Balancing the Benefits of Autonomous Vehicles

Timothy Geary
School of Humanities & Communication
California State University, Monterey Bay
Monterey, CA USA
tgeary@csumb.edu

David Danks
Departments of Philosophy and Psychology
Carnegie Mellon University
Pittsburgh, PA USA
ddanks@cmu.edu

Abstract
Autonomous vehicles are regularly touted as holding the potential to provide significant benefits for diverse populations. There are significant technological barriers to be overcome, but once those are solved, then autonomous vehicles are expected to: reduce fatalities; decrease emissions and pollutants; provide new options to mobility-challenged individuals; reduce travel and commuting times; enable people to use their time more productively; and so much more. In this paper, we argue that these high expectations for autonomous vehicles almost certainly cannot be fully realized. More specifically, the proposed benefits divide into two high-level groups, centered around efficiency and safety improvements, and increases in people’s agency and autonomy. The first group of benefits is almost always framed in terms of rates: fatality rates, traffic flow per mile, and so forth. However, we argue that the absolute numbers for these measures, not the rates; number of fatalities is the key metric, not fatality rate per vehicle mile traveled. Hence, these potential benefits will be reduced, perhaps to non-existence, if autonomous vehicles lead to increases in vehicular usage. But that is exactly the result that we should expect if the second group of benefits is realized: if people’s agency and autonomy is increased, then they will use vehicles more. There is an inevitable tension between the benefits that are proposed for autonomous vehicles, such that we cannot fully have all of them at once. We close by pointing towards other types of AI technologies where we should expect to find similar types of necessary and inevitable tradeoffs between classes of benefits.

Introduction
Autonomous vehicles (AVs), along with many other AI technologies, are forecast to have numerous societal benefits, both quantitative and qualitative. For example, many experts predict that AVs will reduce fatalities (Litman, 2017), emissions (Fagnant and Kockelman, 2014; Greenblatt and Saxena, 2015), traffic (Talebpour and Mahmassani, 2016) and parking congestion (Zhang, Guha-thakurta, Fang, and Zhang, 2015). In the near future, according to the predictions, we will have safer, cleaner, and more livable cities. At the same time, AVs are predicted to increase people’s autonomy and self-determination (Bradshaw-Martin and Easton, 2014; Lenz and Fraedrich, 2016), thereby enabling populations who are currently underserved by transportation options (e.g., those who are visually or mobility impaired) to live fuller lives.

Discussions about the benefits of autonomous vehicles typically treat these various beneficial possibilities as independent: a gain of one type of benefit from AVs is usually assumed (perhaps implicitly) to have no appreciable impact on the likelihood that we will gain some other benefit from AVs. Of course, there are exceptions to this generalization—for instance, analyses of emission benefits frequently also model the traffic impacts (Fagnant and Kockelman 2014)—but there is relatively little explicit discussion of the possibility of trade-offs. This assumption of independence is not necessarily a problem; in particular, it is arguably warranted whenever benefits are expressed as changes in the relevant rates. It is hard to see, for example, why changes in the fatality rate would depend in any interesting way on the increase in transportation options for mobility-impaired individuals.

Nevertheless, we contend in this paper that the presumed independence has masked a deep tension between the various potential benefits of AVs. In particular, we argue that we cannot simultaneously reap all of the benefits of AVs, or at least, all of the benefits that actually matter to us. In the section “Classes of Benefits,” we examine the different benefits proposed for AVs, and show that they largely fall into two groups—either safety and efficiency, or autonomy and self-determination. The former benefits are almost always expressed as improved rates, while the latter emphasize increased transportation opportunities.

As we argue in the “Fundamental Tension” section, though, we should ultimately not care about improved rates of harm, but only improved magnitudes of harm. A reduced rate of vehicular fatalities is cold comfort if there is an increased number of deaths due to an accompanying increase in vehicle-miles traveled (VMT). And if the autonomy and self-determination benefits are realized, then we should expect exactly such an increase in VMT. That is, if we actually
achieve one set of benefits from AVs, then another set will be much smaller, and can even become liabilities. We thus finish in “Responses to the Tension” with a consideration of ways to regulate and develop AV technologies to appropriately balance these two considerations.

**Classes of Benefits**

Technologies such as AVs will undoubtedly impact our lives in numerous ways, as transportation plays a key role in many economic and private domains. Unsurprisingly, there have been many analyses (too many for us to exhaustively cite!) about the potential benefits and drawbacks of AVs. At a high level, we can usefully distinguish between two different types of benefits.

