



HAL
open science

Frankenstein's Problem

Lucy Suchman

► **To cite this version:**

Lucy Suchman. Frankenstein's Problem. Working Conference on Information Systems and Organizations (IS&O), Dec 2018, San Francisco, CA, United States. pp.13-18, 10.1007/978-3-030-04091-8_2 . hal-02083599

HAL Id: hal-02083599

<https://inria.hal.science/hal-02083599>

Submitted on 29 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Frankenstein's Problem

Lucy Suchman¹

¹ Lancaster University, UK

Abstract. This text is based on an invited address presented at IFIP 8.2 'Living with Monsters' in San Francisco, CA, 11 December 2018. Taking the 200th anniversary of Mary Wollstonecraft Shelley's *Frankenstein* as a starting place, I explore questions of autonomy and control with respect to human/technology relations. I consider the ambivalence of these agencies, and recent initiatives in science and technology studies and related fields to reconceptualize the problem as matters of relation and care. While embracing this turn, I reflect as well upon the ambivalences of relation and care, and the need to address the resilient politics of alterity in our figurations (and celebrations) of the monstrous.

Keywords: Technological Autonomy, Control, Care.

1 The Ambivalence of Control

In *Autonomous Technology* [16] Langdon Winner observes that imaginaries of technics-out-of-control have been a persistent preoccupation in modern political thought. Winner's project in the book is to cast a critical eye on this collective anxiety, and at the same time to think through some ways in which questions of control in relation to technological systems could be important for us to engage. Most relevant for the theme of this conference, the book's final chapter, titled 'Frankenstein's Problem,' calls on us to read Mary Wollstonecraft Shelley's original text [8], now celebrating its 200th birthday, as the articulation of a deeply ambivalent relationship to technological power.

To bring the ambivalence of technological power and control into the present moment, I begin with an iconic image from photographer Peter Souza, documenting the meeting of President Barack Obama and members of his national security team in the White House Situation Room [9], during the mission against Osama bin Laden on May 1, 2011 (an event that we now know has itself been mythologized, but that's another story; see [4]). There is much to be said about this photograph, but I read it in this context through questions that I take the figure of Frankenstein to index for us; that is, autonomous technologies-as-monsters, control, responsibility and care. Here we see the audience to an event outside the frame for us as viewers of this image, and distant for them as well, as they watch video feeds transmitted from the scene of operations in Abbottabad, Pakistan. Rather than an autonomous weapon system in which the human has been removed from 'the loop' (I return to that shortly), what we have here is a version of the 'loop' and some of the humans and technical systems that inhabit it, operating under a configuration that might be characterised as a highly complex form of remote control.

But what sense of ‘control’ exactly is in play here? The bodies crowded together in the room look on, mesmerized, apprehensive, but with little hint as to their own responsibility for the events that they are witnessing. Or read another way, it is only their absorption as spectators that implies their sense that they are themselves implicated. They’ve set something in motion; but it’s now out of their hands, and they can only watch it unfold. This is Frankenstein’s problem, then, in a 21st century manifestation.

2 The Monster’s Birth

Shelley’s text addresses themes of creation, neglect, and their consequences. As a technoscientist born out of a youthful infatuation with the alchemists, Victor Frankenstein is subsequently dazzled first by an encounter with lightning, and then with chemistry and the powers of Enlightenment reasoning. Loss of control over one’s desires is a theme early on, as Victor explains “None but those who have experienced them can conceive of the enticements of science” [3:32]. By the early pages of Volume I Victor has discovered, in his own words, “the cause of the generation of life; nay more, I became myself capable of bestowing animation upon lifeless matter” [3:34].

For those of us committed to demystifying what Herbert Simon some 150 years later named “the sciences of the artificial” [9], Victor’s account of how he actually made his creature has little to offer. We learn that having achieved the capacity to bestow life, the creation of in his words “a frame for the reception of it” was a painstaking project. He reports that he spent his time “in vaults and charnel houses,” and experienced “days and nights of incredible labour and fatigue” [3:33-34]. His decision to make the creature somewhat larger than life was based on the practical difficulties of working at a smaller scale. Having formed the determination “to make the being of a gigantic stature; that is to say, about eight feet in height, and proportionably large ... and having spent some months in successfully collecting and arranging my materials, I began” [3:37].

What follows is a vivid account of the details less of the creature’s composition, than of Victor’s own mental, physical, and emotional labours as, in Victor’s words, “with unrelaxed and breathless eagerness, I pursued nature to her hiding places” [3:38]. As is the case throughout the novel, Shelley’s writing is devoted overwhelmingly to relations and structures of feeling, including those between passion and labour, family and friendship and associated feelings of negligence and guilt, comfort and care. Even at the moment that the creature finally comes to life, how that happens is of less interest to Shelley than Victor’s response to it including, infamously, the repulsion that causes him to abandon the creature despite the latter’s attempts to engage him. When the creature re-encounters Victor and persuades him to listen, it describes its own ambivalent encounters with the human world, including its heartbreaking abjection, in extraordinary detail. Even here, however, what enabled the creature’s capacities for language, communication, reflection, injury, rage, and ultimately despair were not a primary concern either for Victor or it seems for Shelley. But they are

crucial questions for our understanding of comparable claims and promises in the contemporary field of humanoid robotics.

