The Concept of Handoff as a Model for Ethical Analysis and Design

Deirdre K. Mulligan and Helen Nissenbaum

The Oxford Handbook of Ethics of AI
Edited by Markus D. Dubber, Frank Pasquale, and Sunit Das

Print Publication Date: Jul 2020 Subject: Law, IT and Communications Law
Online Publication Date: Jul 2020 DOI: 10.1093/oxfordhb/9780190067397.013.15

Abstract and Keywords

This chapter introduces the concept of *handoff*, which offers a lens through which to evaluate sociotechnical systems in ethical and political terms. It is particularly tuned to transformations in which system components of one type replace components of another. Of great contemporary interest are handoff instances in which AI take over tasks previously performed by humans, for example, labelling images, processing and producing natural language, controlling other machines, predicting human action (and other events), and make decisions. Grounded in past work in social studies of technology and values in design, the *handoff* analytical model disrupts the idea that if components of a system are modular in functional terms, replacing one with another will leave ethical and political dimensions intact. Instead, the handoff lens highlights different ways that different types of system components operate and interoperate and shows these differences to be relevant to the configuration of values in respective systems. The handoff lens offers a means to make ethically relevant changes salient that might otherwise be overlooked.

Keywords: handoff analytical model, sociotechnical systems, automation, computational systems, human tasks, AI, ethical analysis

ENTHUSIASM for the new artificial intelligence (AI), derived from machine learning over big data, has meant a sweeping push to insert machine intelligence into wide-ranging systems, producing a raft of “smart” yet often mundane technical objects, as well as AI enhanced systems operating in key societal sectors including finance, military, transportation, criminal justice, and health and welfare.[1] As with automation in prior times, this sweep has also raised doubts and questions, notably, many focused on functional performance and worker displacement. The concept of *handoff* that we have developed guides a different set of questions, namely, how implanting AI[2] affects the ethical and political values embodied in technical systems.

A growing body of work that places technical artifacts themselves—devices and systems—within the scope of ethical analysis, beyond the traditional focus on human action and institutional regulation, has driven progress in understanding technology in ethical terms.

The object of study, according to this understanding, is not a purely material, technical system, performing within a purely human or social context, but is a sociotechnical system whose performance inextricably involves both. Actor-Network (p. 234) Theory (ANT), with its concept of actant, for example, goes even further in this direction, erasing the traditional distinction between human actor, on the one hand, and machine component, on the other. Systems developers may employ diverse nodes[3] in complex actor-networks wherein actants prescribe and delegate behaviors among one another to achieve desired ends. The concept of *handoff*, similarly, assumes a broadened understanding of the technical as, in fact, the sociotechnical, whereby (so-called) technical systems and devices function as they do because of technical and material properties, as well as human behaviors, and economic, social, and political contexts. Unlike ANT, however, handoff illuminates the differences among the different types of actants, if you will, where it considers that these differences are ethically relevant. Applying an ethical lens to technical systems, so conceived, means assessing these diverse dimensions in terms of the contribution they make, or the impact they have, on ethical and political values embodied—potential or enacted[4]—in such systems as a whole. In these assessments, the concept of *handoff* constitutes a useful analytic tool.

The paradigmatic use-case for the Handoff model involves a progression or transformation from one version of a system to another, where the progression involves the replacement of certain components by others. A simple illustration may help. In modern office buildings, lighting is increasingly modulated by motion sensors instead of mechanical, human-operated switches; we would describe this transformation as a handoff of control from a human actor to a programmed motion sensor. We note that often, alongside the motion-sensing control, a traditional interface affords individuals the option of operating a switch in the traditional manner—a paradigmatic example of a parallel configuration within a single system. Although the catalyst for us in developing an analytical framework around the concept of *handoff* was the recent boom in AI based automation, the lighting example shows that it applies generally, to various permutations, including automation involving the replacement of human actors by technical mechanisms (not necessarily AI), one type of machine component by a different type, as when hardware is replaced by software, or even human actors, in one capacity, replaced by other humans in other capacities. Such handoffs occur when, for example, functionality is outsourced, pushed to workers lower on a hierarchy, centralized, or decentralized, and so on. Examples abound.

Taking an ethical perspective on technical systems the concept of *handoff* is particularly useful because it exposes aspects of progressive transformations that may otherwise be overlooked. Those who claim about a given handoff, say, human moderation of content handed off to machines, that the transformed system offers *the same* functionality as the previous may boast, further, that it does so even more reliably, more efficiently, and at lower cost. If there is anything to worry about, goes this account, it is to (p. 235) ensure that content marked as offensive, illegal, or dangerous by the machine roughly meets respective standards. Like others,[5] however, we argue that even were this to hold, reallocation of functionalities among different types of components (or actors) does not necessarily leave the “mass of morality” unchanged: to the contrary, redistribution of functionality,

in itself, may have moral and political repercussions. The Handoff model resists the idea that one can redistribute functions without disturbing the mass of morality, and is designed to reveal the political significance of sociotechnical configurations of function across component actors and the points of inflection among them.

Mapping transformations in terms of the Handoff model shines a spotlight on that which has changed and, by implication, illuminates ethical concerns that these changes raise. It may be that transformed systems embody more positive values, but it may be that replaced components, even performing purportedly the same task, lead to a degradation—such as, dissipated accountability, diminished responsibility, displacement of human autonomy, or acute threats to privacy. In our view, the Handoff model is a critical ameliorative intervention illuminating the structural, political, and ethical stakes of the ongoing transition of control to computational components under the guise of progress and efficiency and often political neutrality.

Catalyst

AI applied in areas such as social media platforms, “smart” cities, healthcare, and the criminal justice system has generated steep and widespread interest. Regulators and journalists interrogate the political implications of algorithms in systems as diverse as Facebook’s advertising platform and risk recidivism software. Governmental bodies set out ethical expectations for AI in self-driving vehicles. Companies develop guidelines and internal structures to address the ethical quandaries posed by AI. Universities grapple with their obligation to produce students who can attend to the social and political entanglements of technical work. Workers within major technology companies oppose the use of their labor toward ethically objectionable ends. This burst of activity and the underlying ethical angst reveal the need for rigorous methods to interrogate the ethical implications of AI.

