Sci Eng Ethics (2014) 20:329–343
DOI 10.1007/s11948-013-9464-1

ORIGINAL PAPER

When Technologies Makes Good People Do Bad Things: Another Argument Against the Value-Neutrality of Technologies

David R. Morrow

Received: 26 April 2013 / Accepted: 13 August 2013 / Published online: 23 August 2013

Abstract Although many scientists and engineers insist that technologies are value-neutral, philosophers of technology have long argued that they are wrong. In this paper, I introduce a new argument against the claim that technologies are value-neutral. This argument complements and extends, rather than replaces, existing arguments against value-neutrality. I formulate the Value-Neutrality Thesis, roughly, as the claim that a technological innovation can have bad effects, on balance, only if its users have "vicious" or condemnable preferences. After sketching a microeconomic model for explaining or predicting a technology's impact on individuals' behavior, I argue that a particular technological innovation can create or exacerbate collective action problems, even in the absence of vicious preferences. Technologies do this by increasing the net utility of refusing to cooperate. I also argue that a particular technological innovation can induce short-sighted behavior because of humans' tendency to discount future benefits too steeply. I suggest some possible extensions of my microeconomic model of technological impacts. These extensions would enable philosophers of technology to consider agents with mixed motives—i.e., agents who harbor some vicious preferences but also some aversion to acting on them—and to apply the model to questions about the professional responsibilities of engineers, scientists, and other inventors.

Keywords Value-neutrality · Technology ethics · Instrumentalism

Introduction

Less than a decade after the Wright brothers launched their first flight, the Italians were using airplanes to kill people in the Italo-Turkish war of 1911–1912. Only a

D. R. Morrow (✉)
Department of Philosophy, University of Alabama at Birmingham, 900 13th Street South, Birmingham, AL 35294, USA
e-mail: davidmorrow@uab.edu

few decades after that, airplanes were instrumental in the Battle of Britain, the firebombings of Dresden and Tokyo, and the dropping of nuclear bombs on Hiroshima and Nagasaki. Few people, presumably, think that early aviation pioneers are responsible for airplanes' military use. After all, they did not build airplanes *for the purpose of* warfare. As many scientists and engineers would say, the airplane is just a "value-neutral" piece of technology; if a technology is used for bad purposes, the blame rests with its users, not with the technology itself.

Scientists and engineers sometimes use this idea—that technologies are value-neutral—to defend unfettered research and development, even of potentially dangerous technologies, such as technologies for intentionally altering the global climate. Those who worry about some technology's misuse ought, on this view, to admonish or constrain those who would abuse it, not those who would invent it.

In this paper, I argue that technologies are not value-neutral in the sense I have just described: in the right circumstances—or perhaps we should say, in the wrong circumstances—a given piece of technology can make good people do bad things. Although there may be other reasons for scientists or society to allow unfettered research into dangerous technologies, the technologies' alleged value-neutrality is not one of them.

My central thesis will come as no surprise to philosophers of technology, many of whom have long denied that technologies are value neutral. Some arguments against value-neutrality target a different kind of "value-neutrality" than I have in mind, showing that values affect the decisions various parties make when developing particular technologies (Van de Poel 2001; Koepsell 2010). Other arguments rely on sophisticated philosophical conceptions of technology to show that technologies, in their sense of 'technology', are not value-neutral, in my sense of 'value-neutral'. For instance, Hans Radder understands a technology as an "artifactual, functional system with a certain degree of stability and reproducibility" (2009, p. 888). Society can maintain the "stability and reproducibility" of an "artifactual, functional system" only by conforming to certain kinds of norms, which vary from technology to technology. So, the very existence of a technology, on Radder's definition, requires conformity to certain norms—or, to put the same thought differently, the adoption of certain values. Still other arguments, which come closest to the arguments in this paper, rely on the idea that certain technologies change the options that are available, attractive, or salient to individuals (Illies and Meijers 2009).

My argument complements and extends the existing arguments against value-neutrality. In the following section, I formulate the Value-Neutrality Thesis more precisely and sketch a general strategy for defending it against some of the existing arguments, including those that are closest to my own. Next, I develop a microeconomic model of technology users' behavior for understanding how a piece of technology affects the world. (Note that this is not a model for evaluating technology's effects in economic terms; it is a model for explaining and predicting how technology did or would alter behavior.) This type of model, which resembles the model developed in Illies and Meijers (2009), might prove helpful in understanding the relations between technologies and society more broadly, independently of concerns about the value-neutrality of technologies. After

explaining the model, I offer my main argument against the Value-Neutrality Thesis. In the final sections, I suggest some ways to extend this microeconomic model and briefly consider what my argument entails for engineers' and other inventors' responsibility for the effects of their inventions.

