Affective Computing

Affect is a core aspect of intelligence. Drives and emotions, such as excitement and depression, are used to coordinate action throughout intelligent life, even in species that lack a nervous system. Emotions are one mechanism that humans evolved to accomplish what needs to be done in the time available with the information at hand—to satisfice. Emotions are not an impediment to rationality; arguably they are integral to rationality in humans. Humans create and respond to both positive and negative emotional influence as they coordinate their actions with other individuals to create societies. Autonomous and intelligent systems (A/IS) are being designed to simulate emotions in their interactions with humans in ways that will alter our societies.

A/IS should be used to help humanity to the greatest extent possible in as many contexts as are appropriate. While A/IS have tremendous potential to effect positive change, there is also potential that artifacts used in society could cause harm either by amplifying, altering, or even dampening human emotional experience. Even rudimentary versions of synthetic emotions, such as those already in use within nudging systems, have already altered the perception of A/IS by the general public and public policy makers.

This chapter of *Ethically Aligned Design* addresses issues related to emotions and emotion-like control in interactions between humans and design of A/IS. We have put forward recommendations on a variety of topics: considering how affect varies across human cultures; the particular problems of artifacts designed for caring and private relationships; considerations of how intelligent artifacts may be used for "nudging"; how systems can support human flourishing; and appropriate policy interventions for artifacts designed with inbuilt affective systems.

Document Sections

Section 1—Systems Across Cultures

Issue: Should affective systems interact using the norms for verbal and nonverbal communication consistent with the norms of the society in which they are embedded?

Background

Individuals around the world express intentions differently, including the ways that they make eye contact, use gestures, or interpret silence. These particularities are part of an individual's and a society's culture and are incorporated into their affective systems in order to convey the intended message. To ensure that the emotional systems of autonomous and intelligent systems foster effective communication within a specific culture, an understanding of the norms/values of the community where the affective system will be deployed is essential.

Recommendations

1. A well-designed affective system will have a set of essential norms, specific to its intended cultural context of use, in its knowledge base. Research has shown that A/IS technologies can use at least five types of cues to simulate social interactions.
2. These include: physical cues such as simulated facial expressions, psychological cues such as simulated humor or other emotions, use of language, use of social dynamics like taking turns, and through social roles such as acting as a tutor or medical advisor. Further examples are listed below:

    a. Well-designed affective systems will use language with affective content carefully and within the contemporaneous expectations of the culture. An example is small talk. Although small talk is useful for establishing a friendly rapport in many communities, some communities see people that use small talk as insincere and hypocritical. Other cultures may consider people that do not use small talk as unfriendly, uncooperative, rude, arrogant, or ignorant. Additionally, speaking with proper vocabulary, grammar, and sentence structure may contrast with the typical informal interactions between individuals. For example, the latest trend, TV show, or other media may significantly influence what is viewed as appropriate vocabulary and interaction style.

    b. Well-designed affective systems will recognize that the amount of personal space (proxemics) given by individuals in an important part of culturally specific

Affective Computing

human interaction. People from varying cultures maintain, often unknowingly, different spatial distances between themselves to establish smooth communication. Crossing these limits may require explicit or implicit consent, which A/IS must learn to negotiate to avoid transmitting unintended messages.

c. Eye contact is an essential component for culturally sensitive social interaction. For some interactions, direct eye contact is needed but for others it is not essential and may even generate misunderstandings. It is important that A/IS be equipped to recognize the role of eye contact in the development of emotional interaction.

d. Hand gestures and other non-verbal communication are very important for social interaction. Communicative gestures are culturally specific and thus should be used with caution in cross-cultural situations. The specificity of physical communication techniques must be acknowledged in the design of functional affective systems. For instance, although a "thumbs-up" sign is commonly used to indicate approval, in some countries this gesture can be considered an insult.

e. Humans use facial expressions to detect emotions and facilitate communication. Facial expressions may not be universal across cultures, however, and A/IS trained with a dataset from one culture may not be readily usable in another culture. Well-developed A/IS will be able to recognize, analyze, and even display facial expressions essential for culturally specific social interaction.

3. Engineers should consider the need for cross-cultural use of affective systems. Well-designed systems will have options innate to facilitate flexibility in cultural programming. Mechanisms to enable and disable culturally specific "add-ons" should be considered an essential part of A/IS development.

