

Unintended by Design: On the Political Uses of “Unintended Consequences”

NASSIM PARVIN¹

GEORGIA INSTITUTE OF TECHNOLOGY

ANNE POLLOCK²

KING'S COLLEGE, LONDON

Abstract

This paper revisits the term “unintended consequences,” drawing upon an illustrative vignette to show how it is used to dismiss vital ethical and political concerns. Tracing the term to its original introduction by Robert Merton and building on feminist technoscience analyses, we uncover and rethink its widespread usage in popular and scholarly discourses and practices of technology design.

Keywords

unintended consequences; technological utopianism; god trick; angel trick; feminist killjoy; feminist STS

An Intrepid Intersectional Feminist Meets a Smart Cities Enthusiast

On a cool, fall afternoon, an intrepid intersectional feminist scholar meets with a colleague from another department for the first time at a coffee shop near campus. They've been introduced to each other because of their shared interest in smart cities, so they dive into the topic after a few minutes. She asks him about what he finds most compelling about smart cities initiatives as she warms her hands on the cup. He explains all the efficiencies, especially in managing the flow of traffic and reducing congestion. “There's so much that we can do through a centralized command and control center!” He beams as he describes what he finds to be the most compelling example: the ability to open up the path for emergency vehicles such as ambulances, which would save lives.

Our intrepid feminist responds: “I see where you are coming from. But there is a lot to be concerned about, too: like the kind of mass surveillance it makes possible, or the susceptibility of such a centralized system to hacking or other breakdowns that would propagate across entire cities in contrast to the more contained and local impact of existing distributed systems. There's

¹Nassim Parvin, Email: nassim@gatech.edu

²Anne Pollock, Email: anne.pollock@kcl.ac.uk

also the environmental and human labor impact of instrumenting cities with sensors, especially if we consider their entire life cycle from extraction of rare minerals to disposal and recycling. And even if we set all these issues aside, we are left with the immediate problem of the lack of coordination with other relevant services in the city: What's the point of getting the patient to the hospital faster if they would still have to wait for a long time once they get there? Wouldn't it be more effective if we made sure healthcare was more accessible both in terms of distance and availability—including insurance for preventative care so that fewer people end up in the emergency room?"

His face has lost all expression, as if to suggest he has heard the story before. Yet he nods in agreement. She stops talking. He finally breaks the awkward silence, and, in a tone both dismissive and defensive, says, "Sure. There are always unintended consequences." With a barely noticeable shrug, he sits back in his chair. Case closed.

"Yeah," she says, feeling overwhelmed (maybe she should not have brought these issues up but she's now in too deep). "But look. These technologies are not new, nor are the issues they raise: exacerbating disparities, environmental impact, surveillance, etc. Why not actively consider and address those issues now as we design and implement the next generation of digital technologies, especially given the scale and impact of cities' investment in them?"

"It is not like they are not considering these issues," he responds loudly. (Perhaps he is raising his voice to be heard over the growing noise of the coffee shop chatter, but our intrepid feminist thinks she hears a note of annoyance in his tone, too.)

"One point of the first phases of the smart cities initiative is to gather data about some of those same issues. But even then, what to do about it is the domain of policy, and we are yet to get a clear direction from them. The city would definitely work within the limits of policy! In the meantime, it is important to remember that we cannot let our city fall behind as so many other cities are already implementing this technology. Oh, and by the way, I am sympathetic to the problems of long wait times in hospitals, but that's an entirely different entity in the city. The budget for the smart cities initiative is separate from that of healthcare and hospitals."

He seems well-meaning, and is convinced that smart city traffic control is the best course of action. From there, the conversation roams around for the remainder of the hour as the shop gradually clears of the afternoon rush. The enthusiast's dismissal of her concerns as "unintended consequences" rings in her ears as they say goodbye. She doubts that they will talk again.

