs will support different and complementary types of abilities than those that are normally available in the context of human-only cognition. So rather than taking over humanity or taking (all) human jobs, we believe that developments in AI will form the basis of a new generation of AAG systems that augment human cognition and support human decision making with the help of computational power.

2.3. Human Cognitive Factors in Interactions with AAGs

Some research has found that humans do not always make the best possible decisions when working together with AAGs. Before listing these examples, however, we note that these are specific studies and that results might differ if the studies were to be replicated with one or more key aspects modified. Previous work has, for example, found that personalized explanations of recommended items (such as cameras or movies) have led to worse decision making, even though users were more satisfied with personalized explanations [Tintarev and Masthoff 2012]. Similarly, allowing users to modify keywords in their news profiles resulted in worse news recommendations [Ahn et al. 2007]. In group decision scenarios, early knowledge about the preferences of other group members had a negative influence on the perceived quality of a group decision [Stettinger et al. 2015]. In the worst case, polarization of views can occur in both algorithms [Bakshy et al. 2015] and in the users themselves, as claimed in Pariser's book on the *Filter Bubble* [Pariser 2011]. The design of AAGs should therefore be based on an understanding of how people make judgments and decisions and how they are influenced by other agents. Simply relying on introspection and intuition about human cognition is not adequate, since research has brought to light many counter-intuitive results. For this reason, this special issue addresses several human-centered factors to consider in decision making, such as a preference for humanlike voices (see Section 3.3) [Clark et al. 2016], overestimation due to a faulty user's model of the system (see Section 3.2) [Knijnenburg and Willemsen 2016], and (over- and under-) reliance on advice (see Section 3.1) [Sutherland et al. 2016].

[2]http://www.darpa.mil/program/explainable-artificial-intelligence, retrieved August 2016.

2.4. Methods for Supporting Interaction with AAGs

These human-AAG systems will leverage humans' pattern matching and their capacity to handle novel situations on the one hand and the computational and analytical power of AAGs on the other. In addition, each decision-making partner (artificial or human) is likely to have different sources of information and ultimately to work on the basis of different premises. A dialogue between the partners can enable a common understanding of the joint decision-making context. Such dialogues can be modeled using formal argumentation. Argumentation has become a popular approach to non-monotonic reasoning—the kind of reasoning in which the partners can draw tentative conclusions, enabling each partner to update his or her conclusion(s) on the basis of further evidence. Suggested application areas have ranged from law [Prakken and Sartor 1997] to practical reasoning in multi-agent systems [Atkinson and Bench-Capon 2007] to decision support systems [Cerutti 2011]. However, it is only recently that studies have investigated the relationship between *human intuition* and formal argumentation theory [Cerutti et al. 2014; Rahwan et al. 2010]. In this special issue, Rosenfeld and Kraus delve deeper into the question of how well formal models and predictive methods align with human deliberative dialogue (see also Section 3.4 for a summary) [Rosenfeld and Kraus 2016].

A joint dialog between humans and artificial advice givers has the potential to help human partners understand the grounds for an AAG's advice and possibly to change their minds on learning new information. This sort of dialogue also allows humans to add or modify information that the advice giver is using (in particular, when the system is missing information [Tintarev et al. 2013; Tintarev and Kutlak 2014]), potentially resulting in the AAG reaching a new conclusion. In the coming years, we foresee an increasing number of these sorts of collaborations between humans and AAG systems.

This sort of joint decision making and *deliberation* between artificial and human decision partners can ultimately help people reach better informed and better considered decisions. Improved methods in conversational agents, using more natural paradigms such as speech, provide a particular opportunity for reaching *joint* conclusions. While intuitive and automatic decisions are often correct, they are also radically insensitive to both the quality and quantity of information that gives rise to impressions and intuitions [Kahneman 2011, p. 87]. A joint discussion could increase the likelihood of activating reasoning processes and questioning prior beliefs. For example, it has been found that actively considering alternative points of view can counteract reliance on the first piece of information offered in a negotiation [Galinsky and Mussweiler 2001].

