
Feedback: Seeing Through Safety

Tuesday, July 11, 2017

Author(s)

Kyle Wesson

*This is the third and final post in a mini-series on the Sci on the Fly blog that explores questions about feedback. Earlier posts are [here \(https://www.aaaspolicyfellowships.org/blog/feedback-turning-directions\)](https://www.aaaspolicyfellowships.org/blog/feedback-turning-directions) and [here \(https://www.aaaspolicyfellowships.org/blog/feedback-checking-out-reviews\)](https://www.aaaspolicyfellowships.org/blog/feedback-checking-out-reviews). This post asks: **How does feedback influence the safety of autonomous vehicles?***

The flight recorder, or “black box,” on board an aircraft is an engineering marvel. It can withstand temperatures of 1,830 degrees Fahrenheit and accelerations of 3,400 g-forces while recording the operation of an aircraft’s critical systems, the crew’s actions, and the aircraft’s location and altitude. The black box is the most sought after piece of equipment after any accident, because it plays such a critical role in aviation’s safety feedback loop. Its data allows investigators to determine what went wrong and how to prevent it in the future. But what happens when such feedback is hidden behind proprietary computer code or obscure algorithms? Does safety suffer?

Critical aviation systems, such as autopilot, whose failure could result in a catastrophic loss must be certified to “Design Assurance Level A” according to an aviation standard called [DO-178B \(https://en.wikipedia.org/wiki/DO-178B\)](https://en.wikipedia.org/wiki/DO-178B). Such systems can take years to code, evaluate, and field after going through rounds of rigorous testing. Even with all of the coding rigor and algorithm development, the black box offers yet another layer of feedback and redundancy. After an accident, an aviation black box helps determine if a critical system encountered a fault or operated according to specification.

Aviation's incredible safety culture — consider that there have been no deaths due to a U.S. airline crash anywhere in the world since 2009 — invites a comparison to the safety culture surrounding autonomous vehicles. As automakers race to achieve [SAE Level 5 \(https://www.sae.org/news/3544/\)](https://www.sae.org/news/3544/), the highest measure of automation, the new tools they have turned increasingly towards are artificial intelligence, machine learning, and proprietary algorithms to “drive” the vehicle. They do so with good reason: SAE Level 5 demands that the automated system can perform all of the same driving tasks under all of the same conditions that a human driver could. It's a difficult and still elusive feat. Yet instead of carrying an aviation-style black box, the autonomous vehicles themselves can become a much different kind of “black box”—the kind whose operation is opaque and potentially impossible to understand.

Despite recommendations in the Department of Transportation [Federal Automated Vehicles Policy \(https://www.transportation.gov/AV\)](https://www.transportation.gov/AV) that “manufacturers and other entities should have a documented process for testing, validation, and collection of ... crash data ... to establish the cause of any such issues,” autonomous vehicles do not carry aviation-style black boxes. Compounding matters, the data and algorithms driving autonomous vehicles are complex and unproven. For example, NVidia recently unveiled an autonomous vehicle whose [machine learning techniques taught itself to drive \(https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/?set=604130\)](https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/?set=604130) by watching human drivers do the same. No one, from the engineers and programmers to the passengers or pedestrians, could be quite sure why the software “drove” the way it did or how it might react in an unpredictable situation. It had, after all, taught itself to drive, not how to explain itself.

These machine learning techniques can be wild and unpredictable, and they certainly do not meet DO-178B specifications. In one example of machine learning gone awry, Google's artificial neural network, which is a type of artificial intelligence, concluded that [dumbbells were always attached to arms \(http://www.businessinsider.com/googles-ai-can-teach-us-about-the-human-brain-2015-7\)](http://www.businessinsider.com/googles-ai-can-teach-us-about-the-human-brain-2015-7) but in grotesque configurations based on pictures that always showed the two connected. Such over-generalization of the real world creates challenges for future safety risk assessment or crash prevention protocols. Moreover, the proprietary nature of the data and algorithms behind autonomous vehicles could limit the role of regulators to establish safety standards based on crash feedback. Indeed, the only way to prove the safety of autonomous vehicles may be to let them [drive on our roads for millions of miles \(https://www.rand.org/pubs/research_reports/RR1478.html\)](https://www.rand.org/pubs/research_reports/RR1478.html). And for that, we would have to let them share the road with us as unproven black boxes.

Yet there is also an analog between autonomous vehicles and ourselves. As opaque as an autonomous vehicle may be, human drivers cannot always explain their actions, may drive while intoxicated, and frequently make poor driving decisions. We can be our own black box at times, and we implicitly accept this risk whenever we take to the road. Perhaps autonomous vehicles aren't so different from ourselves. The question we must answer, then, is are we willing to accept that explanation when our lives are on the line?

Image: Pixabay.com | @HypnoArt

Source URL: <https://www.aaaspolicyfellowships.org/blog/feedback-seeing-through-safety>

List of links present in page

- <https://www.aaaspolicyfellowships.org/blog/feedback-turning-directions>
- <https://www.aaaspolicyfellowships.org/blog/feedback-checking-out-reviews>
- <https://en.wikipedia.org/wiki/DO-178B>
- <https://www.sae.org/news/3544/>
- <https://www.transportation.gov/AV>
- <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/?set=604130>
- <http://www.businessinsider.com/googles-ai-can-teach-us-about-the-human-brain-2015-7>
- https://www.rand.org/pubs/research_reports/RR1478.html