The Value-Neutrality Thesis

Roughly, to call something "value-neutral" is to say that it is not, in itself, good or bad. To say that technologies are value-neutral—that is, to endorse an instrumentalist view of technologies—is to say that no piece of technology is, in itself, good or bad. In this section, I suggest a more precise formulation of the claim that technologies are value-neutral, and I identify a general strategy that someone might use to defend that claim against some of the existing arguments in the philosophical literature.

Stating the Value-Neutrality Thesis

Various writers have construed the claim that technologies are value-neutral in different ways. The core idea, however, is that the effects of any technology depend on the way that technology is used, which is determined by its (potential) users, rather than by the technology itself. Therefore, if a technology has bad effects, it must be because its users have used it in bad ways. The blame for such bad effects rests with the users, who have been reckless, negligent, or malicious. On this view, the technology—inanimate and intentionless—is not even morally evaluable; it is value-neutral. In other words, technologies do not make people bad; when people do bad things with some piece of technology, it is because of some moral failure on the users' part. It might be the case that a particular technology has no morally acceptable use or that, following Radder (2009), its stable and ongoing use requires conformity to norms that are morally objectionable. Defenders of the value-neutrality thesis can accept the existence of such technologies by insisting that only bad people would use or maintain them. Likewise, it might be the case that a technology makes some morally objectionable options available or more salient than other options, as Illies and Meijers (2009) note, but to use this as an excuse for doing something objectionable might just be to admit to a sort of moral weakness—an inability to resist wrongdoing.

For the purposes of this paper, I formulate the idea that technologies are value-neutral as follows:

> Value neutrality thesis:
> *The invention of some new piece of technology, T, can have bad consequences, on balance, only if people have "vicious" T-relevant preferences or if users with "minimally decent" T-relevant preferences act out of ignorance; and the invention of T can have good consequences, on balance, only if people have minimally decent T-relevant preferences or if users with vicious T-relevant preferences act out of ignorance.*

To say that *T* has good (or bad) consequences "on balance" is to say that *T*'s net impact on the world, taking into account both its positive and negative effects, is good (or bad).

I call a preference "vicious" if merely *having* that preference manifests some vice(s), the very having of which is condemnable. For instance, having a preference for the oppression of some gender or some ethnic group manifests the vices of cruelty and callousness. So, *ceteris paribus*, does having a preference for the harming of some innocent person. We may appropriately condemn someone simply for being cruel or callous. Other vicious preferences might manifest greed, narcissism, cowardice, recklessness, intemperance, laziness, or other vices.

I call a preference "minimally decent" just in case it is not vicious, in the sense defined above. This sets the bar low: "Minimal decency" encompasses not only heroic or particularly admirable preferences, such as a preference for standing up to powerful wrongdoers, but also most run-of-the-mill preferences held by most ordinary people, such as a preference to see one's children succeed in their (innocent) endeavors. It even encompasses traits that might be regarded as minor but widespread failings, such as an unwillingness to routinely make major sacrifices for others, inadequate resolve to stick to a difficult diet or exercise regimen in the face of temptation, or the common kind of impatience that leads to irrationally steep discounting of the future. Since we should not condemn people for these failings, they count as "minimally decent" on this definition. (If you think that some of these particular foibles are condemnable, substitute in your own imperfections for which you think others ought not to condemn you; we all have some.)

I call a preference *T*-relevant (in some context) just in case a person's having that preference affects (in that context) whether or how they use *T*. To illustrate: Where *T* is an umbrella, a preference for staying dry over getting wet is *T*-relevant (when it is raining, at least), whereas a preference for pizza over pastries is not (in most contexts). This is because someone who prefers to stay dry might, when it is raining, use an umbrella to keep dry, whereas someone with the opposite preference would not; but except in special circumstances, whether someone prefers pizza over pastries or vice versa does not affect whether or how she uses an umbrella.