In some instances and domains, the main challenge may be one of information overload. Mutlu et al. argue in this special issue that a visual representation is sometimes more suitable for dealing with the information overload problem [Mutlu et al. 2016]. To be appropriate, a visualization has to follow certain known guidelines to find and distinguish patterns visually and encode data therein. A visualization tells a story of the underlying data; yet, to be appropriate, it has to clearly represent those aspects of the data the viewer is interested in. It is important to identify which aspects of a visualization are important to the viewer and how to capture and use those aspects to recommend visualizations. To address these issues, Mutlu et al. introduce a system for adapting visualizations to the viewer (see Section 3.5 for a brief summary).

2.5. Criteria for Evaluating Advice Giving Systems

To increase the likelihood of successful AAG-human synergies, it is helpful to define criteria for success. Many of the criteria previously used to evaluate the interactive intelligent systems above (e.g., intelligent tutors, expert systems, and recommender systems) are also applicable to the broader area of AAGs. We summarize a list of criteria

Table I. Evaluation Criteria for Artificial Advice Givers. The Xs indicate whether the criterion is important from the point of view of the User, the designers of the AAG, or both.

Criterion	Definition	User	AAG
Effectiveness	How good are the user's/system's decisions?	X	X
Efficiency	How fast can the user/system make decisions?	X	X
Learning	How much new and relevant information does the user/system gain?	X	X
Persuasion	Is the user convinced or influenced by the advice given?	—	X
Transparency	Does the user understand how the system works?	X	—
Satisfaction	How usable is the system?	X	—
Scrutability	Can the user correct the system when is wrong?	X	—
Trust	How much confidence does a user have in the system?	—	X

for evaluating AAGs in Table I. Some of these criteria concern different aspects of the success of an interaction *from the user's point of view*, such as subjective satisfaction or transparency. Others express aspects of positive outcomes *from the point of view of an AAG* that is trying to influence the user (e.g., extent of persuasion and trust). Yet a third category may express benefits *from both points of view*, such as effectiveness and learning. In these cases, the agents may be influencing each other with the same aim. For example, the human and agent may have different and complementary information that they need to convey to the other agent.

Of the criteria in Table I, the most intuitive success criterion for AAG-human synergies might arguably be the quality of the outcomes, such as the avoidance of mistakes and the correctness of decisions, or *effectiveness* in decision making. In this special issue, the work of Kniijnenburg and Willemsen evaluated interactions using several metrics such as "ineffectiveness" through proxy metrics such as discontinuation of the experiment [Knijnenburg and Willemsen 2016].

To make the best use of human-AAG collaborations, we also have to consider the barriers to the acceptance of AAGs and their advice. In some of these cases, increased transparency and explanations can be used to alleviate the users' concerns [Herlocker et al. 2000; Tintarev and Masthoff 2015]. Mechanisms for improving transparency have already been introduced to different classes of intelligent systems in the past, in domains from intelligent tutors [Brusilovsky et al. 1996; Dimitrova 2003] to autonomous systems [Cerutti et al. 2014], decision support systems [Bennett and Scott 1985; Kulesza et al. 2015], and recommender systems [Kang et al. 2016; Tintarev et al. 2015; Schaffer et al. 2015; Tintarev and Masthoff 2015]. The need for transparency is also acknowledged by regulatory bodies. For example, the European Union passed a General Data Protection Regulation[3] in May 2016 (effective from 2018) that will also create a "right to explanation" whereby a user can ask for an explanation of an algorithmic decision that was made about them [Goodman and Flaxman 2016]. As information and algorithms become more complex, visualizations may be more suitable than textual interactions. The work by Multlu et al. in this special issue considers that these visualizations will need to consider both the data they represent and the cognitive preferences of different users (see also the summary in Section 3.5) [Mutlu et al. 2016].

It may also be worthwhile to measure *persuasion* or the user's acceptance of the advice itself. Another likely success marker regards the *efficiency* of decisions, measured as the time and effort by the human to make decisions. For example, Knijnenburg and Willemsen measured the number of requests that users made and their time per task [Knijnenburg and Willemsen 2016].

[3]http://ec.europa.eu/justice/data-protection/, retrieved July 2016.

In other instances, it may be more important to consider the user's general *trust* in the system—that is, not their propensity to utilize the advice but their propensity to consider the AAG as a suitable source of information and advice in general. In some systems, it may be more important to focus on the quality of the user's subjective experience while deciding—that is, on his of her *satisfaction* with their interaction with the the AAG. It may also be relevant to consider the effect of interactions over time, such as the extent of the each agent's ability to *learn* from previous interactions. Knijnenburg and Willemsen measured learnability through the proxy metric of the difference in time per task between the first and last task [Knijnenburg and Willemsen 2016].

