Ethical Aspects of Behavior Steering Technology

Philip Brey

This is a preprint version of the following article:

Brey, P. (2006). Ethical Aspects of Behavior-Steering Technology. In P. P. C. C. Verbeek & A. Slob (Eds.), User Behavior and Technology Develepment, (pp. 357-364). Dordrecht: Springer.

For a short period of time in the mid-1970s, a federal law in the United States mandated cars to be designed not to start if seat-belts were not worn. Cars produced during that brief period of history had an electric link between the seats, seat belts and starter. In the seats, there were weight-sensing elements that registered whether a person was using the seat. If so, the seat belt for that seat would have to be fastened or the starter would not work (and a buzzer would sound). This mechanism is an example of *behavior-steering technology*, which is a technology in which one of its main functions is to make users of the technology behave in a way that is not necessarily desired by the user but that is desired by some other party in control of the technology. Usually, such behavior-steering functions are designed to be part of the technology by order of a government or other standard-setting body or by an organization that has commissioned the technology. Some other examples of behavior-steering technologies (BSTs) are gas pedals that increase their resistance the foot of the driver when the car is going above a certain speed limit so as to encourage a more ecomical use of energy, or the heavy weights on hotel keys that induces hotel guests to drop off the key at the receptionist before going outside.

The question that is to be discussed in this essay is under what circumstances, if any, the use of behavior-steering technology is justified from a moral point of view, and under what circumstances its use may become morally problematic. The U.S. seatbelt case illustrates that behavior-steering technologies are sometimes controversial. U.S. car drivers did not appreciate being mechanically forced to wear their seat belts, and many drivers had the mechanism illegally removed. Some people even mounted a court challenge: they felt that the coercive mechanism in place went against their civil liberties. As a result of these protests, the law was repealed, and wearing seat belts became again something that was mandatory but no longer mechanically forced.

So when are behavior-steering functions morally permissible? When is the moral price that has to be paid for the perceived benefits of behavior steering to be too high? In what follows, I will discuss three moral issues need to be considered in

the use of behavior-steering technologies (BSTs): the freedom issue, the technocracy versus democracy issue and the responsibility issue.

1. The freedom issue

The seat belt example suggests that there may be circumstances in which the use of BSTs infringes on freedom rights. Even if the seat belt mechanism at issue can be shown not to infringe on basic freedom rights, it is easy to think up examples of BSTs that do. One could, for example, imagine a car with an on-board navigation system that puts certain areas of town off-limits whenever the city government believes this to be a good idea. Or one could imagine automobiles in fundamentalist countries that only start after an intelligent camera on board has determined the driver to be male, or to be wearing the right state-prescribed clothing. Or one could imagine a car with a mechanical arm inside that hits the driver when he is going over the legal speed limit or swerves on the road. There are also examples of BSTs that clearly do *not* infringe on freedom rights. The hotel key with the heavy ball, for example, can hardly be argued to be a sinister coercive instrument of hotel owners that wrongly restricts the freedom of hotel guests.

So under what circumstances do BSTs restrict human behavior to the extent that basic freedom rights are violated? To answer this question, I will first briefly discuss the classical distinction between two conceptions of freedom, or liberty, that has been proposed by philosopher Isaiah Berlin in his famous essay "Two concepts of liberty". Berlin argues that, historically, the concept of liberty has has two quite different meanings. In the first, 'negative' sense, a person is free to the extent to which his actions are not obstructed or interfered with by others. Your amount of personal freedom, in this sense, is given by the area within which you can act unobstructed by others. In the second, 'positive' sense, a person is free to the extent to which he his his own master, whose life and decisions depend upon himself and not upon external forces of any kind. Such a person is autonomous, or self-determining, and is able to think freely, bear responsibility for his own choices, and able to explain them by reference to his own ideas and purposes. He is his own master, and a slave to no one.

To better understand this contrast, imagine a ruler or king who has unlimited freedom to do whatever he pleases - none of his subjects dare disobey him or get in his way. Yet, all his actions turn out to be strongly conditioned by his father, to which he is psychologically dependent to a large extent: all his choices and actions in life are in fact mere attempts to please his father rather than reasoned, autonomous choices of a free agent. Such a ruler, then, enjoys a large amount of 'negative' freedom but little or no 'positive' freedom. In contrast, a Nelson Mandela, when still locked in his cell during the apartheid regime in South Africa, had very little

'negative' freedom, but his independent, unbroken spirit conferred to him a great amount of 'positive' freedom.