Each half of the Value-Neutrality Thesis includes a clause about acting from ignorance. Someone with minimally decent *T*-relevant preferences could act out of ignorance in using *T* to bring about bad consequences, and someone with vicious preferences could ignorantly use *T* to bring about good consequences. For instance, the inventors and users of the anti-nausea drug thalidomide may have had minimally decent preferences, but because of their ignorance, the drug's disastrous side effects greatly outweighed its benefits. Someone could shoot a random stranger for the thrill of it without realizing that the stranger was about to bomb a market. In general, the sorts of ignorance that might apply here include both the kinds of ignorance that Aristotle discusses—e.g., ignorance of what one is doing, to whom one is doing it, with what one is doing it, etc.—as well as moral ignorance—e.g., about the true nature of well-being. In all of these cases, we can attribute the good or bad effects of the technology to its users' ignorance rather than to the technology itself.

Defending the Value-Neutrality of Technologies

Given this formulation of the Value-Neutrality Thesis, any given technology can be value-neutral even if it has, on balance, bad effects—and even if one could have foreseen those effects in advance. Even something as diabolical as the Nazi gas chambers, for which morally innocent uses are hard to imagine, might be value-neutral on this formulation. Their horrific consequences depended on the vicious preferences of Hitler and his subordinates; if no one had had vicious preferences, the gas chambers would never have been built, much less used.

The basic argument for the Value-Neutrality Thesis is as follows: The mere invention of a piece of technology has no effect on the world except through some agent's use of the technology (or through others' reactions to people's potential use of the technology). When an agent uses an invention in a particular way, he or she intends to bring about some effect. If the effect is, on balance, good, and the agent knew that the effect would be good, then the agent has minimally decent preferences, for desiring good outcomes is minimally decent. If the effect is, on balance, bad, and the agent knew that it would be bad, then the agent has vicious preferences, for desiring bad outcomes is condemnable. If the effect is good (or bad), but the agent did not know that it would be good (or bad), then we should attribute the technology's good or bad effects to the agent's ignorance, not to the technology itself. This covers all of the possible cases: Either the effect is good or bad, on balance, and the agent either knows or does not know the net impact of using the technology. In all of these cases, we should attribute the goodness or badness of the technology's effects to the preferences or ignorance of the user, not to the technology itself. None of this depends on any specific features of any particular technology, and so the argument applies to all technologies. Thus, technologies are value-neutral.

A Microeconomic Model of the Effects of Technologies

To argue against the Value-Neutrality Thesis, I want to sketch a microeconomic model of the ways in which technologies affect individuals' behavior and, through their behavior, affect society for good and for ill. Like the model developed by Illies and Meijers (2009), the model is built on the idea that technologies mediate people's behavior by changing the attractiveness of various options. Unlike Illies and Meijers' model, which relies on an expansive, qualitative notion of "Action Schemes," the present model uses the tools of microeconomics, including decision theory and game theory, to understand, explain, and predict individuals' behavior. While this may be more restrictive in some ways, it offers a simpler, more analytically tractable approach that can easily be scaled up to consider the aggregative and interactive effects of many people's actions. It also allows us to tap into a vast existing body of knowledge about the effects of changing incentives. The main arguments in the next section, for instance, come directly from standard analyses of issues in game theory and behavioral economics. Most of the arguments in the following sections, however, could be reconstructed in Illies and Meijers'

terminology. Reconstructed in this way, those arguments would amount to an extension and precisification of Illies and Meijers' vague argument that technologies are "morally relevant," rather than value neutral, because they alter people's Action Schemes (2009, pp. 437–38).

Let the term 'invention' refer to any technological innovation that enables some people to achieve some goal(s) more effectively or efficiently than before, or enables some people to achieve some goal(s) that were previously unachievable. Inventions, in this sense, cover everything from the proverbial "better mousetrap" to a better Web browser to a spacecraft that enables interstellar travel. Certain things that are intuitively "inventions" are excluded by this definition, such as new but inferior mousetraps. Such inferior inventions are largely irrelevant to the argument in this paper because few people will use them, and so they will have little impact on society. Other things that intuitively may not be inventions, such as new business processes, are included in the definition, but nothing in the argument hinges on the inclusion of such processes.

For present purposes, we can assimilate the kind of invention that enables people to achieve goals that were previously unachievable to the kind that enables people to achieve goals more efficiently. We can do this by thinking of unachievable goals as goals whose attainment requires infinite resources. An invention that enables the achievement of such a goal does so by making it achievable with finite resources. This is a case of enabling the achievement of a goal more efficiently: What used to require infinite resources now requires fewer resources. This has the counterintuitive implication that impossible things, such as time travel, are to be regarded as technically possible but requiring infinite resources; but since this is not meant to be taken literally, there is no harm done by pretending as if it were true to simplify the model.

