Creating Socially Adaptive Electronic Partners

Interaction, Reasoning and Ethical Challenges

M. Birna van Riemsdijk*
Delft University of Technology
Delft, The Netherlands
m.b.vanriemsdijk@tudelft.nl

Catholijn M. Jonker
Delft University of Technology
Delft, The Netherlands
c.m.jonker@tudelft.nl

Victor Lesser
UMASS
Amherst, USA
lesser@cs.umass.edu

ABSTRACT

Technology for supporting people in their daily lives such as personal assistant agents and smart homes carry great potential for making our lives more connected, healthy, efficient and safe by executing tasks on our behalf and guiding our actions. We make two key observations: 1) supportive technology is *inherently social* in the sense that its support to a user is subject to norms from people in the user's social context (e.g., family members and caregivers), and 2) existing supportive technology is *rigid in its realization of this social nature* by hardwiring norms into the technology. This rigidity leads to violation of unsupported norms and inflexibility in dealing with violation of supported norms. In this paper we argue that supportive technology should be able to adapt to diverse and evolving norms of people in unforeseen circumstances, in order to better support people in their daily lives. We conceptualize this vision by proposing the novel concept of a Socially Adaptive Electronic Partner (SAEP), and outlining interaction, reasoning, and ethical challenges that need to be addressed to realize the creation of SAEPs. This requires techniques that span the areas of normative agents, human-agent teamwork, and ethics of AI, putting the multi-agent systems field in a unique position to do this.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents, multiagent systems*

Keywords

Norm-aware agents; Supportive technology; Ethics

1. INTRODUCTION

Technology for supporting people in their daily lives such as personal assistant agents, virtual coaches, location sharing systems, human support robots and smart homes carry great potential for making our lives more connected, healthy, efficient and safe by executing tasks on our behalf and guiding our actions. For example, geofencing[1] is a technology that uses GPS tracking to put a virtual fence around an area that a person (e.g., a child or dementia patient) should not leave. To support safety of the person being monitored, parents or caregivers should be notified immediately if the person does leave the area. Smart homes[2] are equipped with technology for remote monitoring of elderly to encourage independent living. To improve safety, an alarm should be sent to family or caregivers if the system detects irregularities. Virtual coaches are used in healthcare to complement a therapist's guidance. They should motivate and assist the patient to improve the outcome of the therapy [47].

We make two key observations. Firstly, supportive technology is *inherently social* in the sense that its support to a user is subject to norms from people in the user's social context. The purpose of the technology is to enhance the way in which a person's social network can provide support, and the way in which support is provided needs to take into account the user's role in the social networks that the user is a member of. For example, geofencing systems may be viewed as implementing two types of norms:

N1: According to person P1, person P0 should not leave area A.
N2: If P0 leaves A (i.e., violates *N1*) then notify P1 immediately.

Geofencing enhances the way in which P1 can provide support to P0, and because of P1's role in the life of P0 (e.g., parent or caregiver), P1 is the person to be notified if P0 leaves area A. Comparable norms are implemented by smart homes to support elderly. The support that virtual coaches provide should similarly take into account the social relation between therapist and patient.

Our second observation is that existing supportive technology is *rigid in its realization of its social nature* by hardwiring norms into the technology as built-in stimulus-response rules. This makes the software inflexible, which leads to 1) violation of unsupported norms and 2) inflexibility in dealing with (potential) violation of supported norms. Consequently, this inflexibility limits the support that the technology can provide and it comes with the risk of violating important values such as privacy and freedom. For example, the geofencing system only implements norms from P1, and does not support norms from others in the social context of P0 that may conflict with P1's norms. For example, a third person P2 may not want the system to share P0's location with P1 in certain circumstances (e.g., when preparing a birthday surprise for P1). Moreover, the system would provide better support if - instead of implementing *N2* as a rigid stimulus-response rule - it could decide to violate *N2* in certain situations, for example if the child P0 leaves area A because parent P2, who unexpectedly comes home early, decides to take P0 along with them to shop for groceries. Notifying the other parent P1 is not beneficial for the child's safety in this case and would cause unnecessary anxiety for P1.