Further Resources

- G. Cotton, "Gestures to Avoid in Cross-Cultural Business: In Other Words, 'Keep Your Fingers to Yourself!'" *Huffington Post,* June 13, 2013.
- "Paralanguage Across Cultures," Sydney, Australia: Culture Plus Consulting, 2016.
- G. Cotton, *Say Anything to Anyone, Say Anything to Anyone, Anywhere: 5 Keys to Successful Cross-Cultural Communication*. Hoboken, NJ: Wiley, 2013.
- D. Elmer, *Cross-Cultural Connections: Stepping Out and Fitting In Around the World*. Westmont, IL: InterVarsity Press, 2002.
- B. J. Fogg, Persuasive Technology. *Ubiquity*, December 2, 2002.
- A. McStay, Emotional AI: The Rise of Empathic Media. London: Sage, 2018.
- M. Price, "Facial Expressions—Including Fear—May Not Be as Universal as We Thought." *Science,* October 17, 2016.

Affective Computing

Issue: It is presently unknown whether long-term interaction with affective artifacts that lack cultural sensitivity could alter human social interaction.

Background

Systems that do not have cultural knowledge incorporated into their knowledge base may or may not interact effectively with humans for whom emotion and culture are significant. Given that interaction with A/IS may affect individuals and societies, it is imperative that we carefully evaluate mechanisms to promote beneficial affective interaction between humans and A/IS. Humans often use mirroring in order to understand and develop their norms for behavior. Certain machine learning approaches also address improving A/IS interaction with humans through mirroring human behavior. Thus, we must remember that learning via mirroring can go in both directions and that interacting with machines has the potential to impact individuals' norms, as well as societal and cultural norms. If affective artifacts with enhanced, different, or absent cultural sensitivity interact with impressionable humans this could alter their responses to social and cultural cues and values. The potential for A/IS to exert cultural influence in powerful ways, at scale, is an area of substantial concern.

Recommendations

1. Collaborative research teams must research the effects of long-term interaction of people with affective systems. This should be done using multiple protocols, disciplinary approaches, and metrics to measure the modifications of habits, norms, and principles as well as careful evaluation of the downstream cultural and societal impacts.

2. Parties responsible for deploying affective systems into the lives of individuals or communities should be trained to detect the influence of A/IS, and to utilize mitigation techniques if A/IS effects appear to be harmful. It should always be possible to shut down harmful A/IS.

Further Resources

- T. Nishida and C. Faucher, Eds., Modelling Machine Emotions for Realizing Intelligence: Foundations and Applications. Berlin, Germany: Springer-Verlag, 2010.

- D. J. Pauleen, et al. "Cultural Bias in Information Systems Research and Practice: Are You Coming from the Same Place I Am?" *Communications of the Association for Information Systems*, vol. 17,)pp. 1–36, 2006. J. Bielby, "Comparative Philosophies in Intercultural Information Ethics." *Confluence: Online Journal of World Philosophies* 2, no. 1, pp. 233–253, 2015.

- J. Bryson, "Why Robot Nannies Probably Won't Do Much Psychological Damage." A commentary on an article by N. Sharkey

Affective Computing

and A. Sharkey, *The Crying Shame of Robot Nannies.* Interaction Studies, vol. 11, no. 2 pp. 161–190, July 2010.

- A. Sharkey, and N. Sharkey, "Children, the Elderly, and Interactive Robots." *IEEE Robotics & Automation Magazine,* vol.18, no. 1, pp. 32–38, March 2011.

Issue: When affective systems are deployed across cultures, they could adversely affect the cultural, social, or religious values of the community in which they interact.

Background

Some philosophers argue that there are no universal ethical principles and that ethical norms vary from society to society. Regardless of whether universalism or some form of ethical relativism is true, affective systems need to respect the values of the cultures within which they are embedded. How systems should effectively reflect the values of the designers or the users of affective systems is not a settled discussion. There is general agreement that developers of affective systems should acknowledge that the systems should reflect the values of those with whom the systems are interacting. There is a high likelihood that when spanning different groups, the values imbued by the developer will be different from the operator or customer of that affective system, and that end-user values should be actively considered. Differences between affective systems and societal values may generate conflict situations producing undesirable results, e.g., gestures or eye contact being misunderstood as rude or threatening. Thus, affective systems should adapt to reflect the values of the community and individuals where they will operate in order to avoid misunderstanding.

Recommendations

Assuming that well-designed affective systems have a minimum subset of configurable norms incorporated in their knowledge base:

1. Affective systems should have capabilities to identify differences between the values they are designed with and the differing values of those with whom the systems are interacting.

2. Where appropriate, affective systems will adapt accordingly over time to better fit the norms of their users. As societal values change, there needs to be a means to detect and accommodate such cultural change in affective systems.