Thus, without loss of generality, we can say that an invention lowers the cost of achieving some goal(s). The term 'cost' here must be construed as broadly as possible. A more efficient combustion engine lowers both the financial and environmental cost of driving a car. A microwave oven lowers the cost of cooking certain foods by reducing cooking time. A more effective chemotherapy drug lowers the cost of treating cancer by reducing the number of nauseating treatments that are needed to eliminate a tumor—even if the purely financial cost of using the drug is the same as the cost of alternative therapies. The internet lowers the emotional cost of moving far from friends and family, since it makes it is easier to communicate with them, as well as the emotional cost of saying immature, hurtful things to total strangers.

Another way to express the idea that inventions lower the cost of achieving a goal is to say that they reduce the resources required to achieve that goal. Again, the term 'resources' must be construed broadly, so as to include money, time, emotional stamina, and so on. Since those resources could be used to do other things, lowering the amount of resources required to achieve a goal not only makes it cheaper *simpliciter*, but it also makes it cheaper relative to the achievement of other goals. For instance, suppose that a musical theater lover, Iris, lives three hours from New York City. Going to see a show on Broadway therefore requires over nine hours—three to travel into the city, three to watch the show, and three to travel home. Going

to see a local community theater production requires about three hours. Suppose now that someone invents a new, high-speed train that connects Iris's town to New York, cutting the travel time to one hour. Seeing a Broadway show has become cheaper for Iris, in terms of time, relative to seeing a community theater production.

Drawing on microeconomic theory, we can use this insight to explain how inventions change people's behavior. Microeconomic theory assumes that, *ceteris paribus* and within limits explored by behavioral economics, people are more likely to do something (or will do more of it) when it becomes cheaper relative to their other options. We can model this by thinking of people as utility maximizers: Suppose that Iris gains a certain amount of utility from seeing a Broadway show, gains some lesser amount of utility from seeing a community theater show, and loses some utility for each hour spent traveling. If we assume that she wants to maximize her utility, she will choose the community theater over Broadway just in case the utility of the community theater show is greater than the utility of the Broadway show minus the cost (in utility) of the long trip to and from New York. By reducing the travel time required to get to a Broadway show, the high-speed train could cause Iris to change her behavior: She might value community theater more than the Broadway show when the latter requires six hours of travel, but less when the latter requires only two hours of travel.

In general, then, we can explain an invention's impact on the world as follows: An invention changes the costs (broadly construed) that individuals must pay to perform some activity or bring about some outcome. Keeping constant the individuals' preferences over various activities and outcomes and assuming that individuals behave as if they were utility maximizers, these changes in cost sometimes incentivize the individuals to choose different activities and outcomes than they did before. By changing their choices, individuals affect the world for better or worse. Thus, the invention of a new technology causes changes in the general welfare, but the nature of those changes depends on individuals' preferences and choices.

A few comments are in order:

First, this microeconomic model does not entail technological determinism or the autonomy of technology. In the theater example above, the invention of the high-speed train changes Iris's behavior, but only because of Iris's preferences. This new technology is not controlling Iris, forcing her to go to Broadway against her will, or subtly altering her psychology. It is simply changing her incentives so that, given the same preferences as before, she now chooses something different.

Second, although some people might object that a microeconomic model assumes psychological egoism, nothing about the model requires this assumption. The model only requires that we incorporate all of an individual's altruistic and other concerns into his or her utility function, which we can interpret as a theoretical posit that simply reflects his or her preferences, rather than as a direct measurement of some kind of personal satisfaction. Thus, the model allows us to suppose that someone's utility is maximized when she devotes himself almost entirely to helping others.