3. ARTICLES IN THIS SPECIAL ISSUE

Five articles were accepted for the special issue. In this section, we give an overview of these articles, relating them to the concepts introduced so far.

3.1. Effects of the Advisor and Environment on Requesting and Complying with Automated Advice

Sutherland et al. contribute to our understanding of both human factors to consider when designing interactions with AAGs and issues encountered in the evaluation of AAGs [Sutherland et al. 2016]. In order for joint decision making between a human and an artificial advice giver to be effective, it is important that people not only accept advice when it is correct but also reject it when it is incorrect. The complexity of decision tasks, the limited cognitive resources of system users, and a tendency to keep decision effort low are related to the phenomenon of *bounded rationality* [Simon 1955]: Users employ decision heuristics rather than make optimal decisions. Decision making under bounded rationality is a door opener for various types of influences on decision outcomes.

This work investigates factors that may influence the human decision maker—namely the cost of receiving advice (time required for the adviser to give advice), the reliability of the adviser (% of time correct), and the predictability of the environment—on the acceptance of system advice. Sutherland et al. evaluate the the impact of these factors on two types of incorrect joint decision making: (a) the *over-utilization* of suboptimal advisers (accepting advice that should have been rejected) and *under-utilization* of optimal advisers (rejecting advice should have been accepted).

Addressing the challenge of how to evaluate AAGs, Sutherland et al. introduce a digital game environment to evaluate their hypotheses. In addition to a novel virtual evaluation setting, they introduce the *Tarp Technique* for sampling the stimulus space, ensuring that a large range of values are covered as well as ensuring sufficient sampling of the extreme values for the factors (i.e., 100% and 0% accuracy of the adviser, free advice, and completely (un)predictable environments).

3.2. Inferring Capabilities of Intelligent Agents from Their External Traits

Knijnenburg and Willemsen contribute to the special issue an investigation of methods for interaction with agents using *capability* (e.g., level of complexity of linguistic expressions) and *appearance cues* (e.g., humanlike character versus text) [Knijnenburg and Willemsen 2016]. As human interaction with artificial advice givers becomes more natural and similar to how we interact with other humans, it is important to understand when (and how) this naturalness affects the quality of the communication and usability of the system. A first intuition is that a more humanlike interaction is better, and this idea is reflected in many current artificial personal assistants such as Siri, Cortana, Google Now, and Viv. This idea is also in line with the work of Clark et al.

in this special issue (summary in Section 3.3), which found improved interaction for natural speech output compared with artificial speech from the advice giver.

Knijnenburg and Willemsen rightfully revisit the assumption that humanlike is always best by raising the issue of overestimation effects in humanlike agents. They highlight that interfaces of this sort differ from traditional GUIs and that this difference affects users' understanding of how a system works (i.e., the *user's model* of the system [Norman 1986]). While it is widely accepted that users' mental models of systems in turn influence users' behavior, Knijnenburg and Willemsen found evidence that this point also applies in the special case of automated advice givers. More specifically, they found that an agent-based interface had a single integrated model, in which users anthropomorphized (i.e., attributed human form or personality to the system) and inferred a broader set of humanlike capabilities that are not necessarily related to the specific cues displayed by the system.

The system versions with human and capability cues were found to be more usable in terms of efficiency and user satisfaction than the baseline version. On the other hand, the former versions resulted in overestimation. The case for overestimation is supported by several findings. First, more users in these conditions prematurely quit the experiment. Second, users used more complex and natural language (e.g., more first-person references, longer sentences, and more grammatically correct syntax) when using the system versions with more humanlike cues. Finally, users of the humanlike systems tried to make more use of humanlike capabilities like implicit references to context and to earlier parts of the conversation, and they were more likely to ask multiple questions at the same time. It is particularly noteworthy that appearance cues alone could influence the way capabilities were inferred by users. These results are in line with the concerns of critics of artificial agents, such as Shniederman, who have been warning about this type of overestimation since the 1990s (see, e.g., Shneiderman [1997]).

The article also contributes a rich set of criteria for evaluating AAGs. Knijnenburg and Willemsen introduce a number of criteria such as ineffectiveness, efficiency, satisfaction, and learnability. These were measured by proxy through metrics such as discontinuation of the experiment; the number of requests that users made and their time per task; self-reported user satisfaction; and the difference in time per task between the first and last tasks, respectively.