3. Those actions undertaken by an affective system that are most likely to generate an emotional response should be designed to be easily changed in appropriate ways by the user without being easily hacked by actors with malicious intentions. Similar to how software today externalizes the language and vocabulary to be easily changeable based on location, affective systems should externalize some of the core aspects of their actions.

Further Resources

- J. Bielby, "Comparative Philosophies in Intercultural Information Ethics." *Confluence: Online Journal of World Philosophies* 2, no. 1, pp. 233–253, 2015.
- M. Velasquez, C. Andre, T. Shanks, and M. J. Meyer. "Ethical Relativism." Markkula Center for Applied Ethics, Santa Clara, CA: Santa Clara University, August 1, 1992.
- Culture reflects the moral values and ethical norms governing how people should behave and interact with others. "Ethics, an Overview." Boundless Management.
- T. Donaldson, "Values in Tension: Ethics Away from Home Away from Home." *Harvard Business Review.* September– October 1996.

Section 2—When Systems Care

Issue: Are moral and ethical boundaries crossed when the design of affective systems allows them to develop intimate relationships with their users?

Background

There are many robots in development or production designed to focus on intimate care of children, adults, and the elderly[2]. While robots capable of participating fully in intimate relationships are not currently available, the potential use of such robots routinely captures the attention of the media. It is important that professional communities, policy makers, and the general public participate in development of guidelines for appropriate use of A/IS in this area. Those guidelines should acknowledge fundamental human rights to highlight potential ethical benefits and risks that may emerge, if and when affective systems interact intimately with users.

Among the many areas of concern are the representation of care, embodiment of caring A/IS, and the sensitivity of data generated through intimate and caring relationships with A/IS. The literature suggests that there are some potential benefits to individuals and to society from the incorporation of caring A/IS, along with duly cautionary notes concerning the possibility that these systems could negatively impact human-to-human intimate relations[3].

Recommendations

As this technology develops, it is important to monitor research into the development of intimate relationships between A/IS and humans. Research should emphasize any technical and

Affective Computing

normative developments that reflect use of A/IS in positive and therapeutic ways while also creating appropriate safeguards to mitigate against uses that contribute to problematic individual or social relationships:

1. Intimate systems must not be designed or deployed in ways that contribute to stereotypes, gender or racial inequality, or the exacerbation of human misery.

2. Intimate systems must not be designed to explicitly engage in the psychological manipulation of the users of these systems unless the user is made aware they are being manipulated and consents to this behavior. Any manipulation should be governed through an opt-in system.

3. Caring A/IS should be designed to avoid contributing to user isolation from society.

4. Designers of affective robotics must publicly acknowledge, for example, within a notice associated with the product, that these systems can have side effects, such as interfering with the relationship dynamics between human partners, causing attachments between the user and the A/IS that are distinct from human partnership.

5. Commercially marketed A/IS for caring applications should not be presented to be a person in a legal sense, nor marketed as a person. Rather its artifactual, that is, authored, designed, and built deliberately, nature should always be made as transparent as possible, at least at point of sale and in available documentation, as noted in Section 4, Systems Supporting Human Potential.

6. Existing laws regarding personal imagery need to be reconsidered in light of caring A/IS. In addition to other ethical considerations, it will also be necessary to establish conformance with local laws and mores in the context of caring A/IS systems.

Further Resources

- M. Boden, J. Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, P. Newman, V. Parry, G. Pegman, T. Rodden and T. Sorrell, Principles of robotics: regulating robots in the real world. Connection Science, vol. 29, no. 2, pp. 124-129, April 2017.

- J. J. Bryson, M. E. Diamantis, and T. D. Grant, "Of, For, and By the People: The Legal Lacuna of Synthetic Persons." *Artificial Intelligence & Law*, vol. 25, no. 3, pp. 273–291, Sept. 2017.

- M. Scheutz, "The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots," in *Robot Ethics: The Ethical and Social Implications of Robotics,* P. Lin, K. Abney, and G. Bekey, Eds., pp. 205. Cambridge, MA: MIT Press, 2011.

Section 3– System Manipulation/ Nudging/Deception

Issue: Should affective systems be designed to nudge people for the user's personal benefit and/or for the benefit of others?