*M. Birna van Riemsdijk was supported by the NWO Vidi project CoreSAEP.

[1] http://www.brickhousesecurity.com/category/gps+tracking/gps+geofencing.do

[2] http://sine.ni.com/cs/app/doc/p/id/cs-14844

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May, 4–8, 2015, Istanbul, Turkey.*

One may argue that these problems can easily be solved by hard-wiring additional norms as stimulus-response rules into the geofencing system. However, such an ad hoc approach does not address the underlying problem, which stems from the *dynamic nature* of social networks and norms: which specific norms are needed to allow a person's social network to provide support is highly context-dependent. It depends on people's capabilities (which may change over time as with dementia), on the values of people in the relevant social networks (which people may join and leave), and on how these values play out in different situations (what constitutes an unsafe situation may, e.g., depend on the neighbourhood, the activity and the people who are present). As argued by Kaminka in the context of social robots [31], deciding at design time which rules to implement and letting the system follow them blindly will narrow the scope of the interactions, will inhibit the system's ability to deal with unexpected situations, and will entail significant and expensive programming efforts to realize supportive behavior that is good enough for the task at hand.

In contrast to a design time solution, we argue for a fundamentally different approach to development of supportive technology: an approach in which norms are represented explicitly in the software, which enables them to be modified by people and through experiences of the system, as well as reasoned about at run-time. We conceptualize our vision by proposing the novel concept of a *Socially Adaptive Electronic Partner (*SAEP*)* that supports the daily activities of its owner. The key idea is that the set of norms that the SAEP has can be dynamically modified and extended by other people in the owner's social context who have appropriate rights. This open norm concept, in combination with a rich language for expressing norms and a *generic computational framework* for reasoning about how to act in the face of norms, will enable fundamentally more flexible behavior and therefore improved support to people.

We posit that creating SAEPs centers around answering three main questions:

- *Interaction:* how to shape a SAEP's interaction with people about norms?
- *Reasoning:* how can a SAEP reason about and learn how to act in the face of norms and their (potential) violation?
- *Ethical:* to what extent can SAEPs alleviate ethical concerns about the use of supportive technology?

The challenge that we pose to the autonomous agents and multi-agent systems community is development of generic theories, computational models and their empirical validation to answer these questions. The open system [22] in which a SAEP has to function pushes the need for more adaptability and the ability of the SAEP to modify its knowledge base and corresponding behavior than what has been done before in existing limited applications. In this paper we highlight a number of challenges that we feel need to be addressed to answer these questions, and hope that it will inspire others to add their own. We show how existing research from the areas of human-agent teamwork, normative agents, and ethics forms a solid foundation from which we can begin to address these challenges. With this paper we invite the community to join our mission of changing the way we think about supportive technology, and develop the techniques required to make this vision a reality.

2. INTERACTION CHALLENGES

Electronic partners (*ePartners*) and *personal assistant agents* to support users in various tasks have already been proposed in the literature (e.g., [38, 39, 46]). Such research focuses on the relation and interactions between *one* ePartner or personal assistant agent and its user. Interaction challenges that are addressed are for example determining when the ePartner should take an autonomous decision and when it should ask the user (adaptive automation, e.g., [20]), persuading the user to exhibit a particular behavior [26, 30], and user modelling and personalization [38]. In [4] Amir et al. present a vision of Care Augmenting Software PartnERs (CASPERs), in particular for health care, that support care coordination and aid in the communication of medical information to patients and their family. The coordination support of diverse and ever changing teams of caregivers portrays CASPER as the spider in the web. The concept of a SAEP as an ePartner that belongs to a specific owner while accounting for norms from the social context, complements these visions.

The authors of [39] already emphasize the importance in future work of adding values and norms from the social context to the relationship. Also Tambe et al. note as an important lesson they learned from deploying personal assistant agents in office environments that they "must follow the social norms of the human society within which the agents function" [46]. Interaction-Oriented Software Engineering emphasizes the use of norms and social commitments in the design phase of socio-technical systems [15].