**Safety and Efficiency**

The first, and most discussed, benefit of AVs is the potential to dramatically reduce crash and fatality rates. In 2016, Americans drove over 3.2 trillion miles (OHIP, 2016). In the same year, combined fatality rates for both rural and urban driving was estimated at 1.18 fatalities per 100 million miles driven; overall, human-caused motor vehicle crashes resulted in 37,461 fatalities, over 28% of which were due to alcohol use (NHTSA, 2017).

In light of this enormous human cost, the potential safety value of AVs is enormous. Autonomous vehicles would not speed (under most conditions), drink, or get distracted. In fact, analyses have argued that AVs could save nearly 30,000 lives per year in the United States “by shifting the focus from minimization of post-crash injury to collision prevention” (Fleetwood, 2017). Widespread use could be transformative in terms of lives saved, particularly given the principal importance of safety considerations in AV design and deployment (Litman, 2017). Continual improvement of AV technologies should only further drive down the relevant fatality rates, as developers gain more data and experience with the diversity of driving contexts (Kalra and Groves, 2017).

In addition to increased safety, AVs can lower fuel consumption and emissions, thereby protecting the environment and human well-being (Fagnant and Kockelman, 2014; Greenblatt and Saxena, 2015). AVs can drive more smoothly, accelerate/decelerate more efficiently, and reduce fuel consumption by driving the speed limit. With appropriate programming, AVs can take advantage of eco-driving principles to increase efficiency by arguably up to 25%, with corresponding reductions in fuel consumption by 20% and overall delays in traffic by 60% (Igliński and Babiak, 2017). Total greenhouse gasses could, if AVs were electrified rather than fossil fuel-based, be reduced by over 80% (Fulton, Mason, and Meroux, 2017). Of course, some of these benefits interact with one another: for example, AVs are plausibly more likely to be electrified, and so their wide usage provides environmental benefits along two distinct pathways.

Third, AVs are predicted to reduce traffic and therefore travel times. Most traffic is caused by inefficient braking and acceleration, and AVs should reduce those events. In 2014, congestion increased urban Americans’ travel time by 6.9 billion hours, resulting in use of an additional 3.1 billion gallons of gasoline, with a corresponding cost of $160 billion (EIA, 2018). Assuming that AVs are fully autonomous and connected, they can use road space more efficiently and improve traffic flow, thus saving time and fuel (Ji, 2018). More generally, AVs will likely be more consistent and conservative in their driving, which can potentially yield significant reductions in traffic (Talebpour and Mahmassani, 2016). Additionally, if safety benefits are realized, then congestion caused by crashes would correspondingly also be reduced (EIA, 2018).

**Autonomy and Self-determination**

AVs are not simply safer or more efficient drivers. Since no human driver is required, they can provide mobility solutions for people who are otherwise dependent on the support or services of others. For example, elderly, blind, or other mobility-impaired individuals would have the ability to use an AV without coordinating with other people, in contrast to their current dependency (Bradshaw-Martin and Easton, 2014). Alternately, people who might otherwise drive while alcohol impaired would have new transportation options, thereby increasing their ability to function autonomously without placing others at risk. More generally, widespread deployment would likely lead to the development of entirely new patterns of mobility usage (Lenz and Fradrich, 2016).

Even in traditional mobility roles, the introduction of AVs can lead to significant gains in autonomy and self-determination. Reductions in commute times, as well as the ability to productively use that travel time, can enable people to spend more time on tasks that are important or valuable to them. High-level AVs will take over most or all driving tasks, thus freeing up time for other tasks that better advances people’s interests. The net result is an increase in usable (for the person) time. Relatedly, AVs plausibly will reduce people’s overall levels of stress due to reduction of time actually driving, thereby enabling them to be more productive when not in the vehicle. With full autonomous capabilities, users would be able to work, relax, or even sleep in the AV.

**The Fundamental Tension**

Proponents of AVs tout these many benefits as reasons to pursue rapid adoption, even when there are significant uncertainties about their performance in novel contexts or
situations (Kalra and Groves, 2017). We agree that 100\% safety or 100\% reliability are unrealistic goals, and that we should focus on determination of conditions under which AV adoption can provide significant incremental benefits. However, we suggest that efforts to evaluate those conditions have largely asked the wrong questions (or the right questions with the wrong presuppositions).

When we consider the possible impacts of AVs, we contend that we should ultimately worry about the total harms and benefits that result from their use in particular contexts, under specific regulations, and so on. For example, we care about the total number of fatalities, not about the fatality rate (except inasmuch as it carries information about the total number of deaths). We care about the total emissions, not the emissions per vehicle. And so on.