'negative' freedom or freedom from interference

This distinction suggests that BSTs may restrict freedom in two ways: by restricting negative liberty and by restricting positive liberty. Let us first turn to restrictions of negative liberty. These occur when BSTs interfere with the activity of technology users. But, as the hotel key example shows, not every such interference constitutes an infringement on human rights. So when does interferences by BSTs go too far? A first possible line of response to this question would be to argue that interference goes too far when the majority of technology users believes that the BST interferes with their actions to an unreasonable extent, as happened with the seat belt mechanism. Yet, this response is not wholly satisfactory: that a majority believes that a particular technology is unacceptable does not necessarily mean that it is, in fact, unacceptable.

Usually, BSTs are used to promote some socially desirable end, such as sustainability, safety, efficiency, or equity. The price that is sometimes paid for these ends is that limits are imposed on the scope of individual behavior, thus constaining individual liberties. In trading off these positive and negative impacts of BSTs, the importance of promoting positive ends must be weighed against the cost to individual liberty. A BST may then turn out to be justified, even if there is resistance to it. This is the case when, the cost of not using the BST can be shown to be greater than the cost to individual liberty when it is used. For instance, a careful assessment may show that meeting environmental goals requires more efficient use of the automobile, and that this requirement can only be met through certain coercive measures such as speed delimiters built into automobiles because less coercive alternatives have failed. When it is determined in the democratic political arena that sustainability is a greater good than the right of drivers to drive at high speeds, then such coercive methods may turn out to be justified. At the same time, there may be some basic freedom rights that arguably should not be violated under any circumstances: rights such as included in the declaration of human rights, such as the right to freedom of movement, freedom of speech and freedom of assembly. These are ones that are essential to realizing one's individual life plans.

'positive' freedom or autonomy

Let us now turn to 'positive' freedom or autonomy. Autonomy is not threatened by a few BSTs that restrict freedom of movement. Autonomy becomes threatened when essential parts of our lives are conditioned by BSTs, to such an extent that they do not merely interfere with our actions, but go as far as to shape and condition our plans

and goals. Suppose, for example, that automobiles do not just require us to wear a seat belt, but in fact do the driving for us: they are programmed to select routes for us, select driving speeds, stop at stop signs, regulate the internal temperature, and so on. Driving in such 'intelligent' automobiles is not an autonomous activity, but is governed by the dictates of the automobile; the human driver merely follows its choices.

Suppose, now, that other machines like our personal computer, toaster, laundry machine, refrigerator and personal digital assistant are programmed similarly to make our choices for us - which they increasingly are. Living in such a technological culture, in which machines decide for us, we are no longer autonomous decision-makers, but leave it to machines to tell us what is good for us and what we should therefore do. The programmed choices of these machines may sometimes coincide with choices that we would have made ourselves, but at other times may reflect interests of a government or of a company one works for, or may simply be wrong guesses about our own preferences.

A transfer of intelligent thought and decision-making from persons to machines is not necessarily a bad thing; there are many things that we would rather have machines decide on: how hard to suck in air when vacuuming different surfaces, or when to inflate an air bag in an automobile. It is only when machines try to determine more overarching goals of activities that autonomy is significantly eroded: when an automobile tells us where we want to go or how fast we want to go there, or when word processing software changes spelling, grammar and style of our writings without giving us the opportunity to overrule its choices.

2. The technocracy versus democracy issue

BSTs have been accused of of being a technocratic method for shaping society and implementing policy. A technocracy is a society in which political power rests to a significant extent in the hands of scientists, engineers and other experts, who use scientific principles and technological means to attain political ends. A technocratic solution to a policy problem is, therefore, a solution in which technical solutions to problems have taken the place of political decisions. This is customarily held to be desirable for at least two reasons: because technocratic solutions are not democratic, and because they depend on the false idea that social problems can be solved by means of a technological 'fix'.

The first of these criticisms, that technocratic solutions are not democratic, has been held by critics of technocracy such as Jürgen Habermas, Helmut Schelsky, Jacques Ellul and Max Horkheimer. They have argued that science and technology have an internal logic or rationality of their own, that presents itself in an objective form that cannot easily be criticized in a democratic political arena. If scientists and

engineers have determined that solving world hunger requires the genetic engineering of crops and that solving traffic jams requires satellite monitoring of traffic, how can politicians dispute these ideas if they lack the proper expertise? Scientific-technological decision-making and (democratic) political decision-making hence seem to belong to separate spheres; yet, in a technocracy the first increasingly takes the place of the second, and hence politics becomes less democratic.

