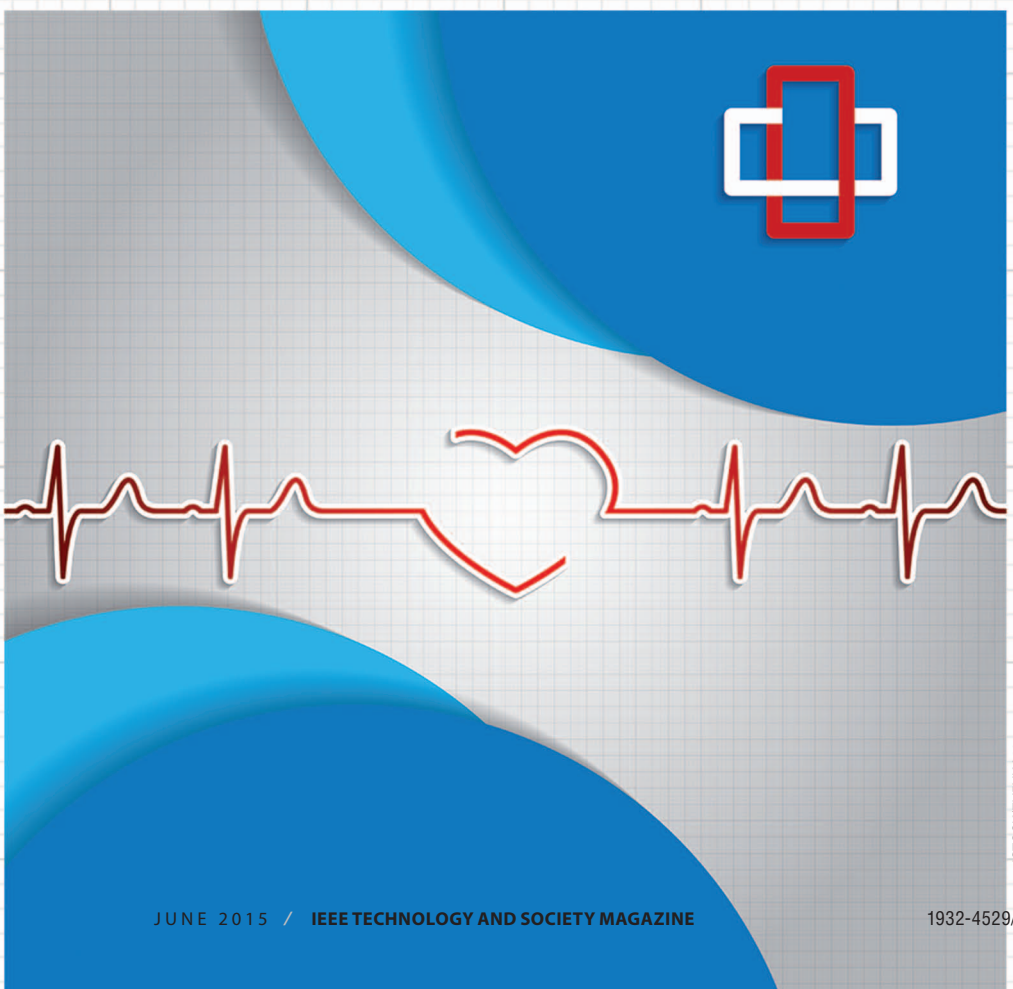
Technology as Moral Proxy

Autonomy and Paternalism by Design

Jason Millar

It is often said that technological artifacts are morally *neutral*, that they are bereft of morality save for whatever we might say about their designers and users. Designers and users, being people, tend to be considered the proper and exclusive focus of our moral attention in the designer-technology-user trio, while technological artifacts, mere tools, cannot contribute anything morally. *Technological neutrality*, a common term for describing this perspective, gives rise to trite statements like, *guns don't kill people, people kill people*. According to technological neutrality, whatever morality is pinned on the artifact is done so in error. Instead, one must refer to the people surrounding an artifact to get an accurate read on the moral claims that can be associated with its use. Thus, technological neutrality supports a strict kind of delineation between people and things: people can be the subjects of a moral analysis; things cannot.

Digital Object Identifier 10.1109/MTS.2015.2425612
Date of publication: 19 June 2015