Third, inventions will often have important strategic effects, not all of which benefit everybody involved. For instance, consider how the microeconomic model

explains the impact of the mechanical tomato harvester that Langdon Winner (1986) discusses. The mechanical harvester greatly reduces a farmer's cost—in money and time—required to harvest a ton of tomatoes, provided that the farmer grows enough tomatoes to make the harvester worth buying. This allows large-scale farmers to sell tomatoes more cheaply than before, depressing the market price for tomatoes. Thus, when large-scale farmers adopt the harvester, it greatly *increases* the cost, in terms of time, to small-scale farmers of picking, say, a hundred dollars' worth of tomatoes. This is because the small-scale farmer, without the benefit of the harvester, must pick more tomatoes, at the same speed, to harvest a hundred dollars' worth than he or she did before. The large-scale farmer's adoption of the harvester could therefore make it more worthwhile for the large-scale farmer to grow tomatoes and less worthwhile for the small-scale farmer to do so. This would explain the dramatic consolidation of tomato-farming that Winner attributes to the harvester's adoption.

These strategic effects, along with certain human limitations highlighted by behavioral economists, fuel my main argument against the Value-Neutrality Thesis.

The Argument Against the Value-Neutrality Thesis

To see why the Value-Neutrality Thesis is false, consider the following two types of cases in which "minimally decent *T*-users"—i.e., agents who have no vicious *T*-relevant preferences—use some invention, *T*, in ways that have bad consequences on balance.

Collective Irrationality: Collective Action Problems

The first type of case involves collective action problems. Roughly, a collective action problem occurs when a group of agents fails to achieve some attainable and collectively desirable goal(s) because the individual agents cannot muster the cooperation necessary to make the pursuit of those goals individually rational for enough members. More precisely, suppose that each member of a group must choose between two types of behavior—which we might call "cooperative" and "uncooperative" behaviors—or choose a behavior along some spectrum whose endpoints we might call "cooperative" and "uncooperative." Each member of the group benefits most when (or to the extent that) most or all members of the group behave cooperatively. However, given the proportion of group members who are currently cooperating, the cooperative behaviors carry costs to the individual that exceed the individual's benefit from his or her own cooperating. Thus, given the current proportion of cooperators, it is not economically rational for any single individual to behave cooperatively, and so too few people cooperate. Unless something can be done to induce a greater proportion of people to cooperate, each individual remains worse off than he or she would be if all or most people cooperated (Olson 1965).

Some inventions can help overcome collective action problems; some can create new ones. Consider, for instance, the invention of more efficient fishing technologies. It is difficult to overfish a large fishery using traditional fishing

methods; it is quite easy to do so using industrial fishing technology. The result is a type of collective action problem often called a "tragedy of the commons" (Hardin 1968). In the long run, the entire fishing community is better off if the total annual catch remains low enough to allow the fish population to rebuild itself each year. To keep the total catch below the relevant threshold, each fishing vessel must "cooperate" in catching no more than its share, which requires catching fewer fish than it could. If, however, too many fishing vessels exceed their share, then no vessel has an economic incentive to limit itself: Cooperating would bring no long-term benefit, given that the others are driving the total catch above the sustainable level, but it would mean giving up the short-term benefit of catching more fish. The predicted (and frequently observed) result is overfishing, which has worse consequences for everyone, on balance and in the long run, than fishing at the maximally efficient level.

Other inventions induce other kinds of collective action problems. Cars induce alienating suburbanization, but bucking the trend of suburbanization brings few benefits unless others do so as well. The invention of cheap fossil fuel energy is leading to climate change, a different kind of tragedy of the commons. The invention of nuclear weapons created one of the classic collective action problems: If the great powers all built nuclear arsenals, the chances of collective annihilation increased, but if no one else built a nuclear arsenal, any country could become a great power by building one itself.

The defining feature of these technology-induced collective action problems is that they involve a behavior, made possible by a new invention, that yields net gains for an individual unless many people are behaving in the same way, but the cessation of which brings no net benefit to the individual who ceases. These problems arise when an invention motivates each individual to begin behaving "uncooperatively." As more people adopt the uncooperative behavior, the aggregate benefit to each individual declines and eventually becomes negative. Once a large group is behaving uncooperatively, however, each individual would lose even more by being one of the few cooperators. Thus, an entire group can become trapped in an equilibrium that is, on balance, worse than the one that existed before the new invention appeared.

Individual Irrationality: Discounting

Most humans discount future benefits relative to present benefits: *Ceteris paribus*, they prefer receiving some benefit sooner rather than later. Humans commonly discount the future steeply enough that they make themselves worse off in the long run by weighting short-term benefits and instant gratification more heavily than larger long-term benefits (Green and Myerson 2004). By increasing the short-term benefit from some activity, some inventions can induce people to switch from an activity with greater net benefits to one with lesser net benefits.