3.3. A Multimodal Approach to Assessing User Experiences with Agent Helpers

This article by Clark et al. contributes to the study of artificial advice givers by evaluating different communication strategies such as vague language and politeness [Clark et al. 2016]. An increasingly common form of artificial advice giver are the automated agent helpers that we interact with in various customer service scenarios. These can be text-based interactions with helpers on websites or interactions with voice-based automated agents on phones. One of the challenges in the study of these systems is the often limited quality of the agent's voice and its lack of appropriate fit to the type of interaction engaged in. Clark et al. contribute several different approaches to this challenge, including a comparison of synthesized and human voices and an analysis of facial expression and gesture as feedback mechanisms. This article is related to Knijnenburg's and Willemsen's article [Knijnenburg and Willemsen 2016] (discussed in Section 3.2) in that the focus is on evaluating automated helper agents with humanlike cues. In contrast, however, the authors of this article present a focused evaluation of *multimodal* aspects of interaction—the effects of voice and video on the interaction experience.

The first approach specifically compares human and synthesized voices in agents using vague language. Here, Clark et al. analyze a 60,000-word text-based corpus of participant interviews to investigate differences in users' attitudes towards automated helper agents in different situations, including analysis of their voices and their use of vague language. Results from the first study indicate that, while the acceptance of vague language is still met with resistance in agent instructors, using a human voice yields more positive results than using the synthesized alternatives.

The second contribution in this article focuses on the development of a novel multimodal corpus of video and text data, which can be used to support multiple analyses of human-agent interactions in agent-instructed assembly tasks. This is a central topic to ACM *TiiS*, since it provides a novel perspective on human interaction with the automated helper agents. In particular, their approach includes an analysis of spontaneous facial actions and gestures during their interaction in the tasks. The authors found that agents are able to elicit these facial actions and gestures of users during interactions, and they posit that further analysis of this nonverbal feedback may help to create a more adaptive advice-giving agent.

Clark et al. conclude the article with a discussion on how the approaches that they explored contribute to furthering the understanding of what it means to interact with software agents. This article is a strong contribution to the special issue theme of methods of supporting interactions with AAGs. This article also contributes to the special issue's focus on considerations for evaluating AAGs, specifically by showing the benefit of leveraging video to glean information from facial expressions and gestures as a feedback mechanism to gauge the performance of artificial advice givers.

3.4. Providing Arguments in Discussions Based on the Prediction of Human Argumentative Behavior

The article by Rosenfeld and Kraus aids our understanding of methods for interaction between human and AAGs [Rosenfeld and Kraus 2016]. In particular, it considers the importance of influence from both system to user and vice versa by contributing a deeper understanding of which formal models of deliberation best represent (natural) human deliberative behavior. The authors compare different approaches to predicting arguments: *formal argumentation theory*, *heuristic approaches*, *machine learning*, and *transfer learning*.

Argumentative dialogs figure importantly in everyday life. Researchers try to understand and explain major properties of such dialogs and the impact of different kinds of argumentative behavior. Rosenfeld and Kraus focus on argumentative dialogs in *deliberations*—for example, a couple is discussing whether to purchase an SUV. Examples of arguments in this context are *the car is too expensive* and related ones such as *good loan programs are available* and *interest rates are very high*.

Data collected within the scope of the empirical study was used to evaluate different combinations of the four above-mentioned approaches to argument prediction. Argumentation theory was not found to be a useful prediction approach, since there seem to be other aspects of argumentation beyond justification that should be taken into account. Of the argument prediction approaches that were compared, machine learning (in combination with corresponding relevance heuristics) was shown to have the best performance in terms of the evaluation criteria mentioned.

3.5. VizRec: Recommending Personalized Visualizations

This article by Mutlu et al. contributes to our understanding of methods of interacting with AAGs [Mutlu et al. 2016]. The authors focus on visualizations as a means for the AAG to convey information to users.

The authors describe and evaluate a range of methods in the VizRec system for recommendation of visualization types for users. Mutlu et al. argue that visualizations have a distinctive advantage in dealing with the information overload problem.