The image of science and technology on which this criticism rests has since been shown to be problematic, however. In contemporary science and technology studies (e.g., Mackenzie and Wajcman, 199; Bijker, Pinch and Hughes, 1987), it is held that science and technology are socially shaped: they are not the result of objective, rational principles but are contingent outcomes of confrontations between humans and nature that involve value- and interest laden choices. Therefore, there is no principled dichotomy between science and technology on the one hand, and politics on the other. This is most obviously so for technology: technological artifacts and systems can be developed to reflect the narrow interests of designers or corporations, but can also be developed in a more democratic manner, to reflect the values and interests of a wide range of people (e.g., Sclove. 1995). Therefore, technological solutions to social problems are not necessarily less democratic than social solutions such as laws and protocols.

Nevertheless, there is a danger that BSTs become undemocratic technologies, because their basic functionality may still be largely decided by engineers rather than by democratic representatives. A democratic design of a BST requires that a procedure is followed that is similar to the development of a new law: a careful consideration of alternatives, and a democratic procedure in which parties with different interests can let their voices be heard. The functional properties of the BST, if not the precise underlying mechanism, must be sufficiently clear to non-experts; this requires an effort of both the experts and laypersons to find a common vocabulary to discuss the political consequences of the technical choices that can be made.

The idea that social problems can be solved by means of a technological 'fix' is perhaps more problematic in this context. The seat belt case shows that engineers and policy makes are often overoptimistic about the usefulness of technologies in solving social problems. As pointed out by Norman (1988), the American seat belt mechanism that required car drivers and passengers to wear seat belt had many problems. For example, the mechanism could not distinguish legitimate case in which the seatbelt should not be buckled from illegitimate ones. If you wanted to carry a heavy package, for example, you were forced to buckle it. And the mechanisms were not reliable, so they often failed, wrongly buzzing, and stopping the engine. Moreover, the mechanism was easy to circumvent by simply buckling the belts and suffing them under the seat. This example shows that BSTs require careful assessment to make sure they are reliable, work right, and distinguish

legitimate violations from illegitimateones. This is very hard to attain for any technological solution to a social problem.

3. The responsibility issue

The final moral issue relating to BSTs is that they may create a *responsibility vacuum*: a situation in which it is fundamentally uncrear who is responsible for actions and their consequences. Many BSTs take away some amount of responsibility from their users. For example, there are automobile that take over the brakes when a car goes into a skid. But this means that in a resulting fatal accident, not just the behavior of the driver can be blamed, but also the behavior of the car itself, and the company who programmed the brakes to respond in a certain manner. In society, individuals are expected to take moral responsibility for their actions, which means that they can be held accountable for them, and can receive praise or blame for them. The use of BSTs complicates the assignment of responsibility, because users may claim responsibility did not rest with them.

4. Conclusion: dealing with the moral issues in behavior-steering technology

When developing a BST, three moral issues must be considered: (1) what are its negative consequences for the individual freedoms of users and can such consequences be morally defended? (2) is there sufficient democratic input during the process of design and implementation of the technology? and (3) what implications does the BST have for the distribution of responsibility between the user on the one hand and the technology and its developers on the other?

Regarding the first issue, I have suggested that BSTs should refrain from violating basic freedom rights but may constrain human action if there is a greater good that is served by this, and I have suggested that BSTs should not erode autonomy by taking away decision-making power over the basic ways in which individuals live their lives. Regarding the second issue, I have argued that BSTs can be designed in more democratic and more technocratic ways, and that for BSTs that are intended to provide policy solutions, an effort should be made for the development process to be as transparent and democratic as possible; also, I have argued that it is desirable to anticipate on side-effects of the use of BSTs in real settings, so as to avoid BSTs from being poorly-thought-out 'technological fixes'. Regarding the final issue, the responsibility issue, it is desirable that the question of how responsibility is distributed in the use of a BST is answered before a BST is put into use: involved actors (notably, the manufacturer and the user) should be clear on who is responsible when the use of the BST results in harmful consequences.

REFERENCES

Berlin, I. (1979). "Two Concepts of Liberty," *Four Essays on Liberty*. Oxford: Oxford University Press, p. 118-172.

Bijker, W., Pinch, T., and Hughes, T., (eds.) (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press,.

MacKenzie, D., and Wajcman, J., (eds.) (1999). *The Social Shaping of Technology, 2nd ed*. Open University Press.

Norman, D. (1988). *The Design of Everyday Things*. New York: Currency.

Sclove, R. (1995). *Democracy and Technology*. New York: The Guilford Press.