ISTOCK/EVRYKA23

1932-4529/15©2015IEEE

Recent work in science and technology studies (STS) and the philosophy of technology challenges technological neutrality. STS demonstrates that designers can intentionally embed moral norms into artifacts to achieve certain ends. Such is the case with grocery cart wheels that lock up when taken too far from a store [1], with bridges designed with low clearances to prevent low-income "bus riders" from accessing upscale beaches in New York state [2], and with cars designed to sound annoying alarms when drivers forget to buckle their seat belts [3]. Such cases demonstrate how artifacts can provide "material answers to moral questions" [4]. A shopping cart with locking wheels answers the question *How far from the store should a shopper be allowed to take this cart?* A low clearance bridge answers the question *Who ought to be encouraged to visit these beaches?* Seat belt alarms answer the question *Should the driver be wearing a seatbelt?* In other cases, artifacts can unexpectedly give rise to new moral possibilities and questions while eliminating others [4]. For example, the introduction of ultrasound imaging into obstetrics imposes deeply moral questions on pregnant women that did not previously exist: *Should I agree to an ultrasound exam?* And *What ought I to do if the exam turns up abnormalities* [4]? In all of these cases it can be considered reasonable to describe some of the moral work as being done by the artifacts, not just by the people surrounding them [2]-[4]. But even if we maintain a heavy degree of technological neutrality in our analysis, at the very least one can say that the examples above demonstrate that designers are able to both raise and answer moral questions by embedding moral norms into technological artifacts, and that those norms can impact users' moral lives, often profoundly.

Designers can intentionally embed moral norms into artifacts to achieve certain ends.

One result of these challenges to technological neutrality is that by demonstrating how designers embed moral norms into technology and how those norms impact users, we can begin to draw attention to the complex moral relationship that exists between designers, artifacts and users [1]-[4]. Analyzing aspects of the relationship allows us to more clearly articulate how some choices that are made during the design of a technology shape a user's moral landscape. As such, we can question the appropriateness of certain design choices from within the context of a moral relationship.

This article argues that in cases where technological artifacts provide material answers to moral questions that arise in the use context, the artifacts can and should be characterized as moral proxies acting on behalf of a person. Because of this, we can accurately characterize the moral link between designers, artifacts and users as a *relationship* of a particularly moral kind. Moral proxies of the human kind have been a topic of analysis in healthcare and bioethics, making them a good starting point for thinking about moral proxies of the artifactual kind. I draw from bioethics and STS literatures to build an analogy between human moral proxies in healthcare and artifactual moral proxies. I then turn my attention to design ethics considerations. If we accept that artifacts can function as moral proxies, it becomes important to recognize that designers can subject users to paternalistic relationships that are ethically problematic. I argue that there are situations in which engineers should adopt design practices that avoid paternalism while bolstering informed consent and user autonomy.

Moral Proxies in Healthcare

In healthcare, the ideal patient qua decision-maker is the fully rational adult capable of understanding healthcare information as it is presented to her, weighing that information in the context of her personal life, and communicating her preferences and decisions to others. However, patients often find themselves in imperfect healthcare circumstances.

It is often necessary to turn to a moral proxy for decision-making (also referred to as a *proxy decision-maker*, or *substitute decision-maker*). A moral proxy is a person responsible for making healthcare decisions on behalf of another, which is necessary when a patient is deemed incapable of making, or communicating, his own healthcare decisions. Incapacity of this sort can be age related — a young child cannot make her own decisions regarding care — or can result from a medical condition — a patient might be unconscious, or unresponsive, or generally unable to comprehend the nature of his condition. If a patient is unable either to comprehend the medical information being provided by healthcare professionals, or is unable to communicate his wishes, he cannot give informed consent to medical interventions. A proxy decision-maker is required in such circumstances.

Generally speaking, spouses, family members, or next-of-kin are considered the most appropriate moral proxies given their presumed intimate knowledge of the patient's life and medical preferences [5]. But this was not always the case. The rise of modern medicine

saw physicians assuming decision-making authority in healthcare both in ordinary contexts involving capable patients and contexts involving incapacitated patients [5]. That trend has been reversed. Despite physicians' considerable knowledge and expertise regarding medical options, family members and next-of-kin are once again viewed as the most ethically appropriate proxy decision-makers.

Because healthcare decisions are among the most intimate and private decisions one can make, few people would easily give up the power to make their own healthcare decisions. In practice, turning decision-making authority over to a moral proxy tends to be a last resort only after every effort made to engage the patient in his or her own decision-making has failed.

Proxy decision-making is not without controversy. In Canada and the U.S., courts have found that moral proxies must strive to make proxy healthcare decisions in the best interests of the patient [5], [6]. Those are decisions that "(the patient) would choose, if he were in a position to make a sound judgment" autonomously [7]. Thus, the ideal proxy decision-maker makes no decisions at all: the ideal proxy acts merely as a conduit for communicating the precise wishes of the patient given the current circumstances. Typically, however, proxy decision-makers find themselves in a position of partial ignorance with respect to the patient's wishes, which complicates the decision-making process. Expectations are such that the proxy, often a family member, is required to disentangle his own preferences from those of the ailing loved one, a difficult requirement [8]. It is equally if not more difficult to determine, as an outside observer, whether or not the proxy has met this requirement. In *Re S.D.*, the case quoted above, parents acting as proxy to their severely disabled seven year old boy refused consent for a routine surgery to unblock a shunt that drained spinal fluid from the brain. In their estimation, their son was enduring a life of suffering that the surgery would only act to prolong. Family and Child Service petitioned the courts on behalf of the boy. Their dispute was over whether or not the parents were able to separate their own interests sufficiently from that of the child's, casting doubt on whether or not their decision was in the best interests of the patient. The parents argued that it was, while the Supreme Court of British Columbia found that it was not. So go the controversies surrounding moral proxies.