They point out that, because visualizations are based on an understanding of visual cognition, people tend to understand them. However, creating proper visualizations requires specific expertise concerning the domain and the underlying data. Mutlu et al. evaluate methods that programmatically produce an appropriate recommendation of a visualization in a particular context.

Mutlu et al. posit that users' perception of visualizations is inherently personal. Accordingly, their investigation focuses on different aspects of user preferences as they are applied to the problem of recommending personalized visualizations. They apply collaborative filtering on a multidimensional scale to estimate aspects of quality for visualizations. In a second approach, tag vectors describing visualizations are used to recommend potentially interesting visualizations based on content. Finally, Mutlu et al. describe a hybrid approach that combines information on what a visualization is about (tags) with information about how good it is (ratings). They present a discussion of the design principles behind their VizRec visual recommender system. They also describe its architecture and the data acquisition approach (with a crowd-sourced study), and they present an analysis of strategies for visualization recommendation.

This article is well aligned with this special issue's theme of methods for supporting interaction with AAGs. First, by simply recommending a particular visualization type to a user, the system can influence that user's perspective on the underlying data. Second, the advice-giving agent, which is the recommender in this case, requires a deep understanding of the end users' needs, beyond the typical standards and guidelines for a nonpersonalized visualization. Mutlu et al. explore this interaction for three categories of recommendation algorithms: collaborative, content-based, and hybrid filtering. Their crowd-sourced experiment examines variance in observed choices for multiple visualization types, and the subsequent discussion provides insight for the design of future systems that provide personalized advice about visualizations.

4. BRIEF SUMMARY OF CONTRIBUTIONS

This special issue covers three themes related to artificial advice givers: (a) human factors to consider when designing interactions with AAGs (e.g., over- and under-reliance, overestimation of capacity), (b) methods for supporting interaction with AAGs (e.g., natural language, visualization, and argumentation); and (c) considerations for evaluating AAGs.

Regarding human factors, the studies in this special issue have found that in some situations (e.g., using an avatar or more complex natural language), increasing the naturalness of the interaction caused overestimation of the system's capacity, which resulted in aborted interaction and other disadvantages [Knijnenburg and Willemsen 2016]. In other situations, more natural interaction such as vague speech was perceived more positively in spoken form than artificially produced speech [Clark et al. 2016]. The issue also confirms a number of factors that contribute to over- and under-reliance on advice, namely the cost of receiving advice, the reliability of the adviser, and the predictability of the environment [Sutherland et al. 2016].

With regard to methods for interaction with AAGs, the special issue contributes novel work in terms of natural language, argumentation, and adaptive visualization. Some decision-making processes require a longer dialogue and deliberation, which is also represented in this special issue as a study that helps us better understand how natural joint decision making relates to formal argumentation models [Rosenfeld and Kraus 2016]. While some advice may be presented verbally (in writing or speech), other advice may be better suited to visual representation in order to minimize cognitive overload [Mutlu et al. 2016]. In addition to domain and data-specific considerations, the way in which visual information is presented can be successfully and automatically tailored to individual human decision makers.

This special issue also describes evaluation criteria and methodologies suitable for evaluating AAGs, such as the digital game environment used by Sutherland et al. [2016] and methods for interpreting facial expressions and gestures in order to gauge the performance of AAGs [Clark et al. 2016].

The articles in this special issue have brought us closer to understanding how to improve the interaction between humans and artificial advice givers. We hope that you enjoy reading it as much as we have enjoyed editing it.

ACKNOWLEDGMENTS

We thank the former and current editors-in-chief of ACM TiiS, Anthony Jameson, Krzysztof Gajos, and Michelle Zhou, for letting this special issue become a reality and for their high commitment to this journal. Special thanks go to the authors for their strong contributions and to the anonymous reviewers for their support in maintaining the high standard of this special issue.

REFERENCES

Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. 2007. Open user profiles for adaptive news systems: Help or harm? In *Proceedings of the International Conference on World Wide Web*. ACM, 11–20.

Xavier Amatriain, Alejandro Jaimes, Nuria Oliver, and Josep M. Pujol. 2011. Data mining methods for recommender systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). Springer, Berlin, 39–71.

Katie Atkinson and Trevor Bench-Capon. 2007. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artif. Intell.* 171, 10 (2007), 855–874.

Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

Tara S. Behrend and Lori Foster Thompson. 2011. Similarity effects in online training: Effects with computerized trainer agents. *Comput. Hum. Behav.* 27, 3 (2011), 1201–1206.