This historic inflection point, with the unspoken imperative to hand off human tasks to machines, in business, government, healthcare, education, in our view, raises an (p. 236) urgent need to characterize and assess potentially destabilizing impacts on values configurations. We already have experienced how latent barriers—physical, economic, time—that served as extralegal protection for privacy are undone by the interjection of machines: for example, drones that alter lines of sight, making fences and property lines insufficient to limit prying eyes; video surveillance systems that can identify individuals in a crowd; and online access to public records that make an individual’s past infractions as salient as her present successes. These experiences should inspire skepticism in the face of all claims of sameness, even if some of these claims prove ultimately to be innocuous. The *handoff* framework offers a guide to maintaining a focus on implicit as well as explicit values as sociotechnical systems evolve. With different types of actors performing different functions, respectively, across versions, the system will call on different modalities of control and regulation—technology, law, ethical norms, and economics. Surely some con-

figurations of functions will provide superior protection for particular values: this is our point of departure and focus of inquiry.

A simple case may illuminate the point. Take sealable envelopes. As a material approach to securing privacy in written correspondence, it achieves this function within a framework of legal protections against tampering, norms against reading private letters, locked letterboxes, and mail slots that bring letters behind locked doors. In other words, although sealable envelopes may qualify as a "privacy enhancing technology," postal privacy is a product of the sociotechnical system of legal, cultural, ethical, and material realities of which it is a part. The societal significance of the sealed envelope is not a function of its paper and glue, alone, or the manufacturing processes that produce it; instead, the character of its embedding within a political economy, politics, ideation, institutional infrastructure, and set of practices is an integral part of how it "works." With the transition to email, initially, federal law was reformed to bolster privacy in the absence of a material envelope; gaps in the law left communications vulnerable. Over time, as remote and indefinite storage of email became the norm, the discrepancies between the privacy afforded to communications by postal and electronic mail were viewed with increasing skepticism and ultimately substantially righted, first through litigation and new laws, and more recently through widespread adoption of end-to-end encryption. While the decision to deploy end-to-end encryption was surely made possible due to improvements in technology, it was driven by a renewed realization, among the public and policymakers, of the ethical significance of unencrypted communications born of the Snowden disclosures, which revealed systematic dragnet surveillance of communications by the U.S. government. The new configuration of communication privacy protection set the stage for renewed "technological drama"[6] around law enforcement access and communications privacy, revealing how various configurations alter the *mass* of privacy.

Details aside, this case shows that even as email gains acceptance as a functional replacement for "snail mail," the entangled reality of communications privacy is destabilized. (p. 237) One might argue that email performs the same function as snail mail, namely, communications among users—albeit more speedily. Lacking the equivalent of a physical envelope, the legal protections, and many of the norms and practices that tacitly and explicitly protect against prying into postal mail, however, the value of privacy needed to be reinserted into a system thus newly configured.

The Handoff model is an instrument for performing analyses, such as these, to reveal ethical issues as they emerge and are disrupted in progressive versions of systems where functions are shifted from one component actor to another (or others). The model (1) sharply reveals how functions are distributed to components (human, computational, mechanical) in alternative sociotechnical systems; and (2) interrogates the value propositions captured in these alternative configurations.

The Handoff Model

Provoked by claims about computational systems taking over tasks previously performed by humans, especially tasks thought to require human intelligence, the concept of *handoff* offers a lens through which to scrutinize them in ethical terms. Outside the purview of scholars and social critics, the common practice of delegating functions performed by humans to machines or from machines of one type to machines of a different type, mostly proceeded with little fanfare.[7] Public imagination and anxiety has been stirred, however, with contemporary forms of automation involving AI taking over human roles—machines that can label ("recognize") images, process ("understand") and produce ("speak") natural language and control other machines (robots) anticipate what we will say and do, and make decisions on the basis of these.

Where function shifts from one type of actor to another, and people are inclined to say that the second is performing *the same* function as the first (same function, different actor), we see a red flag. Before racing to the conclusion, we see a dire need for detailed critical analysis that clearly reveals what stays the same, what does not, and how even seemingly irrelevant differences—flesh and blood versus silicon and metal—makes a difference, for the configuration of ethical values embodied in systems in question. The *handoff* lens draws attention to the backdrop of ethical and political values embodied by respective systems—the systems before and after functional handoff. It decomposes the "how" of the function to understand how it is different and what that means for values. (p. 238) It opens our view not only to what might be the same but what may have changed in the reconfiguration of function across component actors.

To begin, the objects of our analysis are complex technical systems comprising diverse functional components. Because the variable nature of these components may include physical mechanisms, embodied computational subsystems, and even humans, the unit of analysis, strictly speaking is sociotechnical systems, a concept we take as given. Indeed, the sociotechnical is what we mean to cover in the balance of this article, though we mostly revert to the term "system" for the sake of simplicity. Abstractly conceived, a system may be defined in terms of its function, in turn achieved through orchestrated subfunctions performed by a system's component parts, in turn, themselves composed of sub-subsystems (or components), and so on. As such, the model assumes that notions of system and component (or subsystem) are relative terms whose application signals the focus of analysis rather than an ontological commitment.[8] By analogy, we may think of the human body as a system and the organs as component parts; but for the cardiologist, the heart is the system of interest and the chambers, valves, arteries, its components, and so on.

A word on terminology: because systems of interest may comprise multifarious parts, including some that are material and others human, we typically use the term *component* as neutral between the two, though occasionally will use "component-actor" to remind the reader (and ourselves) of this variability.