While these studies point to the importance of accounting for norms when developing personal assistant agents, *how to shape interaction about norms* is an open question. At the heart of the concept of a SAEP is the idea that some people in the social context of the SAEP's owner can dynamically modify its norms. To realize this, interaction challenges need to be addressed:

- *Normative language design:* We need to understand what kind of norms the SAEP should support, i.e., we need to know which normative language to use. Although a plethora of normative models and languages has been developed in the multi-agent systems community [7], determining a suitable combination of normative features for a specific application domain and developing an appropriate domain ontology to instantiate the abstract normative language, is not a trivial matter. This requires understanding who the kind of people that the SAEP should support are, what their lives are like, and what kind of support they need. Moreover, to facilitate reuse across similar domains, we need to understand what 'similar' means in this context, i.e., how to characterize domains with respect to the use of SAEPs.
- *Interaction about norm violation and modification:* Handling norm violations may involve interaction with the SAEP's owner and others in the social context so that they can easily modify the SAEP's norms to conform to their values rather than the developers'. Explicit representation of norms in the SAEP enables such interaction in a principled way rather than through ad hoc mechanisms. Also, by observation and interaction with people, a SAEP can learn and suggest new or modified norms to reflect the actual or intended user behavior. This open norm concept is key to making the SAEP a true partner.

Methods and techniques for addressing these challenges will need to ensure that interaction about norms and the SAEP's behavior is understandable for the target group of users. This involves finding a suitable balance between understandability and expressivity of the normative language as well as sophistication of the reasoning techniques. As noted in [46] in the context of office personal assistant agents, even if the learning was effective and lead to "better" behavior, the fact that the assistant "did not necessarily behave the same way each day could be disconcerting to the users". The interaction design should make it easy for a user to add or modify the collection of norms, understand when violation has occurred or will occur and why the SAEP feels that this is the case. People should be able to easily provide guidance to the SAEP about how to handle the violation, or decide how to handle it themselves by choosing from a list of options provided by the SAEP.

Research on human-computer interaction and human-agent teamwork can provide a starting point for addressing these challenges. For example, Kayal et al. [32] propose a grounded model to describe the social context of family life for the purpose of developing a normative framework to govern the behavior of ePartners for children. Techniques from explainable AI that allow agents to explain their decision making to people [28] may be used as a basis for interaction about normative decisions. User feedback may be used in learning how to respond to norm violation, similar to [3] in which an approach to social adaptation is proposed that allows users to provide feedback to system functioning. Research on preference elicitation [42] can provide a starting point for developing norm elicitation techniques.

The challenge that we pose to the multi-agent systems community is to investigate the fundaments underlying interaction about norms and to build theories and methods that provide guidance to developers on how to shape this interaction. We feel that addressing this challenge is a crucial part of realizing the vision of SAEPs that we put forward, complementing research on formal models for normative multi-agent systems.

3. REASONING CHALLENGES

Research on normative multi-agent systems has yielded a wide variety of normative models [7, 6]. That work focuses on the specification of a normative system from the perspective of the society or institution that imposes the norms with the aim of regulating a collection of autonomous agents. It provides languages for specification of norms, semantics that define when norms are fulfilled and violated with techniques for monitoring this, and notions of normative consequence and normative conflict. Realizing SAEPs also requires computational reasoning techniques that allow the SAEP to decide on what to do when norms are imposed from the social context.

Research on normative agents provides techniques for agents to reason about norms [35]. They have largely been developed for rational self-interested agents that want to comply with norms if benefits from complying outweigh expected costs of violation (e.g., induced by sanctions) or if complying is instrumental in achieving the agent's goals [16, 11, 24, 33, 34, 41, 1, 17]. Internalized emotions for reasoning about whether to comply with norms have also been incorporated in the framework of [17]. Normative agents for simulating human societies [10] have been developed to capture aspects of human normative decision making such as the effect of imitation and norm recognition on norm emergence [5]. Frameworks for generating goals from conflicting adopted norms have been proposed in [14, 12, 18]. Social choice theory [23] and automated negotiation [29] may be used to support people in forming a social agreement about norms that can be adopted by SAEPs.

