

Artificial Identity: Disruption and the Right to Persist

Dina Babushkina
d.babushkina@utwente.nl
University of Twente

Faculty of Behavioural, Management and Social Sciences
Dept. of Philosophy
Enschede, Netherlands

Athanasios Votsis
a.votsis@utwente.nl
University of Twente

Faculty of Behavioural, Management and Social Sciences
Dept. of Governance and Technology for Sustainability
Enschede, Netherlands

ABSTRACT

Anthropomorphism, artificial identity, and the fusion of personal and artificial identities have become commonplace concepts in human-computer interaction (HCI) and human-robot interaction (HRI). In this paper, we argue for the fact that the design and life cycle of 'smart' technology must account for a further element of HCI/HRI, namely that, beyond issues of combined identity, a much more crucial point is the substantial investment of a user's personality on a piece of technology. We raise the fact that this substantial investment occurs in a dynamic context of continuous alteration of this technology, and thus the important psychological and ethical implications ought to be given a more prominent place in the theory and design of HCI/HRI technology.

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models; HCI theory, concepts and models**; • **Applied computing** → **Law, social and behavioral sciences**; *Law, social and behavioral sciences*.

KEYWORDS

anthropomorphism, artificial identity, cognitive extenders, disruptive technology, ethics of technology, smart devices

ACM Reference Format:

Dina Babushkina and Athanasios Votsis. 2018. Artificial Identity: Disruption and the Right to Persist. In *HRI '21: Robo-Identity: Artificial identity and multi-embodiment Co-located with the 2021 International Conference on Human-Robot Interaction (HRI 2021), March 12, 2021, Boulder, CO*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

Anthropomorphism (see e.g. Nyholm 2020; Zebrowski 2020; Damiano and Dumouchel 2018; Airenti et al. 2019; Darling 2017) has become somewhat of a common-place discussion in the fields of Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI). Interestingly, anthropomorphism does not usually have negative connotations in these fields, compared, for example, to

traditional philosophical and psychological discourses. According to our definition, a person may be said to anthropomorphise an object/artefact if (a) he/she is attributing to it typically human features (including conscious and unconscious beliefs, as well as emotional reactions) and (b) guides his/her actions towards it by the belief that it has such features (including treating it as being animate and expect it to respond to such treatment in the manner appropriate for a human). Classic psychological theories would see in this a sign of certain psychological immaturity (as we can see on the example of animism, i.e. the tendency to see natural phenomena as possessing such properties as will), or perhaps even disorder (e.g. delusions), and philosophers would see in it a case of a mistaken belief. In any case, anthropomorphism would be seen as something to be concerned about and to fight with (cf. e.g. Richardson 2019; Turkle 2011; Wilks 2010). Curiously, in HRI and HCI it is often taken as a fact that anthropomorphism has normative implications which have to be taken into account during the design of robotic and AI systems. It is not uncommon to see this sort of implicit reasoning: since people tend to attribute to their devices human features, get attached to them as they would do to humans, and do treat them as they were in some sense like humans, therefore the design process should meet this demand and we should construct devices with more human-like identity.

In this paper we aim to focus on that part of the aforementioned inference which, we think, is generally the right intuition, but is taken to be so for a wrong reason. The intuition is that there is a need to account for the phenomenon of *artificial identity* in robots and computer-based technology, and that such identity is inseparable from personal identity. But it is not the fact of the user's tendency to transfer human features to inanimate objects (not only smart devices, but also plush toys, mountains, rivers, wind, etc) that should guide the design and support the conceptualization of artificial identity (cf. e.g. Seibt et al 2020; Damholdt et al 2020). There is a much better and more ethically weighty reason for that, i.e. *the investment of the user's personality (and often life) into a piece of technology and the disruptive effect that certain types of alternations of such a piece of technology has on its user*. This fact helps better appreciate the uniqueness of modern technology (especially such as various "smart" devices¹) and find the true meaning and locus of their artificial identity.