As noted, systems perform functions, and it is the redistribution of these functions that interests us—across versions, either progressive variations over time or contemporaneously competing with one another. What a system's function is, in general terms, answers the question, "what does this system do?" System-components also perform functions, similarly, answering the question, "what does it do?" and also addressing *how* the component-function or subsystem contributes to the function of the system overall. Further, a system's function can be described at varying levels of abstraction: up a level, in terms of its goals, purposes, or even values; down a level, in terms of *how* it does what it does, as a designer or engineer might explain it. It is worth achieving a degree of precision around these levels, distinguishing goals, purposes, and function from the gritty details of how they are achieved. Nevertheless, it is a mistake to think that the higher (p. 239) order outcomes, including values configurations, are insulated from the hows of implementation, or so the Handoff model says.

At the lower level of "how," an analyst explains how components function and how they function together to produce system function overall. To capture the ways components function together, we posit the concept of *acting on* or *engaging* to describe the interaction of one component on another or others. In our lighting example, we imagine darkness falling and a human (component) flipping a switch, in turn causing lamps to illuminate. Using our newly minted terms, the model describes this series of events as a human *acting on* a switch and a switch acting on a circuit, in turn producing an outcome—"turn on the lights." While the human and the physical switch both *act on* other components, respectively, to fulfill the overall function, the model recognizing that there may be significant differences in how they do so, introduces the construct of *mode* (of acting on, or engaging). Not all social and political theories of technology have emphasized what we have called mode; for example, Larry Lessig primarily sought to emphasize the powers people, institutions, software, and machines have in common, namely, the ability to regulate.[9] Others, however, have recognized that the modes of *acting on* performed by human components and machine components, respectively, typically signal disparate forms of moral responsibility.[10]

For the Handoff model, different values for the *mode* parameter may influence or even determine ethical properties of successive versions of a system. Take physical force, a familiar *mode* of acting on. One physically embodied component-actor may act on another, either forcing or preventing action.[11] The human actor, pushing a button, sets off a causal chain of action resulting in car headlights flashing on. Physical ("material") causation, or —one could say—"brute force" may operate in many different ways, for example, a physical component (or set of objects) may act on another component by constraining its range of action (e.g., a safety overlock) without necessarily causing a particular outcome; there could be far more complex causal interdependencies, as when numerous components function together to produce a complex configuration of outcomes on other components, and so on.

A different *mode* of *acting on*—one might say, more subtle—is *affordance*. As defined by the cognitive psychologist J.J. Gibson, affordances are relational properties of things in the environment whose meaning or significance is derived from their service to the needs or capabilities of respective actor-types (humans, other mammals, invertebrates, etc.).[12] When saying that something is nourishing, is a tool, or serves as secure cover, (p. 240) these properties are affordances in relation to actors of particular shapes, sizes, abilities, and needs. Adapting and widely popularizing this idea, Donald Norman urged designers to exploit (not ignore) affordances to create artifacts that people understand and know how to use because well utilized affordances trigger appropriate cognitive and perceptual reactions in humans.[13] Principles derived from Norman's infamous doors and switches have traveled into realms of digital technologies. One approach a social media site could take is to adopt a policy that permits data extraction and offer an application programming interface that affords data extraction, or adopt technical or legal rules (for example, a prohibition on scraping) that discourage it, in relation to actors with relevant technical know-how. Within the Handoff model, affordances are a *mode* of *acting on* that designers can exploit to suggest a range of possible and desirable actions for a system's successful operation. On the one hand, unlike physical force, affordances are perceived and processed by users (humans) who act—often strategically—accordingly; on the other hand, they systematically elicit predictable behaviors.

In our mini case of the light switch, we observe that the human component physically exerts force on a switch thereby initiating a causal chain resulting in the lights illuminating. Among many possible answers to why the human flipped the switch, one of them celebrates the interface design for successfully exploiting the affordance of "flip-ability"; the human flipped the switch instead of pushing or pulling it. Another plausible answer, however, cites purpose: the human flipped the switch because night had fallen. Different, yet, an answer cites obedience to a rule, for example, when a light so switched, say, on a porch, lighthouse, or skyscraper is required by law. The human chooses to act after having identified conditions or pertinent rules, interpreted them, and decided to act accordingly. The human, as it were, as a free agent, is the prime mover causing the lights to turn on by flipping a switch.

Now, imagine lights whose operation is automated via sensors that detect light conditions and a small computer embedded within the light switch. In this case, in given exterior lighting and possibly other conditions, an algorithm expressed in lines of software code implemented in an embodied computer, physically *acts on* relevant components, resulting in the illumination of lights. The software code (and more abstractly, the algorithm) operates like legal rules. The model does not reify them as component actors; instead, their informational content, expressed as coded instructions, is embodied in material, electronic computers, which act on other system components, and so on.

Without delving into metaphysical questions about the nature of free agency, the Handoff model draws attention to features of the scenarios we have sketched, and differences among them, that are relevant to embodied values. Although one might be tempted to say that the automated light switches are performing *the same task* as human operated

switches, the two involve different modes of acting on: one physical causation, the other human agency. This difference makes a difference, for example, in attributing (p. 241) responsibility (or blame) for human initiated versus sensor-initiated illumination. Affordance lies somewhere in between. Though few would say that humans responding to affordances are not necessarily acting freely, the flourishing areas of usability and design in computational systems attest to the sense that responsibility (and blame) may spread across human actor–components and designer-builders of a system. Norman's famous cases of people pushing doors that should be pulled (and vice versa) and other malfunctions communicate this message; informed analysis of the 1988 tragedy in which human operators on the USS Vincenne downed Iran Air Flight 655 with a surface-to-air missile, revealed that the interface was poorly designed.

In sum: Handoff is an analytical model for exposing ethical and political values embodied in technical systems. Deriving its foundations from bodies of work and related concepts in social studies of technology and values in design, it provides further concepts that are particularly important for the rapid deployment of AI both self-standing and within preexisting systems. It targets and challenges the notion—explicit as well as implicit—that component actors are modular, that one can pluck out a human actor and plug in an intelligent component with no further perturbations. The Handoff model offers a cluster of concepts that are potentially useful for exposing aspects of systems that change in the wake of such replacements, that may be relevant to the configuration of values embodied in the resulting systems, and that may remain invisible under standard ways of characterizing technical systems.