Background

Manipulation can be defined as an exercise of influence by one person or group, with the intention to attempt to control or modify the actions of another person or group. Thaler and Sunstein (2008) call the tactic of subtly modifying behavior a "nudge[4]". Nudging mainly operates through the affective elements of a human rational system. Making use of a nudge might be considered appropriate in situations like teaching children, treating drug dependency, and in some healthcare settings. While nudges can be deployed to encourage individuals to express behaviors that have community benefits, a nudge could have unanticipated consequences for people whose backgrounds were not well considered in the development of the nudging system[5]. Likewise, nudges may encourage behaviors with unanticipated long-term effects, whether positive or negative, for the individual and/or society. The effect of A/IS nudging a person, such as potentially eroding or encouraging individual liberty, or expressing behaviors that are for the benefit others, should be well characterized in the design of A/IS.

Recommendations

1. Systematic analyses are needed that examine the ethics and behavioral consequences of designing affective systems to nudge human beings prior to deployment.
2. The user should be empowered, through an explicit opt-in system and readily available, comprehensible information, to recognize different types of A/IS nudges, regardless of whether they seek to promote beneficial social manipulation or to enhance consumer acceptance of commercial goals. The user should be able to access and check facts behind the nudges and then make a conscious decision to accept or reject a nudge. Nudging systems must be transparent, with a clear chain of accountability that includes human agents: data logging is required so users can know how, why, and by whom they were nudged.
3. A/IS nudging must not become coercive and should always have an opt-in system policy with explicit consent.
4. Additional protections against unwanted nudging must be put in place for vulnerable populations, such as children, or when informed consent cannot be obtained. Protections against unwanted nudging should be encouraged when nudges alter long-term behavior or when consent alone may not be a sufficient safeguard against coercion or exploitation.

Affective Computing

5. Data gathered which could reveal an individual or groups' susceptibility to a nudge or their emotional reaction to a nudge should not be collected or distributed without opt-in consent, and should only be retained transparently, with access restrictions in compliance with the highest requirements of data privacy and law.

Further Resources

- R. Thaler, and C. R. Sunstein, *Nudge: Improving Decision about Health, Wealth and Happiness*, New Haven, CT: Yale University Press, 2008.
- L. Bovens, "The Ethics of Nudge," in *Preference change: Approaches from Philosophy, Economics and Psychology*, T. Grüne-Yanoff and S. O. Hansson, Eds., Berlin, Germany: Springer, 2008 pp. 207–219.
- S. D. Hunt and S. Vitell. "A General Theory of Marketing Ethics." Journal of Macromarketing, vol.6, no. 1, pp. 5-16, June 1986.
- A. McStay, Empathic Media and Advertising: Industry, Policy, Legal and Citizen Perspectives (the Case for Intimacy), Big Data & Society, pp. 1-11, December 2016.
- J. de Quintana Medina and P. Hermida Justo, "Not All Nudges Are Automatic: Freedom of Choice and Informative Nudges." Working paper presented to the European Consortium for Political Research, Joint Session of Workshops, 2016 Behavioral Change and Public Policy, Pisa, Italy, 2016.
- M. D. White, *The Manipulation of Choice. Ethics and Libertarian Paternalism.* New York: Palgrave Macmillan, 2013
- C.R. Sunstein, The Ethics of Influence: Government in the Age of Behavioral Science. New York: Cambridge, 2016
- M. Scheutz, "The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents? " *IEEE Transactions on Affective Computing,* vol. 3, no. 4,pp. 424–433, Sept. 2012.
- A. Grinbaum, R. Chatila, L. Devillers, J.-G. Ganascia, C. Tessier and M. Dauchet. "Ethics in Robotics Research: CERNA Recommendations," *IEEE Robotics and Automation Magazine,* vol. 24, no. 3,pp. 139–145, Sept. 2017.
- "Designing Moral Technologies: Theoretical, Practical, and Ethical Issues" Conference July 10–15, 2016, Monte Verità, Switzerland.

Affective Computing

Issue: Governmental entities may potentially use nudging strategies, for example to promote the performance of charitable acts. Does the practice of nudging for the benefit of society, including nudges by affective systems, raise ethical concerns?

Background

A few scholars have noted a potentially controversial practice of the future: allowing a robot or another affective system to nudge a user for the good of society[6]. For instance, if it is possible that a well-designed robot could effectively encourage humans to perform charitable acts, would it be ethically appropriate for the robot to do so? This design possibility illustrates just one behavioral outcome that a robot could potentially elicit from a user.

Given the persuasive power that an affective system may have over a user, ethical concerns related to nudging must be examined. This includes the significant potential for misuse.

