How common sense fails us on the road: contribution of bounded rationality to the annual worldwide toll of one million traffic fatalities

Michael Sivak *

The University of Michigan Transportation Research Institute, 2901 Baxter Road, Ann Arbor, MI 48109-2150, USA

Abstract

Bounded rationality is a generally adaptive behavior characterized by a rapid decision making process that involves only the most salient aspects of the problem. In the driving context, bounded rationality is evident on a macro (societal) level as well as on a micro (individual driver) level. On a macro level, bounded rationality underligned the development of many of the early, common-sense countermeasures for traffic safety problems. On a micro level, bounded rationality can lead to unexpected driver behaviors and thus to unexpected effects of some common-sense countermeasures. Because the obvious, common-sense countermeasures have already been largely identified (although not always implemented), further progress will require insights rooted in comprehensive theories of driving, and formal study of the implications of these insights. It will require unbounded rationality; it will require real science.

© 2003 Elsevier Science Ltd. All rights reserved.

1. Background

1.1. Basic facts

We were not born to drive. We were born to walk and run. Nevertheless, driving is an exceptionally safe acquired skill. In the US, one would have to drive, on average, 100,000,000 km before being involved in a fatal accident (National Safety Council, 2001). That represents about 2500 trips around the world, or about 130 trips to the moon and back. Nevertheless, because of

---

*Presented April 16, 2002 at the University of Michigan on the occasion of being named a Distinguished Research Scientist.
*Tel.: +1-734-936-1089; fax: +1-734-764-1221.
E-mail address: sivak@umich.edu (M. Sivak).

1369-8478/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S1369-8478(03)00003-2
the enormous amount of driving, it is estimated that worldwide 1,170,000 persons are killed annually in road accidents (World Health Organization, 1999).

Road accidents are the leading cause of death for young persons between 15 and 29 years of age (Murray & Lopez, 1996). Because road accidents involve disproportionate numbers of young people, they feature prominently in the so-called disability-adjusted life years, or DALYs. DALY is a sum of two components, each taking into account expected years of life. One component, years of life lost, relates to fatalities, while the other, years lived with disability, relates to living with diseases or injuries. Road accidents are the fourth leading cause of DALYs, accounting for 4.4% of all DALYs (Murray & Lopez, 1996). (The leading cause of death— ischemic heart disease—accounts for 9.9% of all DALYs (Murray & Lopez, 1996).)

1.2. What is meant by “safety” in the driving context?

There are several different measures of road safety, but all are rates using various denominators. These measures include fatality rates per unit of distance, vehicles, drivers, and population. The selection of the measure influences the answer that we get concerning the current state of affairs as well as progress over time (Sivak, 1996).

The data most frequently touted are those for fatalities per unit of distance traveled. In the US, the fatality rates per unit of distance traveled show a remarkable decrease of 93% between 1923 (the first year for which mileage estimates are available) and 2000 (National Safety Council, 2001). However, the data per population tell a different story. This rate has remained relatively stable over time, with the rate for 2000 only 5% lower than the one for 1923. It is as if we had a certain tolerance for traffic carnage, not in absolute terms but relative to population.

1.3. How does driving compare to other modes of transportation in terms of safety?

Using the fatality rate per distance traveled, in the US, driving is about 15 times as risky as taking a train, and 30 times as risky as taking a bus (National Safety Council, 2001). Although the fatality rate per distance traveled is a reasonable index for ground transportation, the risk of flying is primarily affected by the number of takeoffs and landings. For example, a recent study by Boeing (2001) indicates that 95% of flying fatalities occur either during takeoff and climb after takeoff, or during descent to landing and landing. Only 5% of the fatalities result from accidents that occurred at cruising altitudes. Consequently, the risk of flying depends primarily on the number of flight segments involved in the trip, not on the distance traveled. As a first approximation, a nonstop trip is half as risky as a trip that involves an intermediate stop. A longer nonstop trip is usually substantially safer than a shorter trip that involves an intermediate stop.