We highlight two additional challenges that have received comparatively little attention, and that are particularly relevant in the context of SAEPs:

- *Reasoning about norm compliance:* For a SAEP it is especially important not only to determine whether it *wants* to comply, but also to determine, before it adopts a norm, whether it *can* comply with it. This is important from the perspective of transparency. Since a SAEP may not always be able to guarantee norm compliance, for example if this depends on its owner's behavior, it should also be able to reason about how to handle norm violations.

If a new norm is specified by someone in the social context and the SAEP cannot satisfy it while continuing to satisfy existing norms, it cannot perform its supportive function appropriately. Depending on the purpose of the norm, this may at best be inconvenient and at worst put its owner at serious risk. In either case, the SAEP should be transparent about the extent to which it can guarantee compliance, so that users are aware of potential risks. Determining whether the SAEP can comply with a new norm depends on the SAEP's execution mechanism for integrating norms and agent deliberation, on whether the new norm conflicts with existing norms, and on the capabilities of the SAEP in the current context [48, 49]. The required reasoning and execution mechanisms furthermore depend on the supported normative language, and they should be computationally efficient so that they can be performed at run-time. If the SAEP determines that it currently has no means to satisfy the new norm, it can ask the creator of the norm for a solution. Developing a generic framework for reasoning and interacting about norm compliance requires a formal analysis of these aspects and their interplay.

The SAEP should also be able to reason about norm violation. It should take action as soon as possible (anticipating future norm violations) rather than waiting for the violation to actually occur. The ability to anticipate norm violation allows the SAEP more action options and time either to avoid the violation or to resolve the violation in a more satisfying manner, and allows people to have more time and options to take remedial actions outside the scope of actions available to the SAEP to resolve the violation. To enable a flexible and principled way of handling norm violation, we envision the specification of norms about how to handle violation similar to norms about how to act. The SAEP may take into account the underlying values for which norms were created, for example to ensure safety, in deciding how to handle them. Considering the example from Section 1 where parent P2 unexpectedly comes home early and takes the child to fetch groceries, the SAEP can reason that while a norm is violated (the child leaves area A), safety is not compromised because another trusted person is with the child. This may influence how the SAEP handles the norm to warn P1 if the child leaves area A. It allows the SAEP to act according to the core concerns underlying the norm, instead of to the letter of the norm. The SAEP's ability to reason about norm compliance instead of blindly following norms, enables such flexible behavior.

- *Learning (about) norms:* Allowing people to explicitly specify norms for a SAEP provides them the ability to regulate how it provides support. We envisage that this can be enhanced by enabling the SAEP to learn norms and how to act on them, for example by observing how its owner and people in the social context behave. What type of norms can and should be learned is a topic of research.

For example, norms that increase usability of the SAEP may be learned, while norms that may impact safety might have to be explicitly specified by people. As noted in Section 2, effective interaction of the SAEP with people about newly learned norms has to be studied. The SAEP may also learn how to optimize its adaptation to norms if these permit multiple options, in order to, for example, improve user satisfaction or efficiency. Finally, to allow the SAEP to reason about its capabilities in the current context, it will have to learn a capability model by acting in context. For example, intermittent network connection may compromise SAEP|'s ability to warn trustees in case of problems.

Formal methods such as used in [48] are an important means for developing reasoning techniques that allow the SAEP to determine whether it can comply with norms. To allow the SAEP to reason about norms and take appropriate action in the current context, it needs to constantly monitor the environment to understand whether the context has changed and then assess in this new context whether there will be or already have been norm violations. Existing research on context-aware systems provides sensing and analysis techniques for recognizing physical and social context [44]. Monitoring may be imperfect which may require norm modification to ensure it can be monitored effectively [2]. Techniques for recognizing norms in agent societies [43] and plan recognition [45] research may be used as a basis for learning norms. Contrary-to-duty norms [27] and conflict resolution techniques [50] are relevant for reasoning about norm violation. We need to understand what type of learning technology would be most appropriate for SAEPs. Statistical learning technologies such as reinforcement learning that are based a large corpus of examples generated through experimentation may not be appropriate for SAEPs. However, there is a corpus of work on learning from a few examples (e.g., [51, 37]) that may be used. A key question in applying these ideas would be how much domain knowledge (possibly provided by norms) would be required to use these small example learning approaches.