Pointing out that proxy decision-making is difficult and controversial is not meant to suggest that there is no good way of doing proxy decision-making, or of resolving controversies over what is in a patient's best interests. In the context of this argument, the controversial nature of moral proxies is meant to underscore the moral complexities associated with delegating proxy decision-making powers to someone other than the patient. Though not without its own controversies [9], having the competent patient make her own healthcare decisions is broadly accepted by healthcare professionals, patients and the courts, as morally preferable to having another decide on her behalf.

Technology as Moral Proxy

Consider the following scenario, which is adapted from first-person accounts of living with an Internal Cardiac Defibrillator (ICD). Though Jane is a fictional character, her experiences as described are consistent with those accounts [10]. Jane is at high risk of ventricular arrhythmias, a condition that causes her to unexpectedly experience life-threatening abnormal cardiac rhythms. She found this out a decade ago after having being admitted via ambulance to the emergency room, suffering a heart attack. Jane's cardiologist told her that she was lucky to be alive, and that the symptoms could recur unexpectedly at any time. To increase her chances of surviving similar future cardiac events, her cardiologist recommended that Jane be fitted with an ICD. An ICD is a small implantable device consisting of a power source, electrical leads that are fixed to the heart, and a small processor that monitors the heartbeat to deliver electrical stimuli (shocks) whenever a dangerously abnormal rhythm is detected. The ICD is a small version of the larger defibrillator that the paramedics had used to save Jane's life. Otherwise healthy at the time, Jane agreed to the surgery.

Three uneventful years after her ICD implantation, Jane recalls being in a meeting at work and suddenly feeling lightheaded. She recalls experiencing a sudden "jolt" in her chest, which according to her was the rough equivalent of being kicked in the chest by a horse. The first jolt was followed in quick succession by several others, though she cannot recall how many, each one as traumatic as the first. After several of these shocks her coworkers called an ambulance. Paramedics arrived to find Jane conscious but in shock. In the hospital she was told that her ICD had saved her life after delivering a total of seven series of shocks to her heart. "The ICD performed perfectly," her doctors told her, "it saved your life."

Without the ICD the efforts of several people, and some luck, are required to prevent disaster. Jane could only hope to be in the company of others during a cardiac episode, for starters someone would need to call the paramedics to come help her. The paramedics, assuming Jane is near enough that they are able to arrive in time, must assess the situation, perhaps without the benefit of knowledge of her preexisting heart condition. (Do those who called in the emergency happen to be privy to Jane's heart condition? Is Jane conscious and alert enough to tell the paramedics of her condition when they arrive?) Assuming the paramedics are able to assess her condition accurately they must then prep

both Jane and the defibrillator, and only then can they deliver shocks to Jane's heart. If all goes well, after considerable time, coordination and effort, the human actors have just successfully performed the medical interventions the ICD is capable of performing on its own almost instantaneously. Thus, Jane's ICD all but eliminates the need for humans in the critical path to the medical intervention. It continuously monitors her heart, detects abnormal cardiac rhythms, and delivers potentially life saving shocks before a single human bystander has the chance to come to Jane's assistance.

ICDs are capable of accomplishing the work of several humans, a fact that Bruno Latour refers to as *delegation* (3). Designers delegate the tasks of continuous monitoring, medical assessment and intervention to the ICD. Not only does the ICD perform those functions but also, unlike its human counterparts who are error prone, it performs them with the accuracy and consistency expected of a computer. ICDs also *mediate* our relationship with the physical world, in this case Jane's relationship with time and space (4). With her ICD working diligently in the background, Jane is free to travel farther from medical centers with a certain confidence that in an emergency the time and space between her and medical experts will have less impact on her chances of survival; the ICD promotes Jane's freedom and independence by expanding her geographical safety zones.

In addition to the work that is delegated to the ICD, and the mediating role it plays, answers to deeply moral questions can be delegated to it too (4). An answer to that life and death question that was asked of Jane prior to her implantation — *Would you like to have potentially life-saving electrical shocks administered in the future event that your heart goes into an abnormal rhythm?* — is implicit in the presence of an activated ICD. *Yes! Shock the heart and sustain Jane's life!* The ICD is continuously poised to answer that deeply moral question by the mere fact that it is in Jane, actively monitoring her heart, ready to deliver potentially life-saving electrical shocks at the first sign of an abnormal rhythm. Her ICD instantaneously answers an important healthcare question on her behalf when called upon to act.

The ICD, therefore, can be cast as moral proxy acting on behalf of Jane. In the absence of an ICD, an ideal moral proxy would be required to provide an answer to that same life and death question in an emergency: *Would Jane agree to have potentially life-saving electrical shocks administered given that her heart has gone into an abnormal rhythm?* An activated ICD provides an answer to that question.