Recently, we compared the risks of flying and driving in the US by taking into account that the risk of flying is dependent on the number of flight segments, but, for all practical purposes, is independent of the length of the trip (Sivak & Flannagan, 2003). It turns out that in the US, flying domestically nonstop is safer than driving if the distance is more than 18 km. For one-stop and two-stop flights, the corresponding distances are 36 and 54 km, respectively. Thus, for any reasonable combination of the trip distance and number of flight segments, flying is safer than driving.
In Sivak and Flannagan (2003), we averaged the flight information for 10 years (1992–2001) to get the flying functions, because of large year-to-year variations in flight fatalities. The calculations included the four hijacked planes on September 11, 2001. In the same study, we calculated that disastrous airline incidents on the scale of those of that day would have had to occur about once a month for flying domestically in the US to become equally risky as driving.

2. Bounded rationality

In simple terms, classical or "unbounded" rationality refers to optimization of some function, such as subjective utility, using all relevant information, and unlimited processing resources and processing time (Gigerenzer & Selten, 2001). An important aspect is that all alternatives are considered before a decision is made. Thus, unbounded rationality refers to normative behavior, or how people should act. In reality, however, people do not have all of the relevant information. And they certainly do not have unlimited processing resources and time to reach the optimal decision.

Enter the concept of "bounded rationality". It refers to nonoptimizing, adaptive behavior of real people, and thus it is a descriptive concept. Bounded rationality uses salient information only, relies on limited information processing resources, and can produce a decision without considering all alternatives. The origin of bounded rationality can be traced to the work of Herbert Simon in the 1950s. According to Simon (1957), decisions are reached through a process called satisficing. Decision alternatives are considered through a sequential search. This search is terminated when an alternative is found that meets or exceeds a certain criterion level called the aspiration level. The aspiration level depends on conditions. For example, the aspiration level is lowered if satisfactory alternatives are difficult to reach. As Simon pointed out, "such changes in aspiration level... tend to guarantee the existence of satisfactory solutions" (Simon, 1957, p. 253). The main features of unbounded rationality and bounded rationality are compared in Table 1.

2.1. Bounded rationality and merging into traffic

I will now apply unbounded and bounded rationality to two driving situations. In the first situation, you are at a stop sign, trying to enter traffic on a major road. According to unbounded rationality, you evaluate each opportunity to merge, based on the gap size, the speed of the traffic, the type of vehicle at the tail end of the gap, visibility, surface friction, the acceleration capabilities

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Main differences between unbounded rationality and bounded rationality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unbounded rationality</strong></td>
<td><strong>Bounded rationality</strong></td>
</tr>
<tr>
<td>Optimization</td>
<td>Nonoptimization (satisficing)</td>
</tr>
<tr>
<td>Using all information</td>
<td>Using salient information</td>
</tr>
<tr>
<td>Unlimited information processing resources</td>
<td>Limited information processing resources</td>
</tr>
<tr>
<td>All alternatives are considered</td>
<td>Not all alternatives need to be considered</td>
</tr>
<tr>
<td>Fixed criterion level</td>
<td>Flexible criterion (aspiration) level</td>
</tr>
<tr>
<td>Normative (ideal people)</td>
<td>Descriptive (actual people)</td>
</tr>
</tbody>
</table>
of your vehicle, and so on. Based on all of this information, you classify each gap as safe or unsafe. If a gap is computed to be safe, you use it to merge. Importantly, the decision is not influenced by the length of the time you are waiting to merge.

According to bounded rationality, you set yourself a criterion or aspiration level in terms of the available time gap in traffic. This aspiration level might or might not be entirely safe. For example, based on your prior experience, you know that, if necessary, other drivers would most likely adjust their behavior rather than crash into you. Furthermore, the critical gap, or aspiration level, varies with how long you are waiting for a satisfactory gap. If you are waiting a long time, you are likely to reduce the aspiration level. You will likely accept a shorter gap. Once a gap meets or exceeds your aspiration level (whether or not it is unconditionally safe), you merge into the traffic. Furthermore, in bounded rationality, the value of the aspiration level is not as precisely specified as is the “true safe criterion” in unbounded rationality. This is the case because to get you into the ballpark, you do not need to use all of the relevant information and the calculations need not be precise. Thus, bounded rationality requires much less information and information processing than does unbounded rationality. Thereby, bounded rationality is adaptive. Or at least it is most of the time.