Recommendations

1. As more and more computing devices subtly and overtly influence human behavior, it is important to draw attention to whether it is ethically appropriate to pursue this type of design pathway in the context of governmental actions.
2. There needs to be transparency regarding who the intended beneficiaries are, and whether any form of deception or manipulation is going to be used to accomplish the intended goal.

Further Resources

- J. Borenstein and R. Arkin, "Robotic Nudges: Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being Just Human Being." *Science and Engineering Ethics,* vol. 22, no. 1,pp. 31–46, Feb. 2016.
- J. Borenstein and R. Arkin. "Nudging for Good: Robots and the Ethical Appropriateness of Nurturing Empathy and Charitable Behavior ." *AI and Society,* vol. 32, no. 4, pp. 499–507, Nov. 2016.

Issue: Will A/IS nudging systems that are not fully relevant to the sociotechnical context in which they are operating cause behaviors with adverse unintended consequences?

Background

A well-designed nudging or suggestion system will have sophisticated enough technical capabilities for recognizing the context in which it is applying nudging actions. Assessment of the context requires perception of the scope or impact of the actions to be taken, the consequences of incorrectly or incompletely

Affective Computing

applied nudges, and acknowledgement of the uncertainties that may stem from long term consequences of a nudge[7].

Recommendations

1. Consideration should be given to the development of a system of technical licensing ("permits") or other certification from governments or non-governmental organizations (NGOs) that can aid users to understand the nudges from A/IS in their lives.

2. User autonomy is a key and essential consideration that must be taken into account when addressing whether affective systems should be permitted to nudge human beings.

3. Design features of an affective system that nudges human beings should include the ability to accurately distinguish between users, including detecting characteristics such as whether the user is an adult or a child.

4. Affective systems with nudging strategies should incorporate a design system of evaluation, monitoring, and control for unintended consequences.

Further Resources

- J. Borenstein and R. Arkin, "Robotic Nudges: Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being Just Human Being." *Science and Engineering Ethics,* vol. 22, no. 1, pp. 31–46, 2016.
- R. C. Arkin, M. Fujita, T. Takagi, and R. Hasegawa, "An Ethological and Emotional Basis for Human- Robot Interaction." *Robotics and Autonomous Systems,* vol. 42, no. 3–4 pp.191–201, March 2003.
- S. Omohundro "Autonomous Technology and the Greater Human Good." *Journal of Experimental and Theoretical Artificial Intelligence, vol.* 26, no. 3, pp. 303–315, 2014.

Issue: When, if ever, and under which circumstances, is deception performed by affective systems acceptable?

Background

Deception is commonplace in everyday human-human interaction. According to Kantian ethics, it is never ethically appropriate to lie, while utilitarian frameworks indicate that it can be acceptable when deception increases overall happiness. Given the diversity of views on ethics and the appropriateness of deception, should affective systems be designed to deceive? Does the non-consensual nature of deception restrict the use of A/IS in contexts in which deception may be required?

Affective Computing

Recommendations

It is necessary to develop recommendations regarding the acceptability of deception performed by A/IS, specifically with respect to when and under which circumstances, if any, it is appropriate.

1. In general, deception may be acceptable in an affective agent when it is used for the benefit of the person being deceived, not for the agent itself. For example, deception might be necessary in search and rescue operations or for elder- or child-care.

2. For deception to be used under any circumstance, a logical and reasonable justification must be provided by the designer, and this rationale should be certified by an external authority, such as a licensing body or regulatory agency.

Further Resources

- R. C. Arkin, "Robots That Need to Mislead: Biologically-inspired Machine Deception." *IEEE Intelligent Systems* 27, no. 6, pp. 60–75, 2012.

- J. Shim and R. C. Arkin, "Other-Oriented Robot Deception: How Can a Robot's Deceptive Feedback Help Humans in HRI?" *Eighth International Conference on Social Robotics (ICSR 2016)*, Kansas, MO., November 2016.

- J. Shim and R. C. Arkin, "The Benefits of Robot Deception in Search and Rescue: Computational Approach for Deceptive Action Selection via Case-based Reasoning." *2015 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR 2015)*, West Lafayette, IN, October 2015.

- J. Shim and R. C. Arkin, "A Taxonomy of Robot Deception and its Benefits in HRI." *Proceedings of IEEE Systems, Man and Cybernetics Conference,* Manchester England, October 2013.

Section 4—Systems Supporting Human Potential

Issue: Will extensive use of A/IS in society make our organizations more brittle by reducing human autonomy within organizations, and by replacing creative, affective, empathetic components of management chains?