To further illustrate how an artifact can function as moral proxy, consider Jane's current situation. Just under a year ago she was diagnosed with inoperable cancer. Her initial prognosis suggested she had four to six months to live, and her health has deteriorated to the point that her medical team have discussed palliative measures with her and her family. As a part of those conversations, Jane was asked whether she would want the medical team to attempt to resuscitate her in the event that her heart stopped. Recalling the intense pain she suffered when her ICD fired years ago, and recognizing the gravity of her current medical condition, she decided that no resuscitation efforts should be made, only that she be kept comfortable.

Of course, her "do not resuscitate" (DNR) preference, a deeply moral end-of-life decision, would be ineffective if the medical team failed to alert all of the medical staff by way of adding the important DNR note to Jane's medical chart. In that case, healthcare staff that were not a part of the DNR conversation would have no way of knowing what Jane's preferences were, and would likely assume proxy decision-making powers and attempt CPR in an emergency. Her DNR preferences would be equally ineffective if no one alerted the ICD to her preferences by deactivating it. In that case, quite unaware of the information on her chart, in the event that Jane's heart went into an abnormal rhythm the ICD would assume proxy decision-making power and deliver its pre-programmed series of up to nine painful shocks. With the moral decision to (or not to) attempt cardiopulmonary resuscitation delegated to it, an ICD functions as an efficient moral proxy the moment abnormal heart rhythms are detected, that is, the moment the question whether or not to attempt resuscitation must be answered.

As it stands, Jane is waiting for the medical staff to contact the ICD manufacturer to assist with the deactivation (hers is an older model that the hospital is unequipped to deactivate). She is told the deactivation could take several days to accomplish. Delegating to the ICD her new end-of-life preferences involves highly specialized equipment and some official paperwork. Jane hopes that her last moments in life will not involve several painful reminders of the powerful little device within her. As a moral proxy, her ICD is proving to be somewhat uncooperative.

An Ontological Objection

It is quite odd, one might say, to describe a technological artifact as capable of participating in any kind of relationship. Artifacts do not have the mental stuff required to decide anything, so it is just a mistake to describe them as decision-makers. Thus, ICDs are incapable of moral action, they merely execute someone else's instructions.

Two responses to this objection rescue the artifactual moral proxy. First, human moral proxies can

fail in their duties when they make characteristically human decisions, when bias, emotions, rational failings (among other factors) prevent them from deciding exactly what the patient would want if she could express her own preferences. Considered this way, ideal moral proxies make no decisions at all: the ideal proxy acts merely as a conduit for communicating the precise wishes of the patient given the current circumstances. It could be that well-designed ICDs and other technological artifacts are capable of approximating ideal moral proxies better than humans.

Second, casting the ICD as a moral proxy in an evaluative framework does not commit one to any strong ontological claims about the inner mental lives of artifacts. When it comes to sophisticated automation technologies — robots and the like — some have argued that the combined effects of their embodiment and behavior require us to place them in an ontological category of their own, "somewhere between object and agent" [9]. I have elsewhere adopted the ontological middle ground and argued that it helps us to more completely and accurately characterize what it is that sophisticated automation technologies accomplish in the use context [10]. But in constructing an ethically informed design framework (which I turn to shortly) I can just as easily remain agnostic on the ontological point. In a design context it is just as effective to treat the artifact *as if* it is a moral proxy, if that will enable an appreciation among engineers, designers, and policymakers, of the fact that automation technologies are capable of impacting users, positively or negatively, along characteristically moral lines. In other words, this possibility of treating the artifact *as if* it is a moral proxy enables a more accurate ethical analysis of tecnology, while rendering the objection misplaced.

Autonomy and Paternalism by Design

Patients have not always been considered the most appropriate authority for making decisions about their own healthcare [5], [13]-[15]. Prior to the latter part of the twentieth century, physicians commonly made deeply personal healthcare decisions on behalf of the patient, a practice now termed *paternalism*. The paternalistic healthcare model saw physicians and other health care professionals assuming proxy decision-making power on behalf of patients, who were told what was the "best" healthcare decision given their situation. Indeed, physicians commonly lied to patients as a matter of principle [15], and intervened in patients' care in direct opposition to the patients' expressed preferences [13].

Paternalism stands in contrast to today's accepted standards of practice. Today, healthcare professionals are seen as having an ethical responsibility to provide an appropriate set of healthcare options for the patient to choose from, and to reasonably counsel patients on the benefits and risks of each option so that the patient can make a *free and informed*, in other words an *autonomous*, decision regarding her own care [11]. Healthcare professionals often have in mind one option that stands out as most appropriate, but it is ultimately the patient's responsibility to select healthcare preferences from among the various options presented to him. Intended to respect a patient's autonomy, this model of free and informed consent (informed consent hereinafter) has become the ethical standard by which healthcare decision-making is judged. For consent to be informed, the patient must first be capable of understanding the implications, the consent must be given voluntarily (free from undue pressure or coercion), the decision must be based on having a (reasonably) full set of options to choose from, and consent must be ongoing (it can change at any time) [16]. (Though informed consent practices and requirements differ slightly between Canada and the U.S., they are similar enough for the purposes of this argument.) Paternalistic relationships are problematic in large part because they act to undermine each of the conditions of informed consent, seriously undermining patient autonomy.