The above vignette is a refined composite of our encounters of a similar nature over the years. The point of it is not to call out our interlocutors as individuals, but rather to draw readers' attention to the script of a seemingly mundane argument. The vignette illuminates the uses (and misuses) of the term "unintended consequences" in science and technology discourses and also captures the dismissive affective quality of its use, which serves to remind our intrepid intersectional feminist that she is raising issues that bear on an already-settled case.

The conversation could be centered on any intervention, technical or otherwise. One could raise the problematic qualities of invocations of "unintended consequences" through many examples, including ones from the novel coronavirus as it unfolds during the time of this writing: the intensification of domestic violence experienced by so many women and children who are

locked down with their stressed-out abusers; the deaths of older adults from forms of dementia that are exacerbated by their isolation; the advocacy for steep fines to enforce masking, which would expose those who are already differentially vulnerable to police violence and mass incarceration to more of the same. Single-minded focus on a problem as narrowly defined inhibits more capacious consideration of how we might foster authentic well-being and holistic and inclusive social good.

On the surface, the term “unintended consequences” captures a rather straight-forward idea: the consequences of technologies or other interventions that were unforeseen at the time of their conception or design. Nobody realized at the time that such and such would follow from what they were doing.

But even a quick survey of the uses of the term in public and scholarly discourse suggests a subtle but significant shift in usage: those consequences of technology that can indeed be anticipated in advance but that fall outside of the purview of the specializations that conceive or implement products. As such, the concept is emptied of its substance while doing substantial work—especially for tech companies and technology developers, as well as their enthusiasts. It provides a category that is descriptive of the social, environmental, and political impacts of science and technology as ones that lack prior, deliberate action. Phenomena described as unintended consequences are deemed too difficult, too out of scope, too out of reach, or too messy to have been dealt with at any point in time before they created problems for someone else. The descriptive approach works as a defensive and dismissive strategy.

Theoretical Roots and Evolution of the Term: Anticipation vs. Intention

The phrase “unintended consequences” found widespread usage in contemporary discourses on technology and policy after canonical sociologist of science Robert Merton’s (1936) article titled: “The Unanticipated Consequences of Purposive Social Action.” In this paper, Merton identifies three obstacles that limit one’s ability to *anticipate* the consequences of purposeful action to achieve a particular goal: ignorance, error, and ideological blindness. Ignorance stems from a limitation in knowledge. This may include the inability to predict the outcomes of actions with certainty, complexity in the number of factors and variables that are involved, or ignorance about what type of knowledge could or should be obtained in relation to a given situation. The second obstacle to our ability to anticipate the consequences of actions is error. This includes error in appraisal of a given situation, or error in identifying the range of possible actions available, or error in the assumption that actions that have previously secured certain outcomes will be appropriate for a new situation. The third obstacle is ideological blindness, a term that we might avoid because of ableist connotations, that is explained by Merton with the fabulous turn of phrase “the imperious immediacy of interest.” This category refers to “instances where the actor’s paramount concern with the foreseen immediate consequences excludes the consideration of further or other consequences of the same act. The most prominent elements in such immediacy of interest may range from physiological needs to basic cultural values” (Merton 1936, p. 901).

Science policy analyst Frank de Zwart (2015) traces the problematic conflation of “unintended consequences” and “unanticipated consequences” in Merton’s work and beyond: “unanticipated consequences” has been overwhelmingly replaced in usage with the all-too-common “unintended consequences,” which had been used as its synonym. De Zwart persuasively argues that the conflation is often useful for academic analysts of public policy because it allows them to operate on the assumption that policy makers are ignorant of potential adverse effects, and to write papers whose project is limited to simply uncovering those potential adverse effects. However, policy makers, not unlike technology designers, are often perfectly aware of the potential of their actions to cause harm, and they proceed anyway. Addressing the question of why policy makers take particular actions *despite* knowing about potential harms is a much more challenging—and much more important—project for analysts than settling for unfounded implications of ignorance with the term “unintended consequences.” Here, we draw on feminist technoscience critique to explore additional elements of the allure of the invocation of “unintended consequences.”