Background

If human workers are replaced by A/IS, the possibility of corporations, governments, employees, and customers discovering new equilibria outside the scope of what the organizations' past leadership originally foresaw may be unduly limited. A lack of empathy based on shared needs, abilities, and disadvantages between organizations and customers causes disequilibria between the individuals and corporations and governments that exist to serve them. Opportunities for useful innovation may therefore be lost through automation. Collaboration requires enough commonality of collaborating intelligences to create empathy—the capacity to model the other's goals based on one's own.

According to scientists within several fields, autonomy is a psychological need. Without it, humans fail to thrive, create, and innovate. Ethically aligned design should support, not hinder, human autonomy or its expression.

Recommendations

1. It is important that human workers' interaction with other workers not always be intermediated by affective systems (or other technology) which may filter out autonomy, innovation, and communication.
2. Human points of contact should remain available to customers and other organizations when using A/IS.
3. Affective systems should be designed to support human autonomy, sense of competence, and meaningful relationships as these are necessary to support a flourishing life.
4. Even where A/IS are less expensive, more predictable, and easier to control than human employees, a core network of human employees should be maintained at every level of decision-making in order to ensure preservation of human autonomy, communication, and innovation.
5. Management and organizational theorists should consider appropriate use of affective and autonomous systems to enhance their business models and the efficacy of their workforce within the limits of the preservation of human autonomy.

Affective Computing

Further Resources

- J. J. Bryson, "Artificial Intelligence and Pro-Social Behavior," in *Collective Agency and Cooperation in Natural and Artificial Systems,* C. Misselhorn, Ed., pp. 281–306, Springer, 2015.
- D. Peters, R.A. Calvo, and R.M. Ryan, "Designing for Motivation, Engagement and Wellbeing in Digital Experience," *Frontiers in Psychology*– Human Media Interaction, vol. 9, pp 797, 2018.

Issue: Does the increased access to personal information about other members of our society, facilitated by A/IS, alter the human affective experience? Does this access potentially lead to a change in human autonomy?

Background

Theoretical biology tells us that we should expect increased communication—which A/IS facilitate—to increase group-level investment[8]. Extensive use of A/IS could change the expression of individual autonomy and in its place increase group-based identities. Examples of this sort of social alteration may include:

1. Changes in the scope of monitoring and control of children's lives by parents.
2. Decreased willingness to express opinions for fear of surveillance or long-term consequences of past expressions being used in changed temporal contexts.
3. Utilization of customers or other end users to perform basic corporate business processes such as data entry as a barter for lower prices or access, resulting potentially in reduced tax revenues.
4. Changes to the expression of individual autonomy could alter the diversity, creativity, and cohesiveness of a society. It may also alter perceptions of privacy and security, and social and legal liability for autonomous expressions.

Recommendations

1. Organizations, including governments, must put a high value on individuals' privacy and autonomy, including restricting the amount and age of data held about individuals specifically.
2. Education in all forms should encourage individuation, the preservation of autonomy, and knowledge of the appropriate uses and limits to A/IS[9].

Further Resources

- J. J. Bryson, "Artificial Intelligence and Pro-Social Behavior," in *Collective Agency and Cooperation in Natural and Artificial Systems,* C. Misselhorn, Ed., pp. 281–306, Springer, 2015.
- M. Cooke, "A Space of One's Own: Autonomy, Privacy, Liberty," *Philosophy & Social Criticism,* Vol. 25, no. 1, pp. 22–53, 1999.
- D. Peters, R.A. Calvo, R.M. Ryan, "Designing for Motivation, Engagement and Wellbeing in Digital Experience" *Frontiers in Psychology – Human Media Interaction*, vol. 9. pp 797, 2018.

- J. Roughgarden, M. Oishi and E. Akçay, "Reproductive Social Behavior: Cooperative Games to Replace Sexual Selection." *Science* 311, no. 5763, pp. 965–969, 2006.

Issue: Will use of A/IS adversely affect human psychological and emotional well-being in ways not otherwise foreseen?

Background

A/IS may be given unprecedented access to human culture and human spaces—both physical and intellectual. A/IS may communicate via natural language, may move with humanlike form, and may express humanlike identity, but they are not, and should not be regarded as, human. Incorporation of A/IS into daily life may affect human well-being in ways not yet anticipated. Incorporation of A/IS may alter patterns of trust and capability assessment between humans, and between humans and A/IS.