Unpublished working draft. Not for distribution.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted by ACM, provided that the copies are not made for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
HRI '21, March 12, 2021, Boulder, CO
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00
<https://doi.org/10.1145/1122445.1122456>

¹In the context of this paper, "device" is used in a sense of any piece of technology hosting software. It is assumed that, as far as the phenomenon of artificial identity is concerned, software has the primary importance. This, however, does not mean that hardware does not play a role, especially for robots that are designed to closely interact with humans. But the extent of such a role is a matter of further discussion.

2 WHY NOT PERSONAL IDENTITY?

One claim that we make is that artificial identity should not be conceptualised analogously with personal identity or the identity of animals (DiGiovanna (2017) explores the possibilities of such analogy). We give two reasons for this:

(1) Constructing the notion of artificial identity on the basis of comparison with personal identity is bound to be asynchronous and asymmetric. By asynchronous we mean that if we were to consider the "life cycle of an identity", i.e. the period from the time when the entity can be said to start being a unique individual to the point when it is no more, we are bound not to find the correct correspondence (one reason for that is that there is generally no agreement about the beginning and the end of personal identity). By asymmetric we mean the fact that artificial agents have unique properties (such as the ability to exist on multiple devices) which cannot be un-problematically transferred to human agents, and vice versa.

(2) The ascription of identity to a human or animal appears to entail the *non-derivative right to persist*, i.e. the right to stay the same (continue to exist as the same entity) and not to be subjected to changes significant enough to stop being the same. In other words, animals and humans enjoy the right not to be mutilated physically and altered psychologically. And if that happens, these actions should be condemned as a torture. Humans and animals have such a right in virtue of having their unique identities. Objects, on the other hand, only have the right not to be significantly altered iff they are property (but then again, I am free to do anything I wish to my property, but not someone else's) or of recognized value (such as art objects). Thus, the ascription of identity to objects is crucially different. Now, where does artificial identity belong? Devices do not have the non-derivative right to persist like animals and humans do. This is clear from the ease with which devices are modified, updated, or exchanged for newer models. By default, the software in our computers, tablets and phones is regularly updated in the manner that significantly alters interface, functionality, personal settings, and as a result after the overall performance and workflow.

By this it is not meant to argue that identity may never be ascribed to a device. We merely say that there is a need to find a way to (a) identify conditions under which identity can be ascribed to a it and (b) conceptualize artificial identity as an identity of a device, respecting its unique ontological status.

3 FRAMING ARTIFICIAL IDENTITY

One peculiarity of "smart devices"² is that they are reproducible, interchangeable, and can exist on multiple hardware at the same time. So, where does the unique identity of each token start? The problem is that not all tokens will have an identity. Many will remain nothing more but replicas. So what decides whether a given token (say, my tablet vs yours) will come to have an identity? It is our suggestion that it makes sense to talk of artificial identity

²We use "smart" devices in a broad sense, referring to technology designed to engage and interact with the user, and utilize obtained data to adapt to the user.

when the actions that alter or destroy the device, count—from the perspective of the user—as a non-trivial disruption. In other words, we have to evaluate what effect such manipulations as substituting, significantly altering, updating, overwriting the software of the device has on its user's productivity, cognitive capacities, and work/life routines. In cases when sudden alterations in the device inhibits the user's ability to create and be productive, disrupts the flow of life and lowers her experience of the interaction with the device, we are warranted to say that the token was irreplaceable and thus had obtained a certain identity.

For this to happen you do not need to anthropomorphise your tablet and your laptop. You may develop a strong bond to your connected devices, or even pair with them, without any shadow of belief that they are more than what they are—inanimate artifacts. And still you may be attached to them, depend on them, experience considerable discomfort and feelings of loss where they to be destroyed or their data and software damaged. The reason is that during the history of your interaction with your tablet or laptop *you have invested a part of your own personality* into them by fine tuning the settings, training algorithms with your data, interconnecting various applications and servers due to your specific needs and preferences, feeding information about the events in your life to various systems, and making all that influence, structure, assist, and often regulate various aspects of your life, work, and interactions with others.