Consider, for instance, the introduction of television. Many people find watching television—even mediocre television—to be quite entertaining. The invention of television, therefore, greatly increased the (short-term) utility such people could expect to receive from spending a night at home, passively staring at a box. Contrast

that with other activities with which people might fill their evenings: developing a talent, such playing a musical instrument; participating in a community organization; mastering another language; exercising or playing sports; etc. These activities may promise fewer short-term rewards than television. Learning to play a new instrument or speak another language can be frustrating or dull. So can attending community organization meetings. Yet, the long-term rewards of developing a cherished talent, building social capital, improving one's community, traveling in foreign countries, and maintaining one's physical health are much greater than the long-term rewards of watching most television. If people discount the future steeply enough, however, the short-term rewards of watching television will trump the long-term rewards of those other activities, and the introduction of television will induce them give up those other activities. The sociologist Robert Putnam argues that this phenomenon is partly responsible for the erosion of social capital in the United States in the late twentieth century (2000, pp. 216–46).

In a slightly different context, the invention of certain kinds of food products or processing methods induce similarly short-sighted behavior. Innovations in agricultural production and processing have made it much cheaper to produce meats, highly sweetened drinks, and other high-calorie foods. Consumers—sometimes out of ignorance but often because of steep discounting of future health costs—happily gorge themselves on ever growing portions of cheap junk food. The long-term costs, like many people's waistlines, are enormous and growing.

In many cases, collective action problems deepen discounting-related ones. As more and more people stay home with their televisions and aimless Web browsing, the short- and long-term rewards of community participation decline, increasing everyone's incentive to stay home. As improved agricultural techniques improve yields of, e.g., corn or chickens, thereby lowering prices, each farmer must produce more to maintain his or her own income, thereby lowering prices even further, inducing people to consume even more of it.

Vicious Preferences?

Friends of the Value-Neutrality Thesis might object that the two kinds of cases I identified above depend on vicious preferences. Collective action problems arguably depend on a vicious unwillingness to sacrifice for the public good. Discounting-related problems arguably depend on a vicious "pure time preference"—i.e., a condemnable impatience that is not rooted in anything as reasonable as the time value of money. If people did not have these preferences, then neither type of problem would arise. The preferences are vicious, one might argue, precisely because they lead to individually and socially suboptimal outcomes. Thus, the kinds of problems I identified cannot undermine the Value-Neutrality Thesis because they involve bad outcomes that arise from vicious preferences.

Critics of the Value-Neutrality Thesis could respond in either of two complementary ways. One option is to insist that these preferences are not condemnable. On my characterization of the Value-Neutrality Thesis, a preference counts as vicious if and only if it is appropriate to condemn someone for simply having the preference. A willingness to sacrifice one's own welfare for the public good is often

laudable, and a total unwillingness to sacrifice any amount of one's own welfare for any amount of the public good is condemnable. As many philosophers and undergraduates have argued, however, morality can only demand so much from us; we may sometimes be justified in pursuing our own self-interest or the interests of those closest to us even when doing so does not maximize aggregate utility.[1] Justifications for this vary: Consequentialists may argue that bringing about the best consequences requires devoting ourselves to specific people or projects (Railton 1984). Non-consequentialists may argue that integrity sometimes requires us to cling to our own projects or ideals even when the greater good lies elsewhere (Williams 1973) or that we have the moral prerogative to prefer our own interests over others to some extent or in certain ways (Kamm 2006). Care ethicists may argue that attending to our relationships (Held 2006) or particular others (Noddings 1984) is more important than constant concern for the general welfare. In short, to insist that morality demands constant sacrifice of our own good for the general welfare, even in the face of a collective action problem that renders an individual's contribution all but useless, is to pay a very high price for the Value-Neutrality Thesis, philosophically speaking.

Pure time preference may be hard to defend as a matter of prudence, but it is not clear that it is a moral failing. It is, rather, a psychological foible that sometimes leads us to make poor choices. The question of the *value*-neutrality of technologies is not about whether *causal* responsibility for technologies' bad effects rests with the technologies or their users, but whether the *moral* responsibility is always traceable to some moral failing in the users.