Recommendations

1. Vigilance and robust, interdisciplinary, on-going research on identifying situations where A/IS affect human well-being, both positively and negatively, is necessary. Evidence of correlations between the increased use of A/IS and positive or negative individual or social outcomes must be explored.
2. Design restrictions should be placed on the systems themselves to avoid machine decisions that may alter a person's life in unknown ways. Explanations should be available on demand in systems that may affect human well-being.

Further Resources

- K. Kamewari, M. Kato, T. Kanda, H. Ishiguro and K. Hiraki. "Six-and-a-Half-Month-Old Children Positively Attribute Goals to Human Action and to Humanoid-Robot Motion," *Cognitive Development,* vol. 20, no. 2, pp. 303–320, 2005.
- R.A. Calvo and D. Peters, Positive Computing: Technology for Wellbeing and Human Potential. Cambridge, MA: MIT Press, 2014.

Section 5—Systems with Synthetic Emotions

Issue: Will deployment of synthetic emotions into affective systems increase the accessibility of A/IS? Will increased accessibility prompt unforeseen patterns of identification with A/IS?

Background

Deliberately constructed emotions are designed to create empathy between humans and artifacts, which may be useful or even essential for human-A/IS collaboration. Synthetic emotions are essential for humans to collaborate with the A/IS but can also lead to failure to recognize that synthetic emotions can be compartmentalized and even entirely removed. Potential consequences for humans include different patterns of bonding, guilt, and trust, whether between the human and A/IS or between other humans. There is no coherent sense in which A/IS can be made to suffer emotional loss, because any such affect, even if possible, could be avoided at the stage of engineering, or reengineered. As such, it is not possible to allocate moral agency or responsibility in the senses that have been developed for human emotional bonding and thus sociality.

Recommendations

1. Commercially marketed A/IS should not be persons in a legal sense, nor marketed as persons. Rather their artifactual (authored, designed, and built deliberately) nature should always be made as transparent as possible, at least at point of sale and in available documentation.
2. Some systems will, due to their application, require opaqueness in some contexts, e.g., emotional therapy. Transparency in such systems should be available to inspection by responsible parties but may be withdrawn for operational needs.

Further Resources

- R. C. Arkin, P. Ulam and A. R. Wagner, "Moral Decision-making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust and Deception," *Proceedings of the IEEE,* vol. 100, no. 3, pp. 571–589, 2012.
- R. Arkin, M. Fujita, T. Takagi and R. Hasegawa. "An Ethological and Emotional Basis for Human-Robot Interaction," *Robotics and Autonomous Systems,* vol.42, no. 3–4, pp.191–201, 2003.
- R. C. Arkin, "Moving up the Food Chain: Motivation and Emotion in Behavior-based Robots," in *Who Needs Emotions: The Brain Meets the Robot*, J. Fellous and M. Arbib., Eds., New York: Oxford University Press, 2005.

Affective Computing

- M. Boden, J. Bryson, D. Caldwell, et al. "Principles of Robotics: Regulating Robots in the Real World." *Connection Science,* vol. 29, no. 2, pp. 124–129, 2017.
- J. J Bryson, M. E. Diamantis and T. D. Grant. "Of, For, and By the People: The Legal Lacuna of Synthetic Persons," *Artificial Intelligence & Law,* vol. 25, no. 3, pp. 273–291, Sept. 2017.
- J. Novikova, and L. Watts, "Towards Artificial Emotions to Assist Social Coordination in HRI," *International Journal of Social Robotics,* vol. 7, no. 1, pp. 77–88, 2015.
- M. Scheutz, "The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents?" *IEEE Transactions on Affective Computing,* vol. 3, no. 4, pp. 424–433, 2012.
- A. Sharkey and N. Sharkey. "Children, the Elderly, and Interactive Robots." *IEEE Robotics & Automation Magazine,* vol. 18, no. 1, pp. 32–38, 2011.

Thanks to the Contributors

We wish to acknowledge all of the people who contributed to this chapter.