3 Monstrous agencies

Fast forward two centuries, and on October 11th of 2017 the United Nations General Assembly Economic and Social Council is holding a meeting in New York titled ‘The Future of Everything – Sustainable Development in the Age of Rapid Technological Change.’ Widespread media coverage of this meeting is prompted by the appearance of a robot named Sophia, symptomatic of the latest reanimation of the field of AI [15]. Figured in the image of film icon and humanitarian Audrey Hepburn, according to creator David Hanson, Sophia has become the spokesmodel first and foremost for Hanson Robotics, based in Hong Kong, and secondly for the imminent arrival of the humanoid robot as a transformative global force. Announcing, “I am here to help humanity create the future,” Sophia’s demonstration at the United Nations comprises a closely scripted ‘conversation’ between the show robot (effectively an animatronic manikin) and Deputy Secretary General Amina J. Mohammed. As the robot utters this statement, it slowly raises its arm until its multiply jointed hand is positioned inappropriately close to the Deputy Secretary’s face, in what consequently might be read as a threatening gesture. Mohammed mugs a grimace of perhaps only partially mock horror, eliciting a laugh from her fellow delegates.

Sophia’s software was sourced from the company SingularityNET, a name that gestures towards desires for and fears of humanity’s machinic supersession. One plotting of that trajectory has brought me to the UN myself for the meeting of another body, fittingly for the Frankensteinian narrative based in Geneva. The Convention on Certain Conventional Weapons or CCW is a body charged with establishing “prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects.” Following a report by the UN Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions to the Human Rights Council in 2013 [5], member states agreed to begin discussions at the CCW on what were then christened ‘lethal autonomous weapon systems.’ A coalition of non-governmental organizations headed by Human Rights Watch participated in the CCW’s preliminary series of ‘informal meetings of experts,’ aiming to build support for a pre-emptive ban on weapon systems in which the identification of targets and initiation of attack is put under fully automated machine control. The campaign is premised on the observation that the threat posed by robotic weapons is not the prospect of a Terminator-style humanoid, but the more mundane progression of increasing automation in military systems. A central concern is initiatives to automate the identification of particular categories of humans (e.g. those in a designated area, or who fit a machine-readable profile) as legitimate targets for killing.

As a member of the International Committee for Robot Arms Control, part of the NGO coalition, I presented testimony at the CCW in April of 2016 in a panel on the question of machine ‘autonomy’ [12, see also 14]. My brief contribution focused on the problem of ‘situational awareness,’ accepted in military discourses to be a

precondition for distinction between legitimate and illegitimate targets, a prerequisite to legal killing within the frameworks of International Humanitarian Law. I made a case for the inherent limits of algorithmic approaches to situational awareness, particularly in terms of the requirement of distinction.

My effort to pose an irremediable obstacle to the legality of autonomous weapon systems within the military's own terms required suspension of what would otherwise be profound questions for me about those terms, beginning with the trope of 'autonomy,' treated here as at once the litmus test of the model human subject, and that which is in danger of escaping human control. Further disconcertment comes with the principle of distinction in international laws of armed conflict, where enactments of difference are at their most lethal, in the profoundly gendered, racialized, and irremediably contingent categories of 'us' and 'them' that govern violence in practice. In a new book titled *The Robotic Imaginary* [7], Jennifer Rhee argues that positing the human as originary and as recognisable effectively underwrites the dehumanization of that which can't be known. Following Judith Butler's question regarding differently grievable lives she asks: "Who, in their purported incommensurability, unknowability, unfamiliarity, or illegibility within robotics' narrow views of humanness, is excluded, erased, dehumanized, rendered not-human?" [7:4]. As a case in point she takes the United States drone program, characterized by practices of dehumanization that she contends are embedded in the history of robotics and its various inscriptions and erasures of the human.

The narrowness of robotic views of humanness that Rhee identifies is not only a technical problem, but also indicative of the wider problems of drone visualities. In the past few months these problems have come under some scrutiny thanks to a small but significant rebellion on the part of Google employees against the company's participation in Project Maven, a US Defense Department effort to gain some control over the vast store of video surveillance footage generated by its drone program. Despite protests on the part of the company that the seeing to be automated concerns only classes of objects, it soon became clear that the objects of interest include vehicles, buildings and indeed humans on the ground. Those of us who joined in support of the insurgent Googlers pointed out that further automation of the scopic regimes of the US drone program can only serve to worsen an operation that is already highly contested, and arguably illegal and immoral under the laws and norms of armed conflict [13].