For the types of factors we have considered here, the total magnitude of some harm and benefit will be (roughly) the relevant rate, multiplied by the total usage.\(^1\) Hence, the value of changes in rates cannot be judged without knowledge (or assumption) about whether the total usage changes as well. If the total usage (i.e., VMT in the case of AVs) is unchanged, then a reduction in the rate of some harm will be a positive, as it will reduce the total amount of harm. If the total usage instead increases as a result of the technological change, then this reduction in the rate of harm could actually translate into more total harms, if the usage increase is larger (as a percentage) than the rate reduction.

Consider a simplistic example with numbers similar to the vehicular fatality data (provided in the previous section) for the United States in 2016: a (human-driven) vehicular fatality rate of 1 death per 100 million miles; total VMT of 3 trillion miles per year; resulting in 30,000 vehicular fatalities per year. Suppose that widespread AV deployment would result in a 10\% improvement in the vehicular fatality rate; that is, the new fatality rate would be 0.9 deaths per 100 million miles. This improvement appears to be a significant gain, as we would expect to save 3,000 additional lives each year. However, suppose that this same widespread deployment also leads to increased vehicle usage, whether from people who have limited current use of vehicles (e.g., visually-impaired or elderly individuals) or from shifts to AVs from other forms of transportation (e.g., working in an AV rather than on a train). For concreteness, suppose the VMT increased by 20\% to 3.6 trillion miles per year. In that case, though, we would expect to have 32,400 total deaths. That is, this widespread deployment of AVs—each of which is safer than a human-driven vehicle—would not save lives, but would actually result in 2,400 more deaths each year. And in such a case, we contend that we ought to reject the use of AVs, or at least not adopt them because of some supposed safety benefit. AVs might reduce the fatality rate, but the relevant safety benefit is the fatality magnitude, which depends on the impact of AVs on VMT.

More generally, as noted above, essentially all of the analyses of safety and efficiency benefits either focus exclusively on rates, or assume (perhaps implicitly) that AV usage and VMT remain approximately constant, or even go down. Even analyses that entertain the possibility of an increase in VMT (e.g., Kalra and Groves, 2017) nonetheless place most of their emphasis on situations in which total VMT is flat or reduced. However, if the proposed autonomy and self-determination benefits are actually realized, then we should expect to see an increase, perhaps a quite significant one, in VMT. The net result would be the minimization—perhaps even elimination—of the safety and efficiency benefits (see also Fagnant and Kockelman, 2015). Contrary to the standard analyses, the potential benefits of AVs are not independent, but rather are in tension: increased autonomy and self-determination benefits will (barring appropriate regulation or responses; see next section) likely come at the cost of some safety and efficiency benefits.

This tension within the benefits might seem surprising, but it has natural analogs in other contexts. For example, freeway expansion is often advocated on the grounds that it will (a) reduce congestion, thereby reducing people’s travel times; and (b) increase capacity on key transportation arteries, thereby making the roadways more accessible to more people. However, these two benefits are in tension with one another. If more people actually use those particular roadways (due to increased capacity), then the traffic congestion will not be alleviated and travel times will not decrease. In fact, the “Fundamental Law of Road Congestion” holds that this is exactly what happens: when roadway capacity is increased, the traffic volume (i.e., the VMT) increases to match it, resulting in no improvements in travel times (Duranton and Turner, 2011). Importantly, this tension is not because the benefits are imaginary; road expansion does, in fact, reduce the average travel time if the number of other cars is held fixed. The problem is that the antecedent of the conditional fails to hold (i.e., the number of cars is not fixed), and so the actual travel time is unchanged after the traffic system equilibrates. Similarly, AVs might, in fact, reduce the fatality rate (or other safety and efficiency rate), but not thereby result in any real benefits if they also cause a non-trivial increase in VMT.

We emphasize that we are not arguing that these classes of benefits are necessarily or conceptually incompatible. That is, we are not claiming that tradeoffs (see next section) necessarily must occur. Rather, we are arguing that there is a prima facie tension between the two types of benefits, and so we must engage in difficult empirical investigations to try to determine whether there is some underlying driving conditions, and so we might weight the product by the VMT in each of the different conditions.

---

\(^1\) Of course, our computation could be more sophisticated in various ways. For example, there might be different vehicular fatality rates for different
incompatibility, and if so, the relevant tradeoffs between the classes of benefits. Our paper is a call for more appropriate analyses and debates, not an impossibility proof.