The subject matter of a handoff covers versions of systems, either versions that may be vying for dominance or progressive versions that follow one another as systems creators update existing models over time. A *handoff* analysis focuses on variations in different systems that result from variations in components tasked with "the same" functionality and offers great utility in the rapidly growing area of automation with AI, from access security to content moderation to self-driving cars, and a myriad more.

Access Control through the Handoff Lens: A Case Study

To illustrate an application of the *handoff* framing, we walk through the case of secure access to mobile phones, tracking handoffs across five successive system versions—four actual and one foreshadowed by a collaborator's research. We chose this case because, on the one hand, it is familiar to the point of invisibility, yet, on the other, perhaps because of this, the seemingly innocuous "improvements" in ways that each version produces the *same* functionality over its predecessor, elides differences that make a difference.

Below we explore multiple configurations of the access control function. While they are presented and often thought of as innovative improvements to security and usability, the three configurations currently available in the market place (password, (p. 242) finger-

print, and facial recognition) and the underdeveloped passthoughts, their relation to security and usability, among other values, become more complex as well as user and context dependent when viewed through the *handoff* lens.

In the Beginning …

Originally mobile phones did not include a lock built into the material devices themselves. This did not mean they lacked a built-in access control function. As with other phones, access control was a feature of the system, as it were, whose boundaries were more broadly defined; access to landline devices was controlled by their position in homes or offices, and mobile phones, similarly, on one's person, in purses, pockets, or cars.

User-Selected Passwords

As the services and information on phones grew and became more sensitive and revealing, the industry reached a tipping-point and moved to control access to mobile phones through passwords.

Although, increasingly, users are admonished to construct strong passwords, with nonobvious combinations of numbers, letters, and symbols, mixing upper and lower cases, with frequent updates,[14] the current standard is for users to devise their own passwords. Performing—one might say—*the same* function as a purse or pocket, the password controls access to the phone, though arguably, more effectively because while a stolen purse or picked pocket lays bare the phone's function and content along with the material device, not so with passwords.

With passwords providing access control functionality, the human (component)[15] is responsible for setting up the system by creating a passcode and providing it to the operating system (OS) via a numeric keypad. The operating system saves the human-selected inputs. Once a password is in place, the human *component* must accurately remember and enter the selected digits into the keyboard interface to unlock the phone. The phone affords a keyboard that makes password entry easy, but the OS is exacting, demanding that the input perfectly match—be both accurate and complete—the password recorded.

(p. 243) From Password to Fingerprint

In recent years, mobile phone providers have shifted *how* the *function* of access control is implemented—first to thumbprint and more recently to face recognition. As discussed in the following it is unclear whether these shifts result from technical advancement—for example, improved performance of fingerprint and face matching algorithms or usability, or particular security benefits, or a governing U.S. legal framework, or something else entirely.

The fingerprint, a familiar biometric, followed passwords as a subsystem for controlling access to a mobile device. As with passwords, the human *component* initiates the process by entering the print; unlike passwords, however, users *no longer* select this input; rather they *are* the input, as it were, offering up their body part—thumb/finger—as raw material for the technical component, the *reader*. The fingerprint reader creates a mathematical representation of the fingerprint image, or a template, which it stores. To access the phone, users supply the physical stimulus to be checked against the stored template. In Apple's description, the system "creates a mathematical representation of your fingerprint and compares this to your enrolled fingerprint data [the mathematical representation described above] to identify a match and unlock your device."[16] From each successful access usage, it incrementally updates the mathematical representation to improve matching accuracy. The mathematical representations are fungible in that a new algorithm could be used to generate new mathematical representations.

The *mode* of the human acting upon the phone is not physical force but through the affordances of the fingerprint reader, which is able to sense and perform the logical process of comparing input with a stored set of encrypted templates.

This shift also changes the process of accessing the device. Once a fingerprint-generated password is in place, the *human component* must present a fingerprint in a way that is readable to the phone fingerprint reader—not sweaty, wet, swollen or disfigured, dirty, or oddly angled. Because the password is not the finger itself but the phone's stored representation of it, the same finger may provoke different results—access or denial.

In this configuration the phone demands (*mode*) that the human actor present herself in a manner that is legible to the machine. But the technical actor requires the human only to "be herself"—or close enough to it—in a certain way, not to remember something. To gain access, the human must prove to the machine that she is herself, not that she knows a special secret. Unlike the keypad entries in a password configuration, a fingerprint match is not binary, but is probabilistic in that the phone determines in real time whether the mathematical representation of the current fingerprint constitutes a match with the stored mathematical representation of the prior fingerprint.

In this new configuration the human component no longer knows the password; access is tied to a specific human and can no longer be easily transferred, and the human (p. 244) cannot continually replace the input used to generate the access control because an individual's fingerprints are finite.

From Fingerprint to Face ID

In late 2017 Apple introduced Face ID to replace Touch ID—the fingerprint recognition system. Face ID used iPhone 10's new "TrueDepth camera system," which constructs a 3D map of a person's face. TrueDepth's dot projector projects over 30,000 dots onto the face each time an individual looks at the phone, thereby creating and developing its map

of the person's features. The image and the dot pattern are fed through a neural network to generate a mathematical model of her face.

Some of the shifts that occurred between passwords and fingerprints remain—again the human component is an input, and access is tied to a specific human. Unlike a fingerprint reader, however, which requires contact—and therefore is evident to the human, setting aside issues of volition for later—Face ID is a contactless technology. One human can hold the phone and point it at another human, possibly without their knowledge, to access the phone. A human may be an unwitting input into the authentication system that opens up the phones contents and capabilities for someone else.

From Face ID to Passthoughts

Imagine if we could unlock phones merely by thinking a password—*passthoughts*. A prototype of such a system is under development by John Chuang.[17] With this, the function of controlling access moves deeper into the body. Rather than typing a password, or offering a finger, or face, it is an individual's brain activity that becomes the biometric identifier that is authenticated by the system. Like a fingerprint or face image, thinking a thought generates patterns distinctive enough across individuals that they can be used to uniquely distinguish individuals. In the current research prototype, a human user wears a headset with an electroencephalogram (EEG) resting on the brain's left frontal lobe. Thinking a passphrase produces brainwaves that the EEG registers and compares to an earlier passthought. Like other biometrics a "hit" is defined probabilistically and, not accessible to human users, the human may not know, directly, how close a given passthought is to the stored one to unlock the device successfully. An intriguing merger of a chosen password and embodied biometric, a passthought offers the equivalent of two-factor authentication.