Where does technology fit into the pictures of paternalism and autonomy outlined above? We have already seen how an ICD functions as moral proxy: healthcare decisions are delegated via the artifact's design and settings. Activated, the decision to attempt resuscitation, and the resuscitation attempt are delegated to the ICD. Once deactivated, the ICD both declines resuscitation attempts and respects a DNR decision by not attempting resuscitation. Delegation happens both at the level of the decision and of the intervention. At both levels, some technological correlate is required: to delegate the decision a setting (switch) is required to put the artifact into a particular mode of operation, while delegation of the resuscitation attempt requires more complex functionality (involving electrical leads, algorithms, power sources, and the like) to carry out the operations.

From a moral perspective, the *settings* designed into a technology can be made to do most of the heavy lifting. When technology functions as a moral proxy, the settings that define its mode of operation can reify particular moral decisions; the settings provide the proxy answers to the moral question raised in a use context. As such, an artifact's settings can put it in a mode of operation that corresponds to a patient's preferences, the corollary

being that an artifact's settings can also put it in a mode of operation that confounds a patient's preferences.

Returning to Jane's stubborn ICD, we can say that while it is in a mode of operation Jane has not explicitly endorsed (whether or not it corresponds to or confounds Jane's end-of-life preferences) it is acting paternalistically. We can say this because "paternalism" is meant to describe a particular kind of relationship, and in the case of ICDs there exists just that kind of relationship between a patient and her ICD when the ICD is set without a patient's explicit endorsement. The ICD is capable of intervening in Jane's life and providing material answers to moral questions: it functions as a powerful moral proxy. So long as the ICD is functioning in a mode of operation Jane explicitly endorses, it is acting in Jane's best interests, and the proxy relationship between Jane and her ICD respects Jane's autonomy. When the ICD is functioning without Jane's explicit consent, it acts very much like a healthcare professional poised to assume proxy decision-making authority and impose upon Jane its own answer to what is in Jane's best interests, despite her expressed interests to the contrary. In this latter case, the relationship between Jane and her ICD can accurately be described as paternalistic.

An internal cardiac defibrillator can all but eliminate the need for humans in the critical path to medical intervention.

Thus, there exists a complex moral relationship between designers, artifacts, and users. Designers can embed into artifacts moral norms that offer material answers to moral questions. An activated ICD attempts resuscitation, making it both a technology that promotes resuscitation as a valuable means of extending life, and a technology that answers, *Resuscitate! Yes, shock the heart!* By designing a switch into the ICD (a setting for selecting between modes of operation), designers make the ICD a technology that acknowledges and problematizes the instability of end-of-life decision-making in healthcare. From a purely technical perspective, designers need not have included the switch (it would require further research to determine if designers considered the switch a practical or ethical requirement, or both). From a moral perspective, it is essential to provide a switch. However, the presence of a switch raises an important question about who gets to "set" it and how — who ultimately gets to tell the proxy which way to decide. By designing the ICD in such a way that a patient does not have direct access to the switch, designers embedded another moral norm into the artifact; one that suggests a patient ought not to have direct access to the switch. Each design decision regarding the switch — its presence and who gets access to it — changes the nature of the moral relationship between the artifact and the user.

Further analysis of the presence and nature of an ICD's on/off setting clarifies some ethical aspects of design that engineers should acknowledge in their relationship with artifacts and users. Using the proxy model as a framework for analysis underscores how artifacts can function as deeply moral participants in our lives, as proxies capable of both respecting and usurping our autonomy in surprising ways. A switch, it turns out, is much more than an efficient way of turning things on and off. Without the ability to change the ICD's setting, it would be difficult for the ICD to satisfy the ethical requirement for ongoing consent since the "always on" state of the ICD would eliminate the possibility of changing one's mind. Without the switch it would be difficult to determine in an ongoing manner whose decision is delegated to the ICD: on whose behalf would the proxy currently be acting, the designer's or the patient's?

Add the switch, and suddenly you gain traction on some of the problems associated with paternalistic technological moral proxies. First, you can delegate *different* healthcare preferences to the proxy, enabling the possibility of ongoing consent. You also then have a mechanism for delegating *explicit* healthcare preferences to the artifact. More than the mere presence of the artifact indicating a preference, the setting indicates which preference is the current selection among whatever options are available (resuscitation or no resuscitation in the case of the ICD). Third, the presence of a switch forces one to answer the question, *whom ought to have access to it?* Whenever the relationship between designers, artifacts, and patients is one in which moral proxies are instantiated in the use context via the artifact, the relationship ought to be such that the patient's autonomy is reasonably maximized. Artifacts should not be designed to function in a mode that would subject a patient to a problematic paternalistic relationship. It follows that patients ought to have reasonable access to the settings that put their technologies into modes of operation that provide material answers to moral questions.