Turning to the second defense of the cases presented above, one might argue that even if the relevant preferences are condemnable, they are widespread and ineradicable. Nearly all of us exhibit them from time to time; only in a population of moral and rational saints would they be completely absent. To dismiss collective action problems and discounting-based problems as irrelevant to the value-neutrality of technologies is to insist that technologies would be value-neutral in a world populated by moral and rational saints. Such failings' influence on our actions, therefore, might be seen as part of the background against which technologies operate, just like our biological weaknesses. To blame those failings for a technology's effects, then, is like insisting that the problem with a toxic insecticide is not in the chemical, but in humans' susceptibility to it; while it may be true in some sense, it unreasonably ignores the relevant facts about the beings who will be using the technology. Similarly, it is unreasonable to insist that technologies are value-neutral just because there is some possible world in which humans, equipped with vastly superior moral and intellectual powers, could avoid the bad consequences that result from human use of technologies.

[1] This opens the door to an even more general criticism of the Value-Neutrality Thesis. It is arguably the case that, within limits or in certain contexts (e.g., competitive markets), it is not vicious to prefer to impose costs on others to benefit oneself, one's family, one's country, etc., rather than to forgo those benefits to protect others. In that case, if the costs to others exceed the benefits to the *T*-user (etc.), then the *T*-user could create, on balance, bad effects without acting on vicious preferences. Especially when *T* is generally used by a small elite at the expense of the rest of society, *T*'s aggregate effect may well be large and negative, despite each individual *T*-user's preferences being minimally decent.

In short, the objection that vicious preferences underpin the cases described above fails. The relevant preferences are arguably minimally decent. Furthermore, given the ubiquity and ineradicability of limited selfishness and pure time preference, the Value-Neutrality Thesis becomes unimportant if it means that technologies would not necessarily have bad effects in a world of moral and intellectual superhumans.

Extending the Microeconomic Model of Technological Impacts

This paper focuses on a strict reading of the Value-Neutrality Thesis, on which an invention can have bad effects, on balance, only if its users have vicious relevant preferences. Thus, the paper has focused on cases in which an invention can lead to bad effects even in the absence of vicious effects. The paper's microeconomic model of technologies' social impacts could be extended in various ways to address other kinds of technology-induced problems. While the models in this paper assume minimally decent agents, more realistic models, incorporating agents with more mixed motives and abilities, might provide fruitful avenues for future research. Both of the more realistic models sketched below go some way toward answering the call for an approach to technology ethics that provides a strong theoretical understanding of the interactions between technologies and society (Brey 2010).

One possible extension of the model involves assuming that most agents in society have a genuine but limited aversion to behaving unethically. We could model this by inserting a penalty term into an agent's utility function that applies only when (or to the extent that) the agent acts unethically. This penalty term would represent the agent's degree of conscientiousness. Assume that given the current state of technology, some agent, *S*, behaves ethically because *S*'s conscientiousness reduces her utility from acting unethically below her utility from acting ethically. By increasing the reward from or decreasing the cost of acting unethically, however, a new invention might induce *S* to act unethically. This allows us to model individuals who have some vicious preferences but also have some inclination to resist those preferences, yielding a far more realistic picture of human nature. By building conscientiousness into the model, we can distinguish the role that a particular invention plays in creating bad consequences from the role that the vicious preferences play.

The second extension involves generalizing the second reply to the objection in the previous section. Given that many people do have vicious preferences of some sort, it is sometimes *in principle* possible to predict that an invention will make the world worse off.[2] How often this is possible in practice—especially in the long

[2] This predictive ability is one way in which the microeconomic model's analytic tractability is an important advantage over Illies and Meijers' qualitative approach. Consider, for instance, the case of the energy efficient light bulb (Illies and Meijers 2009, p. 436). Both approaches can explain why introducing such a bulb might actually raise total energy consumption. Given the relevant supply and demand schedules, the microeconomic model can predict the sign of the net effect. That is, it can predict whether the increased use of electric lights will swamp the increased efficiency of the bulbs. Illies and Meijers' model cannot; its strengths lie elsewhere.

run—is a separate question, which I will not attempt to answer here. More common, I suspect, is the kind of situation in which one can predict which of several designs for a particular invention will have the best (or least bad) effects, given people's actual preferences. As a very simple example, imagine an online service that provides information about political candidates' records, policies, etc. Given that the average voter seems to attach relatively little value to learning about political candidates—at least as compared to doing other things—it is fairly obvious that the average voter will not use the service if it provides the information in a way that requires a fair amount of time or technical expertise to understand; it is also fairly obvious that better organized or better funded groups, such as vested interests trying to identify candidates to support or oppose, would use the service. By contrast, if the service made it easy for the average voter to find and understand information about candidates, some voters would use it. The former, hard-to-use option further empowers vested interests; the latter empowers the average voter. Assuming that empowering the average voter leads to better (e.g., more just) outcomes than does empowering vested interests, a seemingly innocuous change in how an information service presents its data can change how that technology affects society. Using the microeconomic model of technological impacts to understand the likely impact of an invention, given people's actual preferences, would allow philosophers of technology to apply the model to questions about engineering ethics and similar subdisciplines.