(p. 245) Access Control through the Lens of Handoff

A typical narrative might celebrate the evolution of these different configuration of access control in mobile operating systems through these four phases: starting with "primitive" physical constraints to more sturdy, logic-based, combinatorial password protection, to sophisticated biometric facial recognition, and finally, even "smarter," brainwave ID. According to this narrative, progression through each version involves a handoff of function from a component-actor of one type to a different type, each one an improvement over the previous. Instead, the *handoff* approach opens a view to potential ripple effects of such replacements: per the focus of the discussion thus far, different types of component actors *act on* one another differently, and these associated differences may have implications for ethical and political values.

In the case of access control, an important feature of physical deprivation or passwords is an ability of phone owners to determine and control the key, investing them the power to delegate access to others.[18] Despite this similarity, however, a significant difference between the two is that the password system, embedded within the logic of the device OS, implicates the OS developers as additional component-actors, thus expanding the boundaries of the system. Access control performed with biometrics also extends a system's boundary beyond the device itself, but unlike password access, it places the users in a different role in relation to the device, namely, "one-user-one-phone," by restricting use to the individual whose biometric (fingerprint, face, or brainwave pattern) is entered as the original key.

Even in this rather limited case, a *handoff* lens exposes ethical and political differences. In the cases of physical and password restraint, device owners have full sovereignty, so to speak, allowing them to delegate usage to others; they allow for a shared, or collective, resource.[19] The move from "something the user does or knows" (password) to "something they are" (biometric) claimed as a usability improvement that relieves users of the need to remember a secret, curtails agency by diminishing both transparency and dimensions of control. Humans choose a password, subject to OS imposed constraints, enjoy a degree of control and understanding of how it functions and sources of its strength (e.g., length and complexity). With biometrics, the OS defines the password and determines its function. Device owners have lost insight beyond how to present themselves and, even then might not grasp failures to unlock, for example, a system glitch or a finger that is too hot, or cold, or damp, and so forth. Prospective passthought systems would seem further to reduce the degree of control as humans find that thoughts are notoriously harder to control than physical action.

(p. 246) Responsibility

Responsibility and accountability closely tie in with control: an actor may only be blamed for harm—in this case, breaches of security—if he or she had a significant hand in controlling the outcome. Breaches due to password failures may fall on device owners for choosing weak passwords or misguidedly sharing a password with others, or on OS providers for failing to build in adequate affordances for users, who can then generate passwords too weak to withstand computational brute force attacks. In fingerprint and Face ID configurations the OS assumes a specific threat model that precludes physical brute force attacks on an individual's wrist to compel connection between the finger and the phone. With this form of attack, an attacker physically forces the body to move in a certain way; thereafter setting in motion a cause and effect set up by the device and OS manufacturers.

As noted earlier, the lens of handoff challenges the typical narrative of technological progress, which implies that advancing from password to fingerprint to face ID is a steady, linear improvement along the trajectory of security.[20] Similarly recent cases involving law enforcement show that the legal framework governing whether and when

government agents can compel individuals to provide access to their mobile devices[21] does not vary linearly along this trajectory.[22] Although police must obtain a warrant before searching a cell phone,[23] once they have it, whether and when they can compel an individual to unlock it turns on the Fifth Amendment. Admittedly, case law continues to evolve, but at present[24] the majority of U.S. courts have concluded that while fingerprints (p. 247) can be compelled in most circumstances, not so with passwords.[25] Existing precedent distinguishes between production of the body,[26] considered nontestimonial, and acts that reveal the contents of the defendant's mind, which are testimonial. Thus, a fingerprint (and, by implication any biometric) can generally be compelled but not a password.[27] This curious distinction demonstrates that features of component actors, which may not affect direct functionality may nevertheless be decisive in a system's politics.

Privacy and Security

Access control is one mode of constraining information flows—to intruders and other unwanted recipients. Setting aside the unchecked information flows among OS, apps, data brokers, and others, against which access security subsystems offer virtually no protection,[28] it is still possible to compare progressive versions against each other. From physical to password-controlled access, an OS might capture physiological metadata, of sorts, potentially revealing gender, health status, and so forth. Other than that, the password itself, particularly if encrypted on a server, incorporates nothing further.[29] (p. 248) Although a fingerprint places irrevocable identifying information in the hands of the device OS, it might offer great protection against external intruders; according to a 2014 survey, passwords were deployed by only 34 percent of all smartphone users,[30] but by 2016, Apple reported that 89 percent of customers with devices supporting fingerprint unlocking were using it.[31]

In the case of face ID, though also a biometric, its application differs from fingerprint in not requiring physical contact for intended use.[32] This means the device may more easily accommodate unlocking by multiple users, potentially returning to the user some of the control offered by passwords. With increasing interest in biometric identification, generally, facial recognition systems, and availability of facial templates to powerful operators (government and commercial) have increasingly alarmed critics.[33] The extent to which biometric systems inappropriately leak characteristics is not necessarily a function of biometrics but, rather, of a system's design, for example, whether templates and processing of input from sensors is performed on the device or centralized on OS, or other third-party servers. A full account, while necessary for the development of a complete analysis, is outside the boundaries of this chapter.