These examples of existing research on normative multi-agent systems and agent reasoning and learning illustrate that there is a solid foundation for developing normative reasoning techniques for SAEPs. At the same time, the specific properties of SAEPs as electronic partners that support people in their daily lives pose specific challenges for the type and level of social adaptivity that is required. To ensure that the level of trust that people place in a SAEP is in line with its capabilities, and to allow people to remain in control of how the SAEP performs its supportive function, we feel that reasoning techniques should allow the SAEP to be *transparent* about its adaptive capabilities and decision making. Moreover, the *open and real-world context* in which SAEPs have to function requires development of generic reasoning techniques that can be used in a computational framework and that are robust in handling unexpected situations. We challenge the multi-agent systems community to develop normative reasoning techniques to facilitate this.

4. ETHICAL CHALLENGES

Supportive technology can be viewed as a type of pervasive and context-aware system. The increasing pervasiveness of mobile devices and sensor technology has sparked a growth in research on such systems and companies developing context-aware applications. These applications are becoming very popular, exemplified by geofencing applications such as the 'Find My Friends' smartphone app and the Life360[3] family locator app with 5-10 million installs.

However, research in value-sensitive design and ethics of technology has shown that the use of such pervasive applications carries the *risk of violating values* such as freedom, responsibility, and privacy [13, 19, 21, 25, 36, 40]. We conjecture that the interaction and computational reasoning techniques that we envision SAEPs to be endowed with, will provide a step towards addressing this problem by enabling the creation of pervasive applications that can adapt their behavior to people's norms (viewed as concretized values) at run time. We posit that the following ethical challenges have to be addressed:

- *Enabling ethical use:* The flexibility of SAEPs in adapting to norms should ensure norm adherence for norms that are relevant for a specific context, and at the same time SAEPs should be able to decide to violate norms if this contributes to a higher value. For example, in case of an accident the SAEP can decide to release medical data of its user to a potential helper, even if that helper has no medical certificate.

There is some evidence in recent ethics literature that such flexible systems may be able to alleviate ethical concerns. In particular, Nissenbaum proposes the framework of contextual integrity for analyzing the extent to which technology violates privacy [40], and states that "information ought to be distributed and protected according to norms governing distinct social contexts" and "the ideal is flexibility in how [systems] are adapted to particular [social] contexts so the flow of information may be tailored according to [...] norms". Arkin discusses challenges for ethical autonomy in the context of warfare [8, 9], which are in line with the challenges described in this article, in particular concerning the need of run-time overriding ethical control (i.e., norms).

- *Preventing unethical use:* Preventing unethical use means that SAEPs can assure that no-one can abuse the system. This can imply that the system prevents the user and people in the social circles of the user from deploying the system in a certain way if this would violate a higher-level moral norm. Unethical use of the geofencing system would be e.g., for school to find out if a child was skipping classes.

Preventing unethical use requires mechanisms to ensure that the system provides responses within rigorously defined ethical boundaries [8, 9]. As we discuss in this article, the challenge here is that ethical boundaries are highly contextual and thus difficult to define at design time. Allowing dynamic modification of norms by people may also yield more opportunities for abuse.

These ethical challenges will probably never be solved completely as the parallel to famous philosophical problems would testify: if we cannot solve a philosophical problem while relaxing on our couch, then how would we expect the system to solve it under pressure while in function? However, we can make progress by changing the attitudes of the people involved and bringing the discussion to a level where people might respect each other's point of view better.