Sharon Wraith Bennett and A. Carlisle Scott. 1985. *The Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley Publishing Company, Chapter 19, 363–370.

Timothy Bickmore and Justine Cassell. 2001. Relational agents: A model and implementation of building user trust. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 396–403.

Peter Brusilovsky, Elmar Schwarz, and Gerhard Weber. 1996. ELM-ART: An intelligent tutoring system on world wide web. In *Intelligent Tutoring Systems*. 261–269.

John M. Carroll and Jean McKendree. 1987. Interface design issues for advice-giving expert systems. *Commun. ACM* 30, 1 (1987), 14–32.

Federico Cerutti. 2011. Decision support through argumentation-based practical reasoning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Vol. 22. 2786–2787.

Federico Cerutti, Nava Tintarev, and Nir Oren. 2014. Formal arguments, preferences, and natural language interfaces to humans: An empirical evaluation. In *Proceedings of the European Conference on Artificial Intelligence*. 207–212.

Leigh Clark, Abdulmalik Ofemile, Svenja Adolphs, and Tom Rodden. 2016. A multimodal approach to assessing user experiences with agent helpers. *ACM Trans. Interact. Intell. Syst.* 6, 4 (2016), Article 29.

Andrew J. Cowell and Kay M. Stanney. 2005. Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *Int. J. Hum.-Comput. Stud.* 62, 2 (2005), 281–306.

Vania Dimitrova. 2003. STyLE-OLM: Interactive open learner modelling. *Int. J. Artif. Intell. Educ.* 17, 2 (2003), 35–78.

Alan Dix. 2016. Human-like computing and human-computer interaction. In *Human Centred Design for Intelligent Environments (HCD4IE) Workshop at the British Human Computer Interaction Conference*.

EPSRC. 2016. Report of human-like computing workshop. Engineering and Physical Sciences Research Council.

David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, and others. 2010. Building Watson: An overview of the DeepQA project. *AI Mag.* 31, 3 (2010), 59–79.

David A. Ferrucci. 2012. Introduction to "this is Watson." *IBM J. Res. Dev.* 56, 3.4 (2012), 1–1.

Adam D. Galinsky and Thomas Mussweiler. 2001. First offers as anchors: The role of perspective-taking and negotiator focus. *J. Person. Soc. Psychol.* 81, 4 (2001), 657.

Bryce Goodman and Seth Flaxman. 2016. European union regulations on algorithmic decision-making and a "right to explanation." In *Workshop on Human Interpretability in Machine Learning at the International Conference on Machine Learning*.

Stephen Hawking, Stuart Russell, Max Tegmark, and Frank Wilczek. 2014. Stephen hawking: "Transcendence looks at the implications of artificial intelligence—but are we taking AI seriously enough?" *The Independent* 2014, 05-01 (2014).

Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. 241–250.

Traci J. Hess, Mark A. Fuller, and John Mathew. 2005. Involvement and decision-making performance with a decision aid: The influence of social multimedia, gender, and playfulness. *J. Manag. Inform. Syst.* 22, 3 (2005), 15–54.

Daniel Kahneman. 2011. *Thinking, Fast and Slow*. Allen Lane, New York.

Byungkyu Kang, Nava Tintarev, Tobias Höllorer, and John O'Donovan. 2016. What am I not seeing? An interactive approach to social content discovery in microblogs. In *Proceedings of the International Conference on Social Informatics (SocInfo)*. 279–294.

Bart Knijnenburg and Martijn Willemsen. 2016. Inferring capabilities of intelligent agents from their external traits. *ACM Trans. Interact. Intell. Syst.* 6, 4 (2016), Article 28.

Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the International Conference on Intelligent User Interfaces*.

Belgin Mutlu, Eduardo Veas, and Christoph Trattner. 2016. VizRec: Recommending personalized visualizations. *ACM Trans. Interact. Intell. Syst.* 6, 4 (2016), Article 31.

Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 72–78.

Raymond S. Nickerson. 1976. On conversational interaction with computers. In *Proceedings of the Workshop on User-Oriented Design of Interactive Graphics Systems at ACM SIGGRAPH*. ACM, 101–113.