Articulating the Boundaries of a System

Smartphones no longer rely on access control provided solely through physical deprivation. Although the *handoff* analysis we sketched implies successive, or competing alterna-

tives, today's reality is that dominant mobile operating systems offer more than one of these approaches, allowing users to choose among them. Instead of lessening the need, a *handoff* analysis may reveal to users relevant differences among options. The transition irrevocably tethers access control functionality to the OS provider. Thus, although the user gets to choose among the three (or, potentially four) alternatives, it is the OS provider that chooses whether and what the user gets to choose both by constraining certain actions and by affording them. Where privacy is a value of concern across progressive or (p. 249) competing versions, we have discussed potential pitfalls of alternatives, for example, password versus biometric or fingerprint versus facial recognition. To some extent, however, privacy is partially constructed by relevant legal frameworks and partially in the hands of the OS provider as a function of design choices, such as, whether biometric templates are stored on the device only or also on central servers, whether encrypted or in the clear, and available by whose choices and under what operations.

The *handoff* lens exposes a critical point about the system, as a whole, that may otherwise be obscured. In the transition from physical deprivation enacted by the user to access control internalized as a subsystem of the OS, the boundaries of the system expand accordingly. While, initially, access control resides outside the technical system, progressive iterations expand the boundaries of the system to include the OS provider as a component actor, fully or partially responsible for the functioning of the access control subsystem. Some might view automation, that is, the insertion of AI (or any mechanic component), as a move to eradicate humans from a system (or subsystem); instead, in the effort to characterize shifts in modes of acting due to automation, a *handoff* analysis suggests that describing such moves as *dis*placements rather than *re*placements of agency yields far more productive insights in service of societal regulation of technological development.

Finally, it can be illuminating to consider the *trigger* for two competing or sequential handoff configurations. *Trigger*—the impetus for the reconfiguration of *function*—often highlights specific values that motivated the reconfiguration or are intended to be implicated by it. The shifts from password to fingerprint to face occurred against a backdrop of technological improvements, steady increase in the range and significance of content stored on mobile phones, heightened awareness of the privacy implications of access to that information, and efforts by the U.S. Federal Bureau of Investigation and intelligence agencies worldwide to develop more permissive legal standards for access to the contents of phones and restrict the strength and require backdoors in encryption in consumer products. The range and significance of content stored on mobile phones and the cost of the phones themselves fueled public pressure on companies to limit the utility of stolen phones. So-called "kill switches," which allow a device owner to remotely disable it, were the primary technology developed to depress thefts, but phone-locking measures were viewed as an additional strategy to suppress theft as they depress resale value.[34] With respect to law enforcement access, Apple products and Apple executives have been at the center of the global maelstrom over individual privacy and law enforcement access. Intelligence and law enforcement agencies have pressed governments and companies to provide them with the capability to read the encrypted contents on phones without the

(p. 250) knowledge or assistance of the user.[35] The relationship between these wide-ranging government actions and shifts in password configurations are unknown, yet Apple has been very vocal about the relationship between device passwords, device encryption, and the balance of power between citizens and the government.[36] And Apple has fought efforts to force product design or redesigns to weaken device level encryption.[37]

The *handoff* lens foregrounds the values at play in these various configurations of controlling access to mobile phones.

The goal has been to demonstrate that the lens offered by handoff affords unique and critical insights into the operation of these systems, in terms of new components and modes of acting, that have dramatic consequences for both human and societal values. In our view, this is a critical ameliorative to a focus on the ongoing transition of control into computational components, instead showing the structural, political, and ethical stakes of those changes. We offer handoff with all humility, acknowledging, first, that there are deep issues about systems and contexts of technology development and use that it does not, and may not ever, be able to capture. Second, as a work in progress, there are undoubtedly factors in the myriad handoffs taking place and still coming from humans to machines that the model does not capture. Here, we hope that experiences applying the model—our own and others—will continue to enrich it and expand its explanatory power.

Bibliography

Akrich, Madeleine, and Bruno Latour. “A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies.” In *Shaping Technology/Building Society: Studies in Sociotechnical Change*, ed. Wiebe Bijker and John Law, 259–64. Cambridge, MA: MIT Press, 1992.

Brownsword, Roger. “Lost in Translation: Legality, Regulatory Margins, and Technological Management.” *Berkeley Technology Law Journal* 26, no. 3 (2011): 1321–66.

Flanagan, Mary, and Helen Nissenbaum. *Values at Play in Digital Games*. Cambridge, MA: MIT Press, 2014.

Friedman, Batya. “Value-sensitive Design.” *interactions* 3, no. 6 (1996): 16–23.

Friedman, Batya, David G. Hendry, and Alan Borning. “A Survey of Value Sensitive Design Methods.” (2017). *Foundations and Trends in Human–Computer Interaction* 11, no. 2 (2017): 63–125.

(p. 251) Latour, Bruno. “Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts.” In *Shaping Technology/Building Society: Studies in Sociotechnical Change*, ed. Wiebe Bijker and John Law, 225–58. Cambridge, MA: MIT Press, 1992.

Lessig, Lawrence. *Code: And Other Laws of Cyberspace*. New York: Basic Books, 1999.

Radin, Margaret Jane. “Regulating by Contract, Regulating by Machine.” *Journal of Institutional and Theoretical Economics* 160, no. 1 (2004): 142–56.

Shilton, Katie, Jes A. Koepfler, and Kenneth R. Fleischmann. “How to See Values in Social Computing: Methods for Studying Values Dimensions.” In *Proceedings of the 17th ACM Computer Supported Cooperative Work & Social Computing*, 426–35. Association for Computing Machinery, 2014.

Surden, Harry. “Structural Rights in Privacy.” *SMU Law Review* 60, no. 4 (2007): 1605–29.

Winner, Langdon. “Do Artifacts Have Politics?” *Daedalus* 109, no. 1 (1980): 121–36.

Notes:

([1]) Research for this chapter has been funded by generous support from the US NSF INSPIRE SES1537324 and the MacArthur Foundation. We are grateful to the Simons Institute for the Theory of Computer Science, where both authors were visitors in the Privacy Program, Spring 2019.

([2]) Throughout this chapter, we prefer the acronym AI, connoting decision and control systems based on models derived from machine learning over big data, instead of the terms “intelligent” or “intelligence” spelled out. As such AI has taken on a constructed meaning, and we can sidestep philosophical questions about whether this is intelligence in any normal meaning.