5. CONCLUSION

In this paper we have put forward a vision of SAEPs as socially adaptive partners of people. We have argued that an important component of this vision is that a SAEP needs to have an explicit representation of norms from people in its social context. The SAEP should be able to reason about these norms and their implications for how the SAEP should act in changing social contexts and to anticipate norm violations given current behavioral trajectories. Further, these norms should be able to be dynamically modified by users with appropriate authority and to represent meta-knowledge about how to respond to potential norm violations, taking into account ethical considerations. We have argued that a key design criterion for SAEPs is that it is easy for an authorized user to modify norms to respond to new, unexpected social contexts, and that the SAEP's interaction and reasoning techniques should allow it to be transparent to people about its capabilities and decisions. Finally, we have addressed the important ethical challenges in realizing this vision and argue that the explicit and dynamic reasoning about norms that we have suggested will help in addressing some of the important ethical considerations in building a SAEP.

[3] https://www.life360.com

REFERENCES

[1] N. Alechina, M. Dastani, and B. Logan. Programming norm-aware agents. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'12)*, pages 1057–1064. IFAAMAS, 2012.

[2] N. Alechina, M. Dastani, and B. Logan. Norm approximation for imperfect monitors. In *Proceedings of the fourteenth international conference on Autonomous agents and multi-agent systems (AAMAS'14)*, pages 117–124. IFAAMAS, 2014.

[3] R. Ali, C. Solis, I. Omoronyia, M. Salehie, and B. Nuseibeh. Social adaptation: When software gives users a voice. In *7th International Conference Evaluation of Novel Approaches to Software Engineering (ENASE'07)*, 2012.

[4] O. Amir, B. J. Grosz, E. Law, and R. Stern. Collaborative health care plan support. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems (AAMAS'13)*, pages 793–796. IFAAMAS, 2013.

[5] G. Andrighetto, M. Campenni, F. Cecconi, and R. Conte. The complex loop of norm emergence: A simulation model. In *Simulating Interacting Agents and Social Phenomena: The second world congress*, volume 7 of *Agent-Based Social Systems*, pages 19–35. Springer, 2010.

[6] G. Andrighetto and C. Castelfranchi. Part III: Norms. In S. Ossowski, editor, *Agreement Technologies*, volume 8 of *Law, Governance and Technology Series*, pages 169–249. Springer, 2013.

[7] G. Andrighetto, G. Governatori, P. Noriega, and L. W. N. van der Torre, editors. *Normative Multi-Agent Systems*, volume 4 of *Dagstuhl Follow-Ups*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2013.

[8] R. C. Arkin. The case for ethical autonomy in unmanned systems. *Journal of Military Ethics*, 9(4):332–341, 2010.

[9] R. C. Arkin and P. D. Ulam. An ethical adaptor: Behavioral modification derived from moral emotions. Technical Report GIT-GVU-09-04, Georgia Institute of Technology. College of Computing, 2009.

[10] T. Balke, S. Cranefield, G. Di Tosto, S. Mahmoud, M. Paolucci, B. T. R. Savarimuthu, and H. Verhagen. Simulation and NorMAS. In G. Andrighetto, G. Governatori, P. Noriega, and L. W. N. van der Torre, editors, *Normative Multi-Agent Systems*, volume 4 of *Dagstuhl Follow-Ups*, pages 171–189. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2013.

[11] G. Boella and R. Damiano. An architecture for normative reactive agents. In *Proceedings of the 5th Pacific Rim International Workshop on Intelligent Agents and Multi-Agent Systems (PRIMA'02)*, volume 2413 of *LNCS*, pages 1–17, 2002.

[12] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.

[13] K. Caine, C. Zimmerman, Z. Schall-Zimmerman, W. Hazlewood, L. Camp, K. Connelly, L. Huber, and K. Shankar. Digiswitch: A device to allow older adults to monitor and direct the collection and transmission of health information collected at home. *Journal of Medical Systems*, 35(5):1181–1195, 2011.

[14] C. Castelfranchi, F. Dignum, C. Jonker, and J. Treur. Deliberative normative agents: Principles and architecture. In *6th International Workshop on Intelligent Agents VI, Agent Theories, Architectures, and Languages (ATAL'99)*, volume 1757 of *LNCS*, pages 364–378. Springer, 2000.

[15] A. K. Chopra and M. P. Singh. Interaction-oriented software engineering: Concepts and principles. *arXiv:1211.4123*, 2012.