4 The Politics of Alterity

As Mary Shelley taught us by breathing life into Frankenstein and his creature, the anthropomorphic artifact is a powerful disclosing agent for the relation between figurations of the human and operations of dehumanization. Dehumanization, Rhee argues, is a kind of evil twin of anthropomorphism, an 'anti-metaphor' that creates a relation of difference between what would otherwise be recognizable as similar and kindred entities. As an antidote she calls for "an understanding of the human through unrecognizable difference, and unfamiliarity, rather than recognition, knowability,

and givenness” [7:5]. The essence of Frankenstein’s problem, which Winner suggests is now a problem for us all, is “the plight of things that have been created but not in a context of sufficient care” [16:313]. But as recent feminist technoscience has taught us – notably Maria Puig de la Bellacasa in her book *Matters of Care* – care itself is “ambivalent in both its significance and its ontology” [6:1]. As a way forward de la Bellacasa writes in support of “committed knowledge as a form of care” [6:16]. In my own writings I’ve proposed continued engagement as an alternative to ‘control’ over the life of technologies (which we know is impossible) (11). Our inability to control something does not absolve us of being implicated in its futures. Rather, our participation in technoscience obliges us to remain in relation with the world’s becoming, whether that relation of care unfolds as one of affection or of agonistic intervention.

But while relationship is necessary, it is not sufficient. The problem of Frankenstein can be understood within a twofold cultural/historical imaginary, comprising on the one hand autonomy read as separateness, and on the other, fantasies of control. The twin figures of the autonomous machine as either the perfect slave or the cooperative partner [1:2], while positing a human-machine relationship, reinstate relations of dominance at worst, instrumentality at best. The former is problematic for its inbuilt injustice, while the justice of the latter is contingent on the projects in which both humans and machines are enrolled. Taken together, these imaginaries work to obscure the politics of alterity that operate through the figure of the monster, as well as the modernist genealogies that shape technology’s contemporary forms. The promise of monsters [2], in contrast, is that they might double back to challenge their makers, questioning the normative orders that are the conditions of possibility for their monstrosity.

References

1. Cardoso Llach, D.: *Builders of the Vision: Software and the Imagination of Design*. New York and Oxford: Routledge (2015).
2. Cohen, J. J.: The promise of monsters. In: Mittman, A.S., Dendle, P. (eds.) *The Ashgate research companion to monsters and the monstrous*, pp. 449-464. New York and Oxford (2012).
3. Guston, D., Finn, E., & Robert, J. S. (eds.) *Frankenstein: Annotated for scientists, engineers, and creators of all kinds*. Cambridge and London: MIT Press (2017).
4. Hersch, S.: The Killing of Osama bin Laden. *London Review of Books*, 21 May (2015).
5. Heyns, C.: Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions. Retrieved from Geneva: http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf (2013).
6. Puig de la Bellacasa, M.: *Matters of Care: Speculative Ethics in More Than Human Worlds*. Minneapolis and London: University of Minnesota Press (2017).
7. Rhee, J.: *The Robotic Imaginary: The human and the price of dehumanized labor*. Minneapolis and London: Minnesota University Press (2018).
8. Shelley, M.: *Frankenstein, or The Modern Prometheus*. London: Anonymous (1818).
9. Simon, H.: *The sciences of the artificial*. Cambridge: MIT Press (1969).

10. Souza, P.: Situation Room, photo taken by White House photographer Pete Souza at 4:06 pm on May 1, 2011. Wikimedia (2011), [https://en.wikipedia.org/wiki/Situation_Room_\(photograph\)](https://en.wikipedia.org/wiki/Situation_Room_(photograph)), accessed 22 September 2018.
11. Suchman, L.: Located Accountabilities in Technology Production. *Scandinavian Journal of Information Systems*, 14(2), 91-105 (2002).
12. Suchman, L.: Situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy. In Geiss, R., Lahmann, H (eds.), *Lethal Autonomous Weapon Systems: Technology, Definition, Ethics, Law & Security* (pp. 273-283). Berlin, Germany: Federal Foreign Office, Division Conventional Arms Control (2016).
13. Suchman, L., Asaro, P., Irani, L. Google's march to the business of war must be stopped. *The Guardian*, 16 May (2018). <https://www.theguardian.com/commentisfree/2018/may/16/google-business-war-project-maven>, accessed 22 September 2018.
14. Suchman, L., Weber, J.: Human-Machine Autonomies. In Bhuta, N., Beck, S., Geis, R., Liu, H.-Y., Kreis, C. (eds.), *Autonomous Weapons Systems* (pp. 75-102). Cambridge, UK: Cambridge University Press (2016).
15. UN News: At UN, robot Sophia joins meeting on artificial intelligence and sustainable development. <https://news.un.org/en/story/2017/10/568292-un-robot-sophia-joins-meeting-artificial-intelligence-and-sustainable> (2017/10/11).
16. Winner, L.: *Autonomous technology: technics-out-of-control as a theme in political thought*. Cambridge, Mass.: MIT Press (1977).