How can we apply these three points to a rudimentary design analysis of ICDs? We can say that beyond any practical requirements it might also satisfy, the presence of a switch allowing the ICD to be either active or inactive is an *ethical design requirement*. We can also say that patients ought to have more direct control

over the settings of their medical devices in order to avoid, whenever reasonably possible, situations where technology is functioning in a mode other than one corresponding to the patient's (user's) moral preferences. Placing undue burdens on the patient (user) that might discourage her from modifying settings, or might prevent or delay the changes from being communicated to the proxy (artifact/device), threaten to subject her to a paternalistic relationship with respect to her device.

Handing over more direct control of device settings to users carries an additional benefit, in that it requires user consent to be more properly informed. Whenever a device functions in a proxy decision-making mode of operation, the user surrenders some moral authority to the device. If a user does so unsuspectingly, say because the device was set in a default mode of operation not fully explained to him, then the informed consent requirements of the proxy relationship might not have been satisfied. This is particularly problematic in cases where the user, if properly informed, would have set the device otherwise from the defaults. Such defaults are always somewhat problematic, given that a basic requirement of all autonomous decisions is that they be informed. Thus, if engineers design devices such that they require certain modes of operation to be explicitly activated by the user, the chances of satisfying the moral requirement of fully informed consent increase in the use context. Users will need to be educated as to which design features have ethical implications in the use context, and will need to make explicit choices about them.

In the case of the ICDs or other medical technologies, one might object to the notion of handing over too much control of device settings to the patient on grounds of patient safety. This might be especially so in cases where liability concerns are at issue. Jane's situation (being unable to directly deactivate her ICD) might be the result of detailed discussions in consultation with lawyers, ethicists, healthcare professionals, patients, and other relevant parties, intended to strike an appropriate balance between safety, security, patients' autonomy (the ease with which they are able to change the settings on their ICDs), and making sure they don't do so accidentally, hastily, while incapacitated, or worse yet unknowingly.

It can also be argued, however, that objections based primarily on patient safety are overly paternalistic. The patient safety objection runs the risk of assuming incorrectly that patients are prone to accidents, hasty judgments, incapacitation of some sort, or other failures that can only be avoided by placing an expert in between the patient, whose life is at stake, and the switch the throwing of which would place the patient at risk against better judgment. After all, paternalism in healthcare stems in part from the notion that the patient is less capable than some other person of making good decisions with respect to her own care, and so must be protected. To be sure, devices must be designed to safeguard against accidental harms. Whether or not devices ought to be designed to safeguard against the deliberate, rational, competent actions of those on behalf of who the devices are properly acting is another matter altogether. On the balance, placing gatekeepers in between the patient and certain settings on his device could be as ethically problematic as placing the same gatekeepers between a patient and his human moral proxy.

The presence of a switch allowing the ICD to be either active or inactive is an *ethical design requirement*.

Beyond Medical Devices

My argument has focused so far on a single medical technology: the ICD. However, we find proxy relationships instantiated by nonmedical technologies. Whenever a technology, medical or otherwise, instantiates a moral proxy relationship in the use context, we can evaluate its design using a proxy analysis.

Self-driving cars (SDCs) provide an excellent example for thinking about how non-medical technologies can function as moral proxies. Perhaps the best-known example of an SDC is the one under development at Google [17]. Google's SDC uses a series of sensors, digital maps, databases, and software to solve the extremely complex problem of driving [18]. To date, Google's SDCs have logged hundreds of thousands of kilometers driving autonomously in regular traffic, with only the occasional need for human intervention [18]. In 2011 the state of Nevada authorized the licensing of SDCs for the state's roads [19]. Florida and California passed similar laws [20]. According to Google, their SDCs are safe: "there hasn't been a single accident under computer control" [20]. Many major auto manufacturers are developing SDCs, while almost every auto manufacturer is implementing semi-autonomous features such as parallel parking and collision avoidance systems. It is expected that SDCs will be on the market within the next decade, and some predict the SDC will dominate the roads by 2040 [21], [22].

Consider the following thought experiment: Steve is travelling along a single-lane mountain road in a self-driving car that is fast approaching a narrow tunnel. Just before entering the tunnel a child errantly runs into the road and trips in the center of the lane, effectively blocking the entrance to the tunnel. The car is unable to brake in time to avoid a crash. It has but two options: hit and kill the child, or swerve into the wall on either side of the tunnel, thus killing Steve [23]. If the decision must be made in milliseconds, the computer will have to make the call [24], [25]. What should the car do?