Of course, in principle, this is true of any piece of technology. Even an artist's chisel may come to possess identity if it becomes his favourite tool, extension of his hand and mind. It is unique because it cannot be easily substituted with a new one. Even repairing the instrument may destroy its unique properties that the artist relies upon to deliver the results that she expects. The instrument is unique because it is fine tuned to the artist, her unique needs and abilities. Without it, the objects created would not have had the same distinct features, would not bear the mark of the craftsman, and, which is equally important, the artist would not have had the same quality of the creative experience. But computer technology and even more so "smart devices", not only have a stronger tendency for that, but altering them tends to be more disruptive for the user. This is due to the fact that these, more often than other types of technology, *become the user's cognitive (and sometimes emotive) extenders* (see more Vold and Hernández-Orallo forthcoming; Hernández-Orallo and Vold 2019; Vold 2018; Clowes 2013; Barr et al. 2015; Clark and Chalmers 1998). The locus of artificial identity is, as Hegel would put it, "for-us". This however does not mean that it exists only in the imagination of a human. We believe that artificial identity is constituted by tangible elements, such as specific settings, modifications of the system, memory, etc. But what we mean by "for-us" is that without the user, whose cognitive capacities it extends, the artificial identity would be meaningless.

What we find interesting about technology is that it is often misleading to talk about it in terms of general principles or *a priori* qualities. The true intrigue is in *conditionals*, in concrete and tangible details of user experience. It is the unique conditions of the specific human-technology interaction environment that creates and shapes the identity of a device. These specific conditions determine whether the device remains an interchangeable copy or becomes a non-disposable, non-interchangeable, unique entity.

This brings us to our main claim:

artificial identity emerges in the interaction of the device with the user, and it is through this interaction that the device obtains its distinct features and the locus of its identity.

Artificial identity, according to this view, consists of

the extension of its user's personality, synthesis with the user's psychological states. It is in this psycho-synthesis that a device develops its unique characteristics.

4 ETHICAL IMPLICATIONS

The account of artificial identity outlined in this paper poses a serious challenge for developers and existing practises of software life cycle such as update schedules. Such practices will have to be modified so that they avoid unsolicited changes in the devices which entail disruptions in the functioning of the user. This is why. Artificial identity is a matter of asymmetric unity of the artificial and human agents. It is, in effect, an extension of human identity. Because the user has invested her personality into the device, what threatens artificial identity exposes the vulnerability of the user. By pairing with the device, investing her personality in tuning the device, she opens herself to the risks of harm in case the device stops being the same (cf. Carter and Palermos 2016; Heersmink 2017). This mutual vulnerability grants the identity extender *its unique right to persist*. We must raise awareness about the unique psycho-synthesis with the artificial agents we live with. Overwriting, updating or substituting the technical component of this psycho-synthesis with a blank copy, is nullifying the user's personal investment into the device. A somewhat similar effect would be produced if a building company would each year return your property to the state it was before you moved in (incl., removing paint from the walls, scraping the floor, and in some cases getting rid of your furniture). For personal assistants, the effects of external intervention into its functioning can be analogous to a case of your partner resetting once a year, completely forgetting you and the history of your relationship.

Given that smart devices are no longer interchangeable but rather function as extenders of the cognitive capacities of users, alteration of which is associated with significant harm to the user, intrusions into the device's software are no longer a morally neutral act (cf. Bostrom, Sandberg 2009; Danaher 2018). There is a need to assess the impact of such resets to the user, and a part of responsible design is to minimize the adverse effects on the user.

ACKNOWLEDGMENTS

Dina Babushkina has received support from University of Twente's Gravitation Programme, Human Research Line. The contribution is part of the research programme Ethics of Socially Disruptive Technologies, which is funded by the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

2021-03-03 12:20. Page 3 of 1-4.