Inventors' Ethical Responsibilities

What does the preceding argument entail about inventors' ethical responsibilities? Some engineers, scientists, and other inventors might fear that by giving up the Value-Neutrality Thesis, they are committing themselves to some kind of societal control—or at least, to some kind of self-censorship in their research and development of new technologies. The need for social oversight of research and development, however, does not follow directly from the rejection of the Value-Neutrality Thesis; further argument is needed. Even if the Value-Neutrality Thesis is false, it might be a bad idea for society to hold inventors liable for their inventions' consequences or to expect or require inventors to perform some kind of cost-benefit analysis before developing new technologies.

There are two reasons to think that a policy of mostly unfettered research and development might be best. First, there are the tremendous epistemic difficulties involved in predicting the effects of a particular technology. Not only would we need reasonable estimates of current people's relevant preferences and of the impact of the users' new behavior, but we would need to know how the invention would affect behavior far into the future, often in interaction with other technologies that cannot now be anticipated. Thus, attempts to guide decision-making in light of some kind of cost-benefit analysis may be futile.

Second, given these epistemic difficulties, any policy on which society penalized inventors when their technologies proved detrimental would have a major chilling effect on technological development. What reasonable engineer would develop a

new technology if she feared that society might misuse it—often in ways that she could not possibly envision—and then punish her for their abuse of her invention?

These arguments are not necessarily decisive. Especially in light of the arguments in the preceding sections, opponents of increased social control over research and development will need to work harder to justify their position. Still, the falsity of the Value-Neutrality Thesis does not *directly* entail particular answers to the major questions of engineering ethics.

Conclusion

Scientists and engineers often disagree with philosophers of technology over whether technologies are value-neutral. The disagreement, however, often focuses on vague notions of value-neutrality and "moral relevance." By focusing on a precise and strong version of the Value Neutrality Thesis, we can see just how hard it is to show that a technology's users are sometimes blameless for the technology's bad effects. The arguments in this paper provide two kinds of cases in which a new technology can induce minimally decent users to bring about bad effects. These cases undermine even the strong version of the Value Neutrality Thesis. Technologies really can make good people do bad things.

Acknowledgments Thanks to Chris Alen Sula and two anonymous referees for helpful comments on earlier drafts of this paper.

References

Brey, P. (2010). Philosophy of technology after the empirical turn. *Techné: Research in Philosophy and Technology, 14*(1), 36–48.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130*(5), 769–792.

Hardin, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248.

Held, V. (2006). *The ethics of care: Personal, political, and global*. New York: Oxford University Press.

Illies, C., & Meijers, A. (2009). Artefacts without agency. *The Monist, 92*(3), 420–440.

Kamm, F. M. (2006). *Intricate ethics: Rights, responsibilities, and permissible harm*. New York: Oxford University Press.

Koepsell, D. (2010). On genies and bottles: Scientists' moral responsibility and dangerous technology R&D. *Science and Engineering Ethics, 16*(1), 119–133.

Noddings, N. (1984). *Caring: A feminine approach to ethics and moral education*. Berkeley, CA: University of California Press.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

Radder, H. (2009). Why technologies are inherently normative. In A. Meijers (Ed.), *Handbook of the philosophy of science, vol. 9: Philosophy of technology and engineering sciences* (pp. 887–921). Amsterdam: Elsevier.

Railton, P. (1984). Alienation, consequentialism, and the demands of morality. *Philosophy & Public Affairs, 13*(2), 134–171.

Van de Poel, I. (2001). Investigating ethical issues in engineering design. *Science and Engineering Ethics, 7*(3), 429–446.

Williams, B. (1973). A critique of utilitarianism. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism: For and against*. Cambridge: Cambridge University Press.
Winner, L. (1986). *The whale and the reactor: A search for limits in an age of high technology*. Chicago: University of Chicago Press.