Similar to the situation involving Jane's end-of-life decisions and the state of Jane's ICD (should it remain active or be deactivated?), this ethical dilemma has no objective answer. But the proxy analysis framework applied previously to the ICD helps us to shed some light on questions surrounding this driverless car scenario. Regardless of which path the SDC takes, we can consider the SDC a moral proxy acting on behalf of whoever set it to take that path. If, for example, the engineers at Google programmed the car to keep going straight, in other words to always protect Steve's life in situations where the vehicle is not legally at fault for the situation, then we can say that the SDC is acting as moral proxy on behalf of Google. But this is morally problematic. It might be the case that Steve considers himself morally obliged to risk his own life in such situations, especially ones involving children. Thus, Steve would find himself subject to a paternalistic relationship: a moral proxy acting on behalf of Google would confound his moral preferences. Indeed, according to the proxy model it is always paternalistic to keep Steve out of the decision-making loop, regardless of whether the car's programming corresponds to his moral preferences. Steve is the morally appropriate decision-maker in this driving context since his life, not the life of the manufacturer, is directly at risk. It follows, as in the healthcare context, that his is the autonomous decision the proxy ought to represent, not Google's.

From a design perspective, this suggests that SDC owners (or passengers) ought to be morally responsible for setting the SDC to act one way or the other in situations like the one described. Designers are not the appropriate decision-makers in these scenarios, and so should reasonably strive to build options into SDCs allowing the choice to be left to the user. As is the case in medical contexts, SDC users ought to be informed of the potential moral consequences of their technology choices. Users should then be asked to make autonomous decisions regarding certain cases where it is reasonably foreseeable that the technology will provide material answers to moral questions, in this case risking either the owner's life or the lives of others in particular use contexts. Making that decision on behalf of the user risks subjecting him to a paternalistic relationship in which his autonomy is unreasonably fettered.

Devices Providing Material Answers to Moral Questions

STS provides numerous examples of how technological devices can provide material answers to moral questions in use contexts. When a device reifies a moral decision it instantiates a moral proxy relationship — a role much discussed in bioethics — acting on behalf of whoever set it to operate in that particular mode. There are established ethical norms surrounding proxy decision-making in healthcare and bioethics. We ought to apply those norms to technological moral proxies. Users should have their autonomy reasonably maximized; the devices ought to act explicitly on their behalf. Designers can maximize user autonomy by designing devices such that users are better informed about the proxy implications of the device, and such that users make explicit choices regarding the settings that reify moral decisions.

Placing requirements on designers may seem misplaced. Why should we? The short answer is that designers — engineers for the most part — are already designing norms into their devices as the ICD example suggests. Whatever decision an engineer makes about the ICD's switch carries moral implications. Design is a social and ethically loaded activity, through and through [2]–[4].

What engineers should want to avoid is accidental paternalism in design. Engineers are inescapably linked to the settings and features they embed in their devices, so they should understand the moral dimensions of those links, and design the devices explicitly using some guiding framework. Accidentally creating devices that subject users to paternalistic relationships is undesirable for users and for engineers.

Moral proxies and informed consent practices provide a conceptual framework that can be applied to the design process to help avoid accidental paternalism. Though such a framework does not promise to help predict or prevent all morally problematic design decisions with proxy-like implications (what process could accomplish such a lofty goal?), recognizing the applicability of the proxy analysis framework to design is a step toward better design.

There is, of course, a practical gap in my argument. One should ask, *how should we actually apply these concepts in the design process*? My goal in this article has been to lay the theoretical underpinnings for a practical approach to performing a proxy analysis.

Therefore, I do not provide a detailed practical approach in this paper. I will, however, briefly sketch a possible way forward. Incorporating a proxy analysis into design activities could involve a version of what Verbeek terms "mediation analysis". His proposal sees engineers engaging users in iterative design activities intended to uncover the kinds of moral implications (answers to moral questions reified by the technology) that I have discussed. Once moral implications are identified, for example once certain proxy relationships are recognized as being instantiated in use contexts, they can be analyzed for their impact on user autonomy. *What is the nature of the proxy relationship? Does the user have a reasonable choice of settings that will avoid paternalism? Is the user adequately informed about the defaults we have designed into the technology that instantiate proxy relationships? Is it feasible to design alternative modes of operation into the device to maximize user autonomy with respect to any particular proxy relationship? Do default settings threaten unjustifiably hamper user autonomy?* These are just a few of the questions that could help guide a proxy analysis in design.

At times, technology presents us with ethical challenges that threaten established design practices. As we (engineers and philosophers) have done when faced with other ethical implications of technology — environmental, social, health — we can confront the recognition that technology can instantiate moral proxy relationships head on and accept our role in the designer-technology-user relationship with its full moral weight. That proxy relationships seem to burden designers with added responsibility should not be cause to abandon the design of all technologies that instantiate them, nor does it allow us to cry foul against those who point to them as problems in need of a solution. A middle ground must be sought that acknowledges proxy relationships for what they are, while seeking a design solution that moves us forward. I have argued here that there is a problem in need of a design solution: technology can function as moral proxy, and can sometimes subject users to problematic paternalistic relationships. I have also sketched a path forward: a proxy analysis can be applied in design to help bolster user autonomy where it should be bolstered, and to avoid subjecting users to unjustifiable paternalism by design. We have given up on paternalism as a principle in healthcare and bioethics and have instead adopted autonomy as a replacement — the doctors still have work and patients appear better off. We should reject paternalism by design while promoting autonomy by design for the same good reasons.