The Affective Computing Committee

- **Ronald C. Arkin** (Founding Co-Chair) – Regents' Professor & Director of the Mobile Robot Laboratory; College of Computing Georgia Institute of Technology
- **Joanna J. Bryson** (Co-Chair) – Reader (Associate Professor), University of Bath, Intelligent Systems Research Group, Department of Computer Science
- **John P. Sullins** (Co-Chair) – Professor of Philosophy, Chair of the Center for Ethics Law and Society (CELS), Sonoma State University
- **Genevieve Bell** – Intel Senior Fellow Vice President, Corporate Strategy Office Corporate Sensing and Insights
- **Jason Borenstein** – Director of Graduate Research Ethics Programs, School of Public Policy and Office of Graduate Studies, Georgia Institute of Technology
- **Cynthia Breazeal** – Associate Professor of Media Arts and Sciences, MIT Media Lab; Founder & Chief Scientist of Jibo, Inc.
- **Joost Broekens** – Assistant Professor Affective Computing, Interactive Intelligence group; Department of Intelligent Systems, Delft University of Technology
- **Rafael Calvo** – Professor & ARC Future Fellow, School of Electrical and Information Engineering, The University of Sydney
- **Laurence Devillers** – Professor of Computer Sciences, University Paris Sorbonne, LIMSI-CNRS 'Affective and social dimensions in spoken interactions' - member of the French Commission on the Ethics of Research in Digital Sciences and Technologies (CERNA)
- **Jonathan Gratch** – Research Professor of Computer Science and Psychology, Director for Virtual Human Research, USC Institute for Creative Technologie
- **Mark Halverson** – Founder and CEO at Human Ecology Holdings and Precision Autonomy
- **John C. Havens** – Executive Director, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems; Executive Director, The Council on Extended Intelligence; Author, *Heartificial Intelligence: Embracing Our Humanity to Maximize Machines*

Affective Computing

- **Noreen Herzfeld** – Reuter Professor of Science and Religion, St. John's University
- **Chihyung Jeon** – Assistant Professor, Graduate School of Science and Technology Policy, Korea Advanced Institute of Science and Technology (KAIST)
- **Preeti Mohan** – Software Engineer at Microsoft and Computational Linguistics Master's Student at the University of Washington
- **Bjoern Niehaves** – Professor, Chair of Information Systems, University of Siegen
- **Rosalind Picard** – Rosalind Picard, (Sc.D, FIEEE) Professor, MIT Media Laboratory, Director of Affective Computing Research; Faculty Chair, MIT Mind+Hand+Heart; Co-founder & Chief Scientist, Empatica Inc.; Co-founder, Affectiva Inc.
- **Edson Prestes** – Professor, Institute of Informatics, Federal University of Rio Grande do Sul (UFRGS), Brazil; Head, Phi Robotics Research Group, UFRGS; CNPq Fellow
- **Matthias Scheutz** – Professor, Bernard M. Gordon Senior Faculty Fellow, Tufts University School of Engineering
- **Robert Sparrow** – Professor, Monash University, Australian Research Council "Future Fellow", 2010-15.
- **Cherry Tom** – Emerging Technologies Intelligence Manager, IEEE Standards Association

For a full listing of all IEEE Global Initiative Members, visit standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ec_bios.pdf.

For information on disclaimers associated with EAD1e, see *How the Document Was Prepared*.

Endnotes

[1] See B. J. Fogg, Persuasive technology. *Ubiquity*, December: 2, 2002.

[2] See S. Turkle, W. Taggart , C.D. Kidd, and O. Daste, "Relational artifacts with children and elders: the complexities of cybercompanionship, *Connection Science,* vol. 18, no. 4, 2006.

[3] A discussion of intimate robots for therapeutic and personal use is outside of the scope of *Ethically Aligned Design, First Edition*. For further treatment, among others, see J. P. Sullins, "Robots, Love, and Sex: The Ethics of Building a Love Machine." *IEEE Transactions on Affective Computing* 3, no. 4 (2012): 398–409.

[4] See R. Thaler, and C. R. Sunstein. *Nudge: Improving Decision about Health, Wealth and Happiness*, New Haven, CT: Yale University Press, 2008.

[5] See J. de Quintana Medina and P. Hermida Justo. "Not All Nudges Are Automatic: Freedom of Choice and Informative Nudges." Working paper presented to the European Consortium for Political Research, Joint Session of Workshops, 2016 Behavioral Change and Public Policy, Pisa, Italy, 2016; and M. D. White, *The Manipulation of Choice. Ethics and Libertarian Paternalism.* New York: Palgrave Macmillan, 2013.

[6] See, for example, J. Borenstein and R. Arkin. "Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being." *Science and Engineering Ethics,* vol. 22, no. 1 (2016): 31–46.

[7] See S. Omohundro, "Autonomous Technology and the Greater Human Good." *Journal of Experimental and Theoretical Artificial Intelligence* 26, no. 3 (2014): 303–315.

[8] See J. Roughgarden, M. Oishi, and E. Akçay. "Reproductive Social Behavior: Cooperative Games to Replace Sexual Selection." *Science* 311, no. 5763 (2006): 965–969.

[9] See the Well-being chapter of this *Ethically Aligned Design, First Edition*.