[16] R. Conte, C. Castelfranchi, and F. Dignum. Autonomous norm acceptance. In *Proceedings of the 5th International Workshop on Agent Theories, Architectures, and Languages (ATAL'98)*, volume 1555 of *LNCS*, pages 99–112. Springer, 1999.

[17] N. Criado, E. Argente, P. Noriega, and V. Botti. Human-inspired model for norm compliance decision making. *Information Sciences*, 245:218–239, 2013.

[18] N. Criado, E. Argente, P. Noriega, and V. J. Botti. Towards a normative bdi architecture for norm compliance. In *Proceedings of 11th International Workshop on Coordination, Organization, Institutions and Norms in Multi-Agent Systems (COIN at MALLOW2010)*, 2010.

[19] A. Czeskis, I. Dermendjieva, H. Yapit, A. Borning, B. Friedman, B. Gill, and T. Kohno. Parenting from the pocket: value tensions and technical directions for secure and private parent-teen mobile safety. In *Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS'10)*, pages 15:1–15:15. ACM, 2010.

[20] T. de Greef, H. Arciszeski, and M. Neerincx. Adaptive automation based on an object-oriented task model: Implementation and evaluation in a realistic c2 environment. *Journal of Cognitive Engineering and Decision Making*, 31:152–182, 2010.

[21] C. Detweiler, A. Huldtgren, and N. Guldemond. Take a stance! designing for healthy consumtion. In *Proceedings of the CHI 2013 Workshop on Personal Informatics in the Wild: Hacking Habits for Health & Happiness*, 2013.

[22] F. Dignum, V. Dignum, J. Thangarajah, L. Padgham, and M. Winikoff. Open agent systems??? In *Proceedings of the 8th International Workshop on Agent-Oriented Software Engineering (AOSE'07)*, volume 4951 of *LNCS*, pages 73–87. Springer, 2008.

[23] U. Endriss. Social choice theory as a foundation for multiagent systems. In *Proceedings of the 12th German Conference on Multiagent System Technologies (MATES-2014)*, volume 8732 of *LNAI*, pages 1–6. Springer-Verlag, September 2014. Extended abstract corresponding to an invited talk.

[24] M. S. Fagundes, H. Billhardt, and S. Ossowski. Normative reasoning with an adaptive self-interested agent model based on Markov Decision Processes. In *Proceedings of the 12th Ibero-American Conference on AI (IBERAMIA'10)*, volume 6433 of *LNCS*, pages 274–283. Springer, 2010.

[25] J. Fahlquist. Responsibility and privacy – ethical aspects of using GPS to track children. *Children & Society*, 2013.

[26] B. Fogg. *Persuasive Technology, Using computers to change what we think and do*. Morgan Kaufmann, San Fransico, 2003.

[27] J. Hansen, G. Pigozzi, and L. van der Torre. Ten philosophical problems in deontic logic. In G. Boella, L. van der Torre, and H. Verhagen, editors, *Normative Multi-agent Systems*, number 07122 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

[28] M. Harbers. *Explaining Agent Behavior in Virtual Training*. PhD thesis, 2011.

[29] K. V. Hindriks and C. M. Jonker. Creating human-machine synergy in negotiation support systems: Towards the pocket negotiator. In *HuCom'08*, 2008.

[30] C. Horsch, W.-P. Brinkman, R. Eijk, and M. Neerincx. Towards the usage of persuasive strategies in a virtual sleep coach. In *Proceedings of UKHCI 2012 Workshop on People, Computers and Psychotherapy*, 2012.

[31] G. A. Kaminka. Curing robot autism: A challenge. In *Proceedings of the twelfth international joint conference on autonomous agents and multiagent systems (AAMAS'13)*, pages 801–804, 2013.

[32] A. Kayal, W.-P. Brinkman, R. Gouman, M. A. Neerincx, and M. B. van Riemsdijk. A value-centric model to ground norms and requirements for epartners of children. In *Coordination, Organizations, Institutions, and Norms in Agent Systems IX (COIN'13)*, volume 8386 of *LNCS*, pages 329–345. Springer, 2014.