Author Information

The author is with the Department of Philosophy, Queen's University, Kingston, Ont., Canada. Email: jason.millar@queensu.ca.

Acknowledgment

The author thanks Ian Kerr and Sergio Sismondo for their many intellectual interactions over the years that have left me a better philosopher, professor, engineer, and person. Peter Asaro and Ryan Calo also offered generous feedback on versions of this work, as did the many participants at We Robot 2014 and the 2014 IEEE International Symposium on Ethics in Engineering, Science, and Technology.

References

[1] I. Kerr, "Digital locks and the automation of virtue," in *From "Radical Extremism" to "Balanced Copyright": Canadian Copyright and the Digital Agenda*, M. Geist, Ed. Toronto, Canada: Irwin Law, 2010.
[2] L. Winner, *The Whale and the Reactor*. Chicago, IL: Univ. of Chicago Press, 1986.
[3] B. Latour, "Where are the missing masses: The sociology of a few mundane artefacts," in *Shaping Technology/Building Society: Studies in Sociotechnical Change*, W.E. Bijker and J. Law, Eds. Cambridge, MA.: M.I.T. Press, 1992, pp. 225–258.
[4] P-P. Verbeek, "Materializing morality: design ethics and technological mediation," *Sci. Tech. Hum. Values*, vol. 31, pp. 361-380, May 2006.
[5] E-H. Kluge, "Consent and the incompetent patient," in *Readings in Biomedical Ethics, A Canadian Focus*, 3rd ed., E-H Kluge, Ed. Toronto, Canada: Pearson Prentice Hall, 2005, pp. 146–148.
[6] N. Cantor, *Making Medical Decisions for the Profoundly Mentally Disabled*, Cambridge, MA: M.I.T. Press, 2005.
[7] *Re S.D.* (1983). 3 W.W.R. 618 (B.C.S.C.).
[8] E-H. Kluge, "After 'Eve': Whither proxy decision making," in *Readings in Biomedical Ethics, A Canadian Focus*, 3rd ed., E-H Kluge, Ed. Toronto, Canada: Pearson Prentice Hall, 2005, pp. 186–194.
[9] R. Calo, "Robotics and the new cyberlaw," *California Law Rev.*, vol. 103, (to be published).
[10] J. Millar, and I. Kerr, "Delegation, relinquishment and responsibility: The prospect of expert robots," in *Robot Law*, R. Calo, M. Froomkin and I. Kerr, Eds. Northampton: Edward Elgar, to be published.
[11] H. Draper, and T. Sorell, "Patients' responsibilities in medical ethics," in *The Bioethics Reader, Editors' Choice*, R. Chadwick, H. Kuhse, W. Landman, U. Schüklenk, and P Singer, Eds. Malden, MA: Blackwell, 2007, pp. 73–90.
[12] A. Pollock, "The internal cardiac defibrillator," in *The Inner History of Devices*, S. Turkle, Ed. Cambridge, MA: M.I.T. Press, 2008, pp. 98–111.
[13] P. Murray, "The history of informed consent," *Iowa Ort. J.*, vol. 10, pp.104–109, 1990.
[14] O. O'Neill, *Autonomy and Trust in Bioethics*. Cambridge, U.K.: Cambridge Univ. Press, 2002.
[15] A.R. Jonsen, *A Short History of Medical Ethics*. Oxford, U.K.: Oxford Univ. Press, 2008.
[16] College of Nurses of Ontario, Practice guideline: consent, 2009.
[17] T. Vanderbilt, "Let the robot drive," *WIRED*, p. 86, Feb 2012.
[18] E. Guizzo, "How Google's self-driving car works," *IEEE Spectrum*, Oct. 2011.
[19] M. Slosson, "Google gets first self-driven car license in Nevada," *Reuters.com*, May 2012.
[20] E. Hayden, "Speeding into the future: Self-driving cars are now legal in California," *Time: Newsfeed*, Sept. 2012.
[21] D. Newcomb, "You won't need a driver's license by 2040," *Wired.com: Autopia*, Sept. 2012.
[22] L. Laursen, "Self-driving car rules will lag tech, think tanks predict," *IEEE Spectrum*, Jan. 2014.
[23] J. Millar, "You should have a say in your robot car's code of ethics," *WIRED.com*, Sept. 2014.
[24] G. Marcus, "Moral machines," *New Yorker*, Nov. 2012.
[25] P. Lin, "The ethics of autonomous cars," *The Atlantic*, Oct. 2013.