We contend that we simply do not know (at the current time) whether the autonomy and self-determination benefits will result in a sufficiently large increase in VMT to offset significant safety benefits, nor whether any extra VMT will be higher or lower risk than normal, nor whether specific groups or communities are likely to be disproportionately affected, nor answers to a range of other relevant questions about this tension. Arguments that depend on a list of benefits without consideration of tensions are overly simplistic. But similarly, arguments that we must choose between safety and autonomy depend on empirical questions that (to our knowledge) remain open.

Responses to the Tension

This tension between the classes of proposed benefits has largely gone unrecognized in the literature on the pros and cons of AVs. A few analyses have acknowledged the possibility that safety and efficiency benefits can be offset if VMT increases (Kalra and Groves, 2017; Litman, 2017), but discussion of that possibility has largely been relegated to footnotes. Thus, one important response is to adapt and improve the debates about AVs. We cannot simply list and compare the different potential benefits and potential drawbacks, as it may not be possible to achieve all of the benefits simultaneously. The total benefits may be less than one might have hoped by considering each class of benefits in isolation, and so we must be more nuanced in our arguments for and against AVs.

That being said, we need not stop with only shifts in rhetoric and argumentation. One challenge that we face is that many of the relevant empirical facts may not be known (or even knowable) in advance of regulatory and legal approval of AVs. As argued above, the scope and size of the tension between classes of benefits is ultimately an empirical matter. However, predicting the usage levels of novel technologies is a notoriously challenging task: there are various frameworks for predicting adoption of novel technology (e.g., the Technology Acceptance Model; Davis, 1989), but their accuracy is mixed, particularly for technology such as AVs that potentially have wide-ranging impacts.

Moreover, this uncertainty about the VMT impact of autonomy and self-determination benefits means that we are unlikely to be able to develop an empirically accurate mathematical model of the relevant tradeoffs. And if we do not have the knowledge to construct a reliable cost-benefit analysis, then we arguably do not have the appropriate information or bases to make the standard, simple approve/deny regulatory decisions about AVs. Rather, we propose that we should consider three different, not mutually exclusive, changes to the likely regulatory processes and standards for AVs.

First, even if regulatory agencies prioritize the minimization of safety- and efficiency-related harms, then they should set a higher bar for demonstrated improvements in safety and efficiency. Various proposals have been made for what degree of improvement in (expected) safety rates should be required before AVs are approved for widespread public use (e.g., Fagnant and Kockelman, 2015), including some arguments that manufacturers should only have to demonstrate that AV accident rates are approximately equivalent to those of human drivers (Kalra and Groves, 2017). However, the autonomy and self-determination benefits imply that current VMT is only a lower bound on vehicle usage.

We have good reasons to think that the actual VMT will be significantly higher. Hence, equal safety and efficiency rates will likely lead to greater total harms, unless there turn out to be no autonomy benefits whatsoever. That is, rate equivalence to human performance is highly unlikely to be the rate at which total harms are the same with or without AVs. Instead, if we want to reduce total harms, then we should require AVs to exhibit significantly better-than-human safety and efficiency rates, precisely since we expect usage to increase. Of course, just how much better depends on empirical questions that we cannot yet answer, but we can certainly conclude that “comparable to human performance” is not the correct standard for regulatory approval (again, if regulators have the goal of minimizing total harms).

A second option is that regulators might approve the use of AVs based on one of the currently-proposed bars for safety and efficiency benefits (e.g., slightly better than human performance), but then actively monitor and manage AV usage to ensure that VMT does not grow faster than safety and efficiency rates are reduced. For example, regulators might tie the number of approved AVs, or the permissible total AV mileage, to improvements in safety rates (e.g., each one percentage point improvement in safety allows the licensing of another 10,000 AVs).

This type of active monitoring and dynamic approval would be unusual in the transportation space, but is more common in other domains such as pharmaceuticals (London and Danks, 2018) or water usage rights in parts of the United States. Implementation of such a system would require the creation of a significantly more intrusive monitoring system by the relevant regulatory agencies, but would potentially allow for the dynamic balancing of the potential tradeoffs.

\(^2\) Of course, there might be multiple reasons that we ought not regulate AVs in a simple binary manner (Danks and London, 2017; London and Danks, 2018), not solely the difficulty of estimating possible changes in VMT levels and distribution.
inherent in the tension between the two classes of proposed benefits.