Exculpatory: Critique without Accountability

The loss of precision that comes with using the term “unintended consequences,” instead of the original “unanticipated consequences,” is deeply consequential. The replacement of “unintended” for “unanticipated” implies that all unintended consequences were unanticipated; thus, we lose sight of the category of consequences that *were* anticipated—even as they may be unintended. At an interpersonal level, it works to facilitate social relations, like the one described in the vignette above, by dulling the sharp edge of critique. Concerns are acknowledged by enthusiasts, described under a nonthreatening descriptor, and then set aside for others, be they ethics boards or policy makers, to deal with after the fact.

The conflation of intention and anticipation elides the distinction between them, and allows enthusiasts (of a particular policy, technology, or what have you) to abdicate responsibility for the perfectly foreseeable consequences of particular decisions, by positioning those consequences as outside their purview. This is an intriguing companion to the “god trick” identified by Donna Haraway (1988), one that we might call the “angel trick.” The “god trick” refers to putatively objective knowledges in which those who claim objectivity do so as though they see everything from nowhere rather than from a necessarily situated perspective. Objectivity, there, derives from its (im)partiality. We might extend this by proposing a related maneuver, wherein one lays claim to the position of god’s angel. The angel is a henchman, just following god’s orders, and the “angel trick” operates by claiming dominion only over a more limited sphere. God may be all-knowing and all-powerful, but, even if angels might be aware of other domains, to the extent that their knowledge bears on the actions they take, it is highly circumscribed, and so are their powers. When the angel encounters all else with the claim of total powerlessness and/or incompetence, that abdication might be naïve or might be devious, but the trick operates through a strategic absolution that looks a lot like modesty. The reliance on the valence of “unanticipated consequences” that lurks beneath “unintended consequences”

provides a way to acknowledge the limitations of knowledge while denying responsibility for the shortcomings of decisions made on the basis of that knowledge. Indeed, because it operates on the presumption of ignorance, it can also be used to argue for more fact gathering and variable input as the panacea. Both the “god trick” and its companion make an untrue claim to innocence on the part of the designer. However, as Haraway argues, noninnocence is the fundamental condition of all knowledge claims and of all critique. Whether claiming omniscience, as in the god trick, or clinging to a narrowness of scope of action, as in its companion., those to take or support actions that have knowable potential to harm should be held accountable.

More broadly, this category of consequences is significant given that the conflation of unintended and unanticipated consequences is, as expected, largely in the service of the smooth operation of power. It may be viewed positively in that by qualifying those consequences as “unintended,” analysts and critics can expose, discuss, or correct such issues without having to assign blame on those who may be responsible for or to them, or having to get into discussions about responsibility that may be too difficult or politically costly. At the same time, the conflation helps those in positions of power to avoid accountability, since they don’t have to own the consequences of their choices. It becomes possible for designers and advocates to claim shared values with critics—such as abhorrence of structural inequalities including sexism, racism, or ableism—without taking responsibility for ensuring that the design of technologies upholds those values (see Shelby 2020). (Note, too, that, should the consequences turn out to be positive, their ownership seems readily claimed.) It is true that we cannot know the intentions of those making decisions about technological design and policy. We can call into question, however, the claimed inability to anticipate consequences, tracing its roots in zealous techno-utopic values and interests that reinforce inequality, extraction, and exploitation. “Unintended consequences” operates in a subtle but insidious way to uphold the status quo, especially institutional structures and knowledge regimes that have produced those negative impacts—inclusive of the entrenched privileging of expertise in narrowly technical domains—while avoiding accountability.