([3]) Bruno Latour, “Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts,” in *Shaping Technology/Building Society: Studies in Sociotechnical Change*, ed. Wiebe Bijker and John Law (Cambridge, MA: MIT Press, 1992), 225–258.

([4]) Katie Shilton, Jes A. Koepfler, and Kenneth R. Fleischmann, “How to See Values in Social Computing: Methods for Studying Values Dimensions,” in *Proceedings of the 17th ACM Computer Supported Cooperative Work & Social Computing* (Association for Computing Machinery, 2014), 426–435.

([5]) See Roger Brownsword, “Lost in Translation: Legality, Regulatory Margins, and Technological Management,” *Berkeley Technology Law Journal* 26, no. 3 (2011): 1321–1366; Margaret Jane Radin, “Regulating by Contract, Regulating by Machine,” *Journal of Institutional and Theoretical Economics* 160, no. 1 (2004): 142–156; Lawrence Lessig, *Code: And Other Laws of Cyberspace* (New York: Basic Books, 1999); Harry Surden, “Structural Rights in Privacy,” *SMU Law Review* 60, no. 4 (2007): 1605–1629; Orin S. Kerr, “Compelled Decryption and the Privilege against Self-Incrimination,” *Texas Law Review* 97, no. 4 (2019): 767; Julie E. Cohen, “Pervasively Distributed Copyright Enforcement,” *Georgetown Law Journal* 95, no. 1 (2006): 1–48.

([6]) Bryan Pfaffenberger, “Technological Dramas,” *Science, Technology and Human Values* 17, no. 3 (1992): 282–312.

([7]) See, for example, Janet Morrissey, “When Robots Ring the Bell,” *New York Times* (November 7, 2018); James Vincent, “Economists Worry We Aren’t Prepared for the Fallout from Automation,” *The Verge* (July 2, 2018), https://www.theverge.com/2018/7/2/17524822/robot-automation-job-threat-what-happens-next; Yuki Noguchi, “Recruiters Use ‘Geofencing’ to Target Potential Hires Where They Live and Work,” *National Public Radio* (July 7, 2017), https://www.npr.org/sections/alltechconsidered/2017/07/07/535981386/recruiters-use-geofencing-to-target-potential-hires-where-they-live-and-work?t=1560452691647.

([8]) Terminology presented a dilemma. We use the generic term *component* to apply to both human and nonhuman parts of the *sociotechnical system*. While the term component does not naturally apply to human *actors*, for our purposes it is important to be able to refer in like manner to human and nonhuman components of a system. Actor-Network-Theory (See, for example, Bruno Latour, “Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts,” in *Shaping Technology/ Building Society: Studies in Sociotechnical Change*, ed. Wiebe Bijker and John Law [Cambridge, MA: MIT Press, 1992], 225–58, which most certainly has influenced us, came up with *actant* as a way out of the dilemma, but our preference is to not adopt theoretical jargon, which can be off-putting for general readers. Going forward, we will mostly stick with the term *component* and sometimes will revert to actor, or subsystem. In addition to human actors and physical objects that can be or constitute system components, we allow for the possibility of groups and institutions as components.

([9]) Lessig, *Code: And Other Laws of Cyberspace*.

([10]) See, for example, Karen Yeung, “The Forms and Limits of Choice Architecture as a Tool of Government,” *Law & Policy* 38, no. 3 (2016): 186–210; Brownsword, “Lost in Translation”; Surden, “Structural Rights in Privacy”; Cohen, “Pervasively Distributed Copyright Enforcement.”

([11]) Remaining at the intuitive level, for the moment, we must look past the fact that there is nothing simple about causation, as Aristotle well demonstrated!

([12]) James J. Gibson, “The Theory of Affordances,” in *The Ecological Approach to Visual Perception* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1986), 127–143.

([13]) Donald A. Norman, “Affordance, Conventions, and Design,” *Interactions* (1999): 38–42.

([14]) Research casts doubt on actual security benefits of these practices. See Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano, “The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes,” in *Proceedings of the IEEE Symposium on Security and Privacy* (Oakland: IEEE, 2012), 1–15.

([15]) We temporarily set aside a key question about the legal relationship between the human actor and the device—the user/owner, or owner who is not the user, or user who is not the owner—all of which may have significance for the composite values output due to legal distinctions.

([16]) "About Touch ID Advanced Security Technology," Apple Support, https://support.apple.com/en-gb/HT204587 (accessed June 14, 2019).

([17]) John Chuang, "Passthoughts: User Authentication Using Brainwaves," http://people.ischool.berkeley.edu/~chuang/passthoughts/ (accessed June 14, 2019).

([18]) In some situations there may be legal constraints on such sharing, but we will set those aside for now.

([19]) One may relate this scenario to the shift from physical books to e-books where the configuration of access is altered, away from traditional personal property to a model that is far more limited.

([20]) The likelihood of false positives—the wrong biometric opening the device—has, according to Apple, been greatly reduced by the introduction of Face ID. Where Touch ID, with a single enrolled finger, had a 1 in 50,000 chance of unlocking with the wrong fingerprint. "About Touch ID advanced security technology," Apple Support, https://support.apple.com/en-gb/HT204587 (accessed June 14, 2019). Face ID, with a single enrolled appearance, has approximately a 1 in 1,000,000 chance of opening with the wrong face. "About Face ID advanced technology," Apple Support, https://support.apple.com/en-us/HT208108 (accessed June 14, 2019).

([21]) A phone user and owner may be distinct, bur for our purposes we focus on the limited case where owner and user are the same.

([22]) For this analysis we consider U.S. law. For a thorough discussion of this issue from conflicting viewpoints, see Kerr, "Compelled Decryption and the Privilege Against Self-Incrimination"; and Laurent Sacharoff, "What Am I Really Saying When I Open My Smartphone?: A Response to Professor Kerr," *Texas Law Review Online Edition* 97 (2019), available at https://texaslawreview.org/what-am-i-really-saying-when-i-open-my-smartphone-a-response-to-orin-s-kerr/; Orin S. Kerr and Bruce Schneier, "Encryption Workarounds," *Georgetown Law Journal* 106, no. 4 (2018): 989–1019; and Laurent Sacharoff, "Unlocking the Fifth Amendment: Passwords and Encrypted Devices," *Fordham Law Review* 87 no. 1 (2018): 203–251.