2.2. Bounded rationality and driver self assessment

The assessment of our own capabilities as drivers is relevant whether we think of driver behavior in terms of unbounded or bounded rationality. However, unbounded rationality assumes that this self-assessment is accurate. (Why would it not be, given access to all of the relevant information and unlimited information-processing resources?) But, to the extent that self-assessment might not be accurate, it would play a critical role in bounded rationality.

When asked to evaluate their driving skills, most drivers consider themselves to be better than the average driver. We found this with subjects from the US, Germany, and Spain (Sivak, Soler, & Tränkle, 1989). For example, when asked to rate themselves on a 5-point scale on how predictable they are as drivers, 0 percent of Americans rated themselves as “very unpredictable” (one extreme), while 30 percent rated themselves as “very predictable” (the other extreme). (The analogous percentages were 5 and 28 percent in Germany, and 0 and 22 percent in Spain.) The subjects gave similar, strongly asymmetrical distributions of responses concerning how relaxed, wise, and considerate they were as drivers. The same basic results were obtained in several other studies performed in 18 additional countries (Cauzard & Wittink, 1998; Groeger & Brown, 1989; McCormick, Walkey, & Green, 1986; Svenson, 1981). It is obvious that, as a consequence of the overestimation of one’s driving-related skills, bounded rationality can result in decisions that may be too risky.

3. Human factors: common sense and bounded rationality

The field of human factors is sometimes characterized as being unscientific, based on nothing more than common sense. Common sense, in turn, shares several key aspects with bounded rationality. First, to use common sense is to consider only the most salient features of the problem. Second, in using common sense, it is not important to examine all possible alternatives. Third, the
first alternative that meets the needs of the situation (the aspiration level) is accepted. Using common sense is designed to arrive at a quick solution.

Many early safety countermeasures could, indeed, be viewed as based on common sense. However, I will argue that as motorization progressed, common-sense countermeasures have been supplemented (and in some cases replaced) by countermeasures that required going well beyond common sense. Furthermore, additional common-sense countermeasures were clearly identified, but they are encountering implementation problems. Examples include lower nighttime speed limits, mandatory helmet-wearing laws not only for motorcyclists but also for bicyclists, and stricter laws concerning permissible blood alcohol concentration.

3.1. Traffic safety countermeasures: common-sense-based and science-based

I will now present an example of a common-sense vehicle countermeasure, and contrast it with a noncommon-sense countermeasure whose development required considerable expertise, insight, and formal validation. Soon after the introduction of headlamps on vehicles, people realized that they could be glaring to oncoming drivers. Consequently, a two-beam headlighting system was born: with high (driving) beams to be used when no oncoming traffic is nearby and low (passing) beams to be used otherwise. The problem with this simple, common-sense solution is that low beams do not provide enough illumination to drive safely at speeds above about 70 km/h. (This is because at higher speeds the braking distance is longer than the visibility distance.) That creates a complex human factors conflict between maximizing visibility and minimizing glare.

After more than fifty years of tweaking the light distribution within low-beam headlamps, it is clear that a fully satisfactory compromise between visibility and glare is not possible within the constraints of traditional incandescent illumination. A real solution would require a major conceptual change. An example of a potentially promising, noncommon-sense solution is polarized headlighting. Polarized lighting dates back to the 1920s. It “is based on the principle that polarized light is freely transmitted through a polarizing medium, the axis of which is parallel to the direction of the light’s polarization, but is [virtually] extinguished when the axis is at 90°” (Land, 1968, p. 334). One potential embodiment of this approach involves using in front of the headlights a polarizing material (the polarizer) with its axis at, say, 45° measured clockwise from the vertical. The driver views the road through a polarizing material (the analyzer) whose axis is set in the same direction as the polarizer, and consequently all of the polarized light is available. However, the polarized light from opposing headlamps is polarized at 45° counterclockwise from this head-on perspective. Thus, the light reaches the driver’s analyzer at an angle of 90° with respect to its axis, and is therefore almost completely extinguished.