Disciplinary: Acknowledgment without Commitment

Another reason that the notion of unintended consequences takes too much for granted is that, in some sense, this notion positions the baseline situation as outside the scope of impact. As Merton (1936, p. 895) writes: “Rigorously speaking, the consequences of purposive action are limited to those elements in the resulting situation which are exclusively the outcome of the action, i.e., those elements which would not have occurred had the action not taken place.” However, as Merton points out in the next sentence, “Concretely, however, the consequences result from the interplay of the action and the objective situation, the conditions of action.” Not to consider the conditions of the hospital waiting room when deciding to give vehicles preferential treatment through the streets is a deliberate decision. But breaking from the artificial narrowness of scope, the work of design is also the work of reimagining social structures and organizations.

The issues raised by the intrepid feminist introduced above can indeed be identified and anticipated but are dismissed by the smart cities enthusiast because they fall outside of the

interests, values, skills, or institutional and cultural structures of technology research and development. In the above example, in the absence of a systemic view of the problem and assessment of a range of products, services, and policy interventions that might improve life in the city inclusive of (but not limited to) travel times for emergency vehicles—such as the implementation of a public transportation system that would decrease the need for single occupancy vehicles and thus reduce congestion—we are left with a technical intervention driven by the capacities of the technology under the auspices of particular powerful technology representatives within municipalities (Sadowski and Bendor 2019).

And what about the affective qualities of the conversation with the techno-utopian who is excited about the potential to move emergency vehicles seamlessly through the city, and his invocation of “unintended consequences” to dismiss our intrepid intersectional feminist’s raising of broader social concerns? An invocation of perfectly anticipatable problems interferes with the glee of the neat technological fix, and as such, the feminist “kills joy,” as in Sara Ahmed’s (2010, p. 581) evocative terms: “The feminist killjoy spoils the happiness of others; she is a spoilsport because she refuses to convene, to assemble, or to meet up over happiness.” The disciplinary power of the phrase “unintended consequences” surfaces in its power to undermine and dismiss the epistemological value of what is being raised, while claiming ownership of it as being already known. The issues the feminist is raising are rendered uninteresting and out of place thus actively turned away (chill out with all the non-sequiturs, Cassandra!). What she knows is packaged in a simple two-word phrase: the boring and known category of “unintended consequences” that effectively denies the relevance of what she has to say.

Ideological: Ethics without Imagination

All too often, it is argued that the perfectly predictable entrenchment of existing inequalities is outside of design’s purview, but as Ruha Benjamin (2019, p. 79) argues in her compelling critique of the role of technology in perpetuating and intensifying inequalities: “A narrow investment in technological innovation necessarily displaces a broader set of social interests.” Figuring consequences as unintended putatively acknowledges known possibilities for failure, while at the same time undermining the kinds of knowing, and even spheres of knowledge production, that clearly demonstrate the inadequacy of cherished cultural assumptions that structure how we frame and solve problems.

The ideological mindset here may be broadly referred to as techno-utopianism. A more nuanced characterization of it, however, surfaces familiar but still under-appreciated feminist critiques of the dominant view of science and technology. Among them is the all-too-common misconception of technological development as inherently neutral, and the consequent decoupling of ethics and design, while paradoxically claiming technological progress as a moral commitment. In practice, technological pursuits are legitimized no matter how dangerous, leaving it to ethicists and policy makers to fine tune their features in extreme cases, adjust their conditions of use, or somehow otherwise address the fallout. In doing so, the concept of unintended consequences promotes the reductive view of ethics as a regulatory practice as

opposed to seeing design itself as a practical form of ethics that is fundamental to the conception and development of technologies (e.g., Whitbeck 1996; Pantazidou & Nair 1999; Verbeek 2006).