A third possibility is that regulators might choose to prioritize the autonomy and self-determination benefits and approve AVs relatively quickly, perhaps even before they are safer or more efficient that human drivers. This response could enable appropriate populations to realize the increased mobility and autonomy benefits of AVs immediately, even if those benefits potentially come with some increased harms due to AV accidents. That is, regulators might decide that the autonomy and mobility expansion benefits are sufficiently strong that we should actually tolerate an increase in the number of vehicular fatalities, at least in the short run.

This potential line of response might seem surprising, since regulatory actions typically emphasize safety and efficiency benefits. However, there are many cases in which laws and regulations accept a certain level of risk or harm in order to achieve other substantive benefits, including autonomy and self-determination. One can rarely eliminate the possibility of harms, and so the relevant question is what level of risk is acceptable given the other hoped-for benefits. We propose that regulators could decide that the autonomy benefits are worth slightly increasing current risk levels, at least for certain populations. Of course, this potential risk increase could be minimized by only approving AVs for use by groups who would disproportionately benefit from their use. For example, a regulatory agency might allow only vision- or mobility-impaired individuals to use AVs, not those without relevant impairments. Such a policy would aim to realize the bulk of the autonomy benefits while minimizing the potential increase in harms.\(^3\)

These three different regulatory responses are not mutually exclusive, but could be used in conjunction with one another. Of course, there could be significant legal or political barriers to the use of one or another response in a particular country, community, or context. For example, there might be substantial resistance to approving AVs for use by only a subset of the population in a region or country. In practice, regulators might be highly constrained by relevant legal, social, and cultural norms. Nonetheless, we believe that it is valuable to see that there are different options available to regulators, even when key empirical questions cannot be answered in advance of (limited) approval and deployment.

One key feature shared by all of these responses is the need for regulators to either (a) prioritize one class of benefits, or (b) attempt to dynamically balance them. We might wish that regulators would not be placed in this position, but unfortunately these different classes of benefits are in tension with one another. Decisions about how to weight values should presumably be made only after substantial public discussion and comment. Hence, we have yet another reason to shift the rhetoric and arguments about AVs towards acknowledgment of this potential tension: open discussion is required for regulators to obtain the necessary information and feedback for them to make principled decisions about the best ways to try to regulate this new technology, with its complicated set of potential benefits.

**Conclusions**

Most decisions about widespread approval or licensure are based on cost-benefit analyses of some sort: if the new technology has more benefits than costs—whether those are economic, social, psychological, or other—then regulators typically approve the system, perhaps with restrictions on the contexts or populations of use. We have here argued that this cost-benefit analysis is significantly more complicated for AVs than has traditionally been assumed. The benefits of increasing autonomy and self-determination themselves carry a cost in the form of increased VMT, and so increased opportunities for accidents, traffic, and emissions. We must use a more sophisticated understanding of the underlying psychology of technology use and adoption in order to better estimate the likely changes in key parameters of our analyses.

When framed in this way, one can quickly see that this tension is not unique to AVs, but rather arises for a host of other kinds of AI systems. For example, medical diagnostic AIs are often praised for their potential to have greater accuracy, and also to provide more widespread access to accurate diagnoses (since they can be deployed in regions where there are few human doctors). But the latter benefit of access implies much greater usage of the system, and so a corresponding increase in the number of false positives, which themselves impose a cost on the healthcare system.\(^4\) Of course, regulators might decide that the increase in access is worth some additional costs and burdens on the system, but that is a trade-off of values that must be considered. We ought not model the accuracy improvement benefits in isolation from the access improvement costs.

Our observations and arguments here also point towards the need for improved models of technology adoption and usage. While such predictions are quite difficult, there could be significant gains from even incremental improvements in our predictive accuracy. The costs and benefits of novel technologies frequently interact with one another, and our

---

\(^3\) Of course, if AVs are particularly risky for passengers, then the regulators would be putting disproportionate risk on already-disadvantaged populations, which might provide a compelling contrary reason to deny early approval for these groups.

\(^4\) This possibility is particularly worrisome if the increased usage occurs in low-risk populations, since there might be few people who actually benefit from learning a true positive diagnosis.
approval decisions should be based on more complex analyses that capture these interdependencies.

Acknowledgments

This paper was made possible in part by support from the McNair scholars program and CSU-Louis Stokes Alliance for Minority Participation (CSU-LSAMP) Program to TG, and an Andrew Carnegie Fellowship (Carnegie Corporation of New York) to DD. The statements made and views expressed are solely the responsibility of the authors.

References