REFERENCES

- Airenti, G., Cruciani, M., Plebe, A. (Eds.). (2019). The Cognitive Underpinnings of Anthropomorphism. Frontiers Media SA.
- Barr, N., Pennycook, G., Stolz, J., Fugelsang, J. (2015). The brain in your pocket: evidence that smartphones are used to supplant thinking. *Computers in Human Behavior* 48 (2015), 473–480.
- Bostrom, N., Sandberg, A. (2009). Cognitive enhancement: methods, ethics, regulatory challenges. *Science and Engineering Ethics* 15, 3 (2009), 311–341.
- Clark, A., Chalmers, D. (1998). The extended mind. *Analysis* 58, 1 (1998), 7–19.
- Carter, J. A., Palermos, S. O. (2016) Is having your computer compromised a personal assault? The ethics of extended cognition. *Journal of the American Philosophical Association*, 2(4), pp. 542–560.
- Clowes, R. W. (2013) The Cognitive Integration of E-Memory. *Rev.Phil.Psych.* 4, 107–133.
- Damholdt, M. F., Vestergaard, C. Seibt, J. (2020). Testing for 'anthropomorphization': a case for mixed methods in Human-Robot Interaction. In C. Jost, B. Pévédic, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, M. Grandgeorge, N. Mirnig (Eds.), *Human-robot interaction : evaluation methods and their standardization* (pp. 203–227). Springer. Springer Series on Bio-and Neurosystems Vol. 12.
- Damiano, L., Dumouchel, P. (2018). Anthropomorphism in Human–Robot Co-evolution. *Frontiers in Psychology*, 9, 468.
- Danaher, J. (2018). Toward an ethics of AI assistants: An initial framework. *Philosophy and Technology*, 31(4), 629–653.
- Darling, K. (2017). "Who's Johnny?" Anthropomorphic Framing in human–robot interaction, integration, and policy. In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press.
- DiGiovanna, J. (2017). Artificial identity. In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press.
- Hernández-Orallo, J., Vold, K. (2019). AI Extenders: the ethical and societal implications of humans cognitively extended by AI. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)*. Association for Computing Machinery, New York, NY, USA, 507–513.
- Nyholm, S. (2020). Humans and robots: ethics, agency, and Anthropomorphism. Rowman and Littlefield.
- Heersmink, R. (2017). Extended mind and cognitive enhancement: Moral aspects of cognitive artifacts. *Phenomenology and the Cognitive Sciences* 16, 1, 17–32
- Richardson, K. 2019. The Human Relationship in the Ethics of Robotics. *AI and Society*, 34(1): 75–82
- Seibt, J., Vestergaard, C., Damholdt, M. F. (2020). Sociomorphing, not anthropomorphizing: towards a typology of experienced sociality. In M. Nørskov, J. Seibt, O. Quick (Eds.), *Culturally Sustainable Social Robotics—Proceedings of Robophilosophy 2020*. IOS Press Amsterdam. *Frontiers of Artificial Intelligence and Its Applications*, 51-67.
- Turkle S. (2011). *Alone together*. New York: Basic Books.
- Vold, K. (2018). Overcoming deadlock: scientific and ethical reasons to embrace the extended mind thesis. *Philosophy and Society* 29 (2018), 475. Issue 4.

233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290

291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348

349 Vold, K., Hernández-Orallo, J. (forthcoming) “AI extenders and
350 the ethics of mental health.” In *Ethics of Artificial Intelligence in*
351 *Brain and Mental Health*. Springer.

352 Wilks Y., (ed) (2010). *Close engagements with artificial compan-*
353 *ions*. Amsterdam: John Benjamins Publishing Co.

Zebrowski, R. L. (2020). Fear of a bot planet: anthropomorphism,
humanoid embodiment, and machine consciousness. *Journal of*
Artificial Intelligence and Consciousness, 07(01), 119–132.

Unpublished working draft.
Not for distribution.