Donald A. Norman. 1986. Cognitive engineering. In *User Centered System Design: New Perspectives on Human-Computer Interaction*, Donald A. Norman and Stephen W. Draper (Eds.). Erlbaum, Hillsdale, NJ, 31–61.

Eli Pariser. 2011. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin, New York.

Henry Prakken and Giovanni Sartor. 1997. Argument-based extended logic programming with defeasible priorities. *J. Appl. Non-Classical Log.* 7, 1–2 (1997), 25–75.

Lingyun Qiu and Izak Benbasat. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *J. Manag. Inform. Syst.* 25, 4 (2009), 145–182.

Leo R. Quintanar, Charles R. Crowell, and Patrik M. Moskal. 1987. *Social Ergonomic and Stress Aspects of Work With Computers*. Amsterdam: Elsevier Publishers, Chapter The Interactive Computer as a Social Stimulus in Human-Computer Interactions.

Iyad Rahwan, Mohammed I. Madakkatel, Jean-François Bonnefon, Ruqiyabi N. Awan, and Sherief Abdallah. 2010. Behavioral experiments for assessing the abstract argumentation semantics of reinstatement. *Cogn. Sci.* 34, 8 (2010), 1483–1502.

B. Reeves and C. Nass. 1996. *The Media Equation: How People Treat Computers, Televisions, and New Media Like Real People and Places*. Cambridge University Press, New York, NY.

Ariel Rosenfeld and Sarit Kraus. 2016. Providing arguments in discussions on the basis of the prediction of human argumentative behavior. *ACM Trans. Interact. Intell. Syst.* 6 (2016), Article 30. Issue 4.

J. Ben Schafer, Joseph Konstan, and John Riedl. 1999. Recommender systems in e-commerce. In *Proceedings of the ACM Conference on Electronic Commerce*. ACM, New York, NY, 158–166.

James Schaffer, Prasanna Giridhar, Debra Jones, Tobias Höllerer, Tarek Abdelzaher, and John O'Donovan. 2015. Getting the message? A study of explanation interfaces for microblog data analysis. In *Proceedings of the International Conference on Intelligent User Interfaces*. New York, 345–356.

Ben Shneiderman. 1997. Direct manipulation versus agents: Paths to predictable, controllable and comprehensible interfaces. In *Software Agents*, Jeffrey M. Bradshaw (Ed.). AAAI Press, Menlo Park, CA, 97–106.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.

Herbert A. Simon. 1955. A behavioural model of choice. *Quart. J. Econ.* 69, 1 (1955), 99–118.

Derek Sleeman and John Seely Brown. 1985. Intelligent tutoring systems. *Artif. Intell.* 26, 2 (1985), 233–238.

Martin Stettinger, Alexander Felfernig, Gerhard Leitner, and Stefan Reiterer. 2015. Counteracting anchoring effects in group decision making. In *Proceedings of the International Conference on User Modeling, Adaptation, and Personalization*. 118–130.

Steve Sutherland, Casper Harteveld, and Michael Young. 2016. Effects of the advisor and environment on requesting and complying with automated advice. *ACM Trans. Interact. Intell. Syst.* 6, 4 (2016), Article 27.

Nava Tintarev, Byunkyu Kang, and John O'Donovan. 2015. Inspection mechanisms for community-based content discovery in microblogs. In *Proceeding of the Joint Workshop on Interfaces and Human Decision Making for Recommender Systems at the ACM Conference on Recommender Systems*. 21–28.

Nava Tintarev and Roman Kutlak. 2014. Explanations—making plans scrutable with argumentation and natural language generation. In *Proceedings of the International Conference on Intelligent User Interfaces (Demo Track)*. 29–32.

Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems: Methodological issues and empirical studies on the impact of personalization. *User Model. User-Adapt. Interact.* 22 (2012), 399–439.

Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook* (2nd ed.). Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, Berlin.

Nava Tintarev, Nir Oren, Roman Kutlak, Matt Green, Judith Masthoff, K. van Deemter, and W. Vasconcelos. 2013. SAsSy—scrutable autonomous systems. In *Do-Form: Enabling Domain Experts to use Formalised Reasoning, a Symposium at the (British) Society for the Study of Artificial Intelligence and the Simulation of Behaviour*.

Janet H. Walker, Lee Sproull, and R. Subramani. 1994. Using a human face in an interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 85–91.

Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, CA.

Received October 2016; accepted November 2016