([23]) Riley v. California (Riley II), 134 S. Ct. 2473, 2495 (2014).

([24]) Under U.S. law an individual accused of a crime can "take the Fifth," and refuse to testify against herself. The Fifth Amendment of the U.S. Constitution declares that "No person shall … be compelled in any criminal case to be a witness against himself," which applies to *acts* that are "testimonial"—have communicative aspects—not just spoken words. There is a good argument that communicating a password to a phone is protected,

and the majority of courts that have examined the issue have reached that conclusion. See Doe v. United States, 487 U.S. 201, 210 n. 9 (1988) (stating in dicta that compelling someone to reveal the combination to his wall safe is testimonial for purposes of the Fifth Amendment); Wayne R. LaFave et al., 3 Criminal Procedure § 8.13(a) (4th ed. 2017) ("[R]equiring the subpoenaed party to reveal a passcode that would allow [the government] to perform the decryption … would require a testimonial communication standing apart from the act of production, and therefore make unavailable the foregone conclusion doctrine.").

([25]) Several state courts have concluded that the Fifth Amendment privilege against self-incrimination does not protect against compelled disclosure of a fingerprint to unlock a seized cellphone, because fingerprints are not a testimonial communication. State v. Diamond, 2018 WL 443356 (Minn. 2018); Commonwealth v. Baust, 89 Va. Cir. 267 (Va. Cir. Ct. 2014); Florida v. Stahl, 206 So. 3d 124, 135 (Fla. Dist. Ct. App. 2016). There are instances where compelling a fingerprint may be testimonial, for example, where it speaks to the ownership of a device as in *In re Application for a Search Warrant*, 236 F. Supp. 3d 1066, 1073 (N.D. Ill. 2017). Holding that compelling production of fingerprints from *all* people present at the execution of a search warrant to unlock seized devices raised Fifth Amendment concerns, but noting that generally ownership is a foregone conclusion and therefore the fingerprint not testimonial. Most recently a federal magistrate judge in the U.S. District Court for Northern District of California concluded that biometrics are testimonial holding that "Government may not compel or otherwise utilize fingers, thumbs, facial recognition, optical/iris, or any other biometric feature to unlock electronic devices," *In re Of*, Case No. 4-19-70,053 KAW, at *9 (N.D. Cal. Jan. 10, 2019).

([26]) Doe v. United States, 487 U.S. 201, 210 (1988) ("[A] suspect may be compelled to furnish a blood sample; to provide a handwriting exemplar, or a voice exemplar; to stand in a lineup; and to wear particular clothing").

([27]) Orin Kerr, "The Fifth Amendment and Touch ID," *Washington Post* (October 21, 2016), https://www.washingtonpost.com/news/volokh-conspiracy/wp/2016/10/21/the-fifth-amendment-and-touch-id/.

([28]) Helen Nissenbaum, "Contextual Integrity Up and Down the Data Food Chain," *Theoretical Inquiries in Law* 20, no. 1 (2019): 221–256.

([29]) Surely passwords can be birthdates, names of children, favorite sports team, etc.

([30]) "Smart Phone Thefts Rose to 3.1 Million in 2013," *Consumer Reports* (April 2014), https://www.consumerreports.org/cro/news/2014/04/smart-phone-thefts-rose-to-3-1-million-last-year/index.htm (accessed June 14, 2019).

([31]) Mikey Campbell, "Average iPhone User Unlocks Device 80 Times per Day, 89% Use Touch ID, Apple Says," *Apple Insider*, https://appleinsider.com/articles/16/04/19/average-iphone-user-unlocks-device-80-times-per-day-89-use-touch-id-apple-says (accessed June 14, 2019).

([32]) One can imagine scenarios for fingerprints that don't require contact by the relevant human—a severed finger or a print manufactured—but those are not the "normal" use case.

([33]) See, for example, Timothy Williams, "Facial Recognition Software Moves from Overseas Wars to Local Police," *New York Times* (August 12, 2015); Catie Edmondson, "An Airline Scans Your Face. You Take Off. But Few Rules Govern Where Your Data Goes," *New York Times* (August 6, 2018); Joshua Rothman, "In the Age of A.I., Is Seeing Still Believing?" *The New Yorker* (November 5, 2018).

([34]) Brian X. Chien, "Smartphones Embracing 'Kill Switches' as Theft Defense," *New York Times Bits Blog* (June 19, 2014), https://bits.blogs.nytimes.com/2014/06/19/antitheft-technology-led-to-a-dip-in-iphone-thefts-in-some-cities-police-say/. Chien describes kill switches and legislation to require them, noting that "[p]olice and tech companies have tried harder over the last year to educate consumers on additional security measures to protect phones, like setting up passcodes, which can make it harder to gain access to devices so that they can be erased and resold."

([35]) See, for example, Statement of Sally Quillian Yates, Deputy Attorney General, Department of Justice, and James B. Comey, Director, Federal Bureau of Investigation, "Going Dark: Encryption, Technology, and the Balance between Public Safety and Privacy," S. Comm. on the Judiciary, 114th Cong. (July 8, 2015). The statement presents the argument that ensuring that technology allows the government to exercise lawful access is "not asking to expand the Government's surveillance authority, but … ensur[ing] that [they can] obtain electronic information and evidence pursuant to the legal authority."

([36]) Richard Lawler, "Tim Cook Outlines Apple's View on Privacy, Encryption in MSNBC Interview," *Engadget* (April 6, 2018), https://www.engadget.com/2018/04/06/tim-cook-revolution-interview/.

([37]) *Id*.

Deirdre K. Mulligan

Deirdre K. Mulligan, Associate Professor, School of Information; Faculty Director, Berkeley Center for Law and Technology, University of California, Berkeley

Helen Nissenbaum

Helen Nissenbaum, Professor, Information Science, Cornell Tech