[33] F. López y López, M. Luck, and M. d'Inverno. Constraining autonomy through norms. In *The First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, pages 674–681. ACM, 2002.

[34] F. López y López, M. Luck, and M. d'Inverno. Normative agent reasoning in dynamic societies. In *The Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'04)*, pages 732–739. IEEE Computer Society, 2004.

[35] M. Luck, S. Mahmoud, F. Meneguzzi, M. Kollingbaum, and T. J. Norman. Normative agents. In S. Ossowski, editor, *Agreement Technologies*, volume 8 of *Law, Governance and Technology Series*, pages 209–220. Springer, 2013.

[36] K. Michael, A. McNamee, and M. Michael. The emerging ethics of humancentric GPS tracking and monitoring. In *Proceedings of the International Conference on Mobile Business (ICMB'06)*. IEEE Computer Society, 2006.

[37] S. Muggleton. Inductive logic programming. *New Generation Computing*, 8(4):295–318, 1991.

[38] K. Myers and N. Yorke-Smith. Proactivity in an intentionally helpful personal assistive agent. In *Proceedings of AAAI 2007 Spring Symposium on Intentions in Intelligent Systems*, 2007.

[39] M. Neerincx, M. Neerincx, and T. Grant. Evolution of electronic partners: Human-automation operations and epartners during planetary missions. *Journal of Cosmology*, 12:3825–3833, 2010.

[40] H. Nissenbaum. *Privacy in Context: Technology, Policy and the Integrity of Social Life*. Stanford University Press, Stanford, California, 2010.

[41] S. Panagiotidi and J. Vázquez-Salceda. Norm-aware planning: Semantics and implementation. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Volume 03*, pages 33–36. IEEE, 2011.

[42] A. Pommeranz, J. Broekens, P. Wiggers, W.-P. Brinkman, and C. M. Jonker. Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation. *User Modeling and User-Adapted Interaction*, 22(4-5):357–397, 2012.

[43] B. T. R. Savarimuthu, S. Cranefield, M. Purvis, and M. Purvis. A data mining approach to identify obligation norms in agent societies. In *Proceedings of the 6th international conference on Agents and data mining interaction (ADMI'10)*, pages 43–58. Springer, 2010.

[44] D. Schuster, A. Rosi, M. Mamei, T. Springer, M. Endler, and F. Zambonelli. Pervasive social context - taxonomy and survey. *ACM Transactions on Intelligent Systems and Technology*, 9(4):1–22, 2012.

[45] G. Sukthankar, C. Geib, H. H. Bui, D. Pynadath, and R. Goldman, editors. *Plan, Activity and Intent Recognition: Theory and Practice*. Morgan Kaufman, 2014.

[46] M. Tambe, E. Bowring, J. P. Pearce, P. Varakantham, P. Scerri, and D. V. Pynadeth. Electric elves: What went wrong and why. *AI magazine*, 29.

[47] M. Tielman, W.-P. Brinkman, and M. A. Neerincx. Design guidelines for a virtual coach for post-traumatic stress disorder patients. In T. Bickmore, S. Marsella, and C. Sidner, editors, *Intelligent Virtual Agents*, volume 8637 of *LNCS*, pages 434–437. Springer, 2014.

[48] M. B. van Riemsdijk, L. Dennis, M. Fisher, and K. V. Hindriks. Agent reasoning for norm compliance: a semantic approach. In *Proceedings of the twelfth international joint conference on autonomous agents and multiagent systems (AAMAS'13)*, pages 499–506. IFAAMAS, 2013.

[49] M. B. van Riemsdijk, L. Dennis, M. Fisher, and K. V. Hindriks. A semantic framework for socially adaptive agents: Towards strong norm compliance. In *Proceedings of the fourteenth international joint conference on autonomous agents and multiagent systems (AAMAS'15)*. IFAAMAS, 2015. This volume.

[50] W. W. Vasconcelos, M. J. Kollingbaum, and T. J. Norman. Normative conflict resolution in multi-agent systems. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 19(2):124–152, 2009.

[51] X. Zhang and et al.. An ensemble architecture for learning complex problem-solving techniques from demonstration. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4):1–38, 2012.