The invocation of “unintended consequences” becomes a way to marginalize the profound ethical questions at the heart of technological design choices. We might easily picture how conversations like the one above could unfold differently to serve as a starting point of an imaginative exchange of ideas. There is already an opening in the encounter for transcending disciplinary boundaries. Both of the interlocutors are committed to improving life in the city. They have material and social capital. Why not think together about whether and what alternate approaches might look like? Or what other kinds of connections, interventions, or experiments may be possible? The ready-made script of “unintended consequences” instead works to shut down the conversation by assuming that capacious considerations are unproductive and unnecessary impediments to problem solving as narrowly defined. We are presented with false tradeoffs and seemingly inevitable compromises—such as the need to accept mass surveillance for the assumed benefit of efficiency in taking patients to hospitals—as opposed to opening up the problem space in ways that could accommodate other possibilities (Parvin, publishing as JafariNaimi et al., 2015). Why relegate the difficult ethical issues to some other entity at some other time after the fact when those same issues can foster collaborative commitments and creative approaches?

Conclusion

The term “unintended consequences” is a barrier to, rather than a facilitator of, vital discussions about design. In the encounter of the smart cities enthusiast and the intrepid feminist, the rhetorical move of deflecting blame by invoking unintended consequences ends the conversation. This provides an opportunity to understand the counterproductivity of invocations of unintended consequences in discussions of technological design more broadly. Questions of innocence and blame are not at stake in such conversations, necessarily, but the default retort works to deny accountability, imagination, or commitment. The overemphasis on intent—understood predominately in its individualistic interpretation—forecloses consideration of the complexity of social systems in such a way as to lead to quick technical fixes. Instead of rendering criticism in terms of blame or lack thereof, we can see it as an opening for meaningful intellectual engagement. Alternatively, attentiveness to the uses and misuses of “unintended consequences” could serve as a starting point for fostering imaginative and collaborative engagements imbued with possibility.

Nassim Parvin is an Associate Professor of Digital Media in the School of Literature, Media, and Communication at the Georgia Institute of Technology. Her research explores the ethical and political dimensions of design and technology.

Anne Pollock is a Professor of Global Health and Social Medicine at King's College London. She is currently completing her third book, *Sickening: Racism, Health Disparities, and Biopolitics in the 21st Century*.

Acknowledgements

We would like to thank the external reviewers for their helpful feedback, as well as *Engaging STS* Managing Editor Katie Vann for additional generous suggestions that helped to sharpen the argument.

References

- Ahmed, Sara. 2010. "Killing Joy: Feminism and the History of Happiness." *Signs* 35: 3 (Spring): 571–594.
- Benjamin, Ruha. 2019. *Race After Technology*. Cambridge, UK: Polity Press.
- De Zwart, Frank. 2015. "Unintended but Not Unanticipated Consequences." *Theory and Society* 44, no. 3: 283–297.
- Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3: 575–99.
- JafariNaimi (Parvin), Nassim, Lisa Nathan, and Ian Hargraves. "Values as Hypotheses: Design, Inquiry, and the Service of Values." *Design Issues* 31, no. 4 (2015): 91–104.
- Merton, Robert K. 1936. "The Unanticipated Consequences of Purposive Social Action." *American Sociological Review* 1, no. 6: 894–904.
- Pantazidou, Marina, and Indira Nair. 1999. "Ethic of Care: Guiding Principles for Engineering Teaching & Practice." *Journal of Engineering Education* 88, no. 2: 205–212.
- Sadowski, Jathan, and Roy Bendor. 2019. "Selling Smartness: Corporate Narratives and the Smart City as a Sociotechnical Imaginary." *Science, Technology, & Human Values* 44, no. 3: 540–563.
- Shelby, Renee. 2020. "Value-Responsible Design and Sexual Violence Interventions: Engaging Value-Hypotheses in Making the Criminological Imagination." In *Routledge Handbook of Public Criminologies*, edited by Kathryn Henne and Rita Shah, 286–298. New York: Routledge.
- Whitbeck, Caroline. 1996. "Ethics as Design: Doing Justice to Moral Problems." *Hastings Center Report* 26, no. 3: 9–16.
- Verbeek, Peter Paul. 2006. "Materializing Morality: Design Ethics and Technological Mediation." *Science, Technology, and Human Values*. 2006, 31, no. 3: 361–380.