This elegant approach to the dilemma of visibility vs. glare has some problems to overcome before it could replace current headlighting. Some of the unresolved problems concern light losses due to polarization, and transition issues with a mixture of polarized and unpolarized lighting in the fleet of vehicles on the road. It is mentioned here not because it is ready to be implemented, but because it is a potentially viable solution not based on common sense.

Other paired examples of common-sense and science-based vehicular countermeasures (see Table 2) include getting rid of sharp fins on the outside of vehicles to reduce pedestrian injuries vs. developing pedestrian airbags (outside airbags that deploy to provide soft structures for impending collisions with pedestrians); mandating minimum light transmittance of windshields
Table 2
Examples of common-sense and science-based countermeasures

<table>
<thead>
<tr>
<th>Common-sense countermeasure</th>
<th>Science-based countermeasure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low beam and high beam headlighting</td>
<td>Polarized headlighting</td>
</tr>
<tr>
<td>Getting rid of sharp fins</td>
<td>Pedestrian airbags</td>
</tr>
<tr>
<td>Minimum windshield transmittance</td>
<td>See-through A pillars</td>
</tr>
<tr>
<td>Dash-mounted displays</td>
<td>Head-up displays</td>
</tr>
<tr>
<td>No nearby trees and poles</td>
<td>Collapsible poles and channeling guardrails</td>
</tr>
<tr>
<td>Visual screening</td>
<td>Attentional screening ( UF_{ov} )</td>
</tr>
</tbody>
</table>

vs. developing see-through A pillars (the front pillars supporting the roof); and displaying information in the dashboard vs. providing the same information via a head-up display (in which virtual image appears outside the windshield, superimposed upon the external scene). In the driving environments, analogous pairs include removing poles and trees that are near the roadway vs. designing collapsible poles and guardrails that redirect vehicles back into traffic. Finally, in the driver domain, an example includes screening of visual capabilities vs. screening of peripheral attention, with the latter playing an increasingly diagnostic role with advanced age (e.g., Ball, Owsley, Sloane, Roenker, & Bruni, 1993).

4. Common-sense “solutions” that do not necessarily work as intended

4.1. Behavioral adaptation: risk compensation

Some common-sense “solutions” do not work as intended. One reason for this is behavioral adaptation. The extreme version of the theory of behavioral adaptation, called risk compensation or risk homeostasis, posits that drivers adjust their behavior to maintain a constant level of perceived risk. For example, this theory postulates that drivers reduce their speed when driving on ice in such a way that the resulting risk of being involved in an accident remains approximately the same as on dry pavement. Although there are serious reservations about this extreme version of behavioral adaptation, there is experimental evidence showing that drivers do adjust their behavior in response to perceived changes in risk, and that these behavioral changes are in the direction of maintaining a constant level of risk (Adams, 1985; Wilde, 1989).

Because of behavioral adaptation, the benefits of accident countermeasures of which drivers are aware are frequently smaller than would be expected a priori, as shown by the following three examples. The first example concerns studded tires. If one were to drive with studded tires at the same speed as without them, the increased friction with the studded tires would result in increased safety margins. However, it is possible that drivers would drive faster with studded tires, negating some or all of the safety benefits accrued by the increased friction. The results from a study by Rumar, Berggrund, Jernberg, and Ytterbom (1976) showed some (but incomplete) behavioral compensation. Specifically, on icy roads, drivers with studded tires drove somewhat faster than those without studded tires. Nevertheless, there was still a net benefit of having studded tires, as measured by the resulting safety margins.
The second example of behavioral adaptation deals with driver behavior when using ABS (antilock braking systems). These systems are designed to prevent the lockup of vehicle wheels during hard braking, and therefore allow effective steering even on slippery surfaces (Farmer, 2001). Consistent with behavioral adaptation, a recent study found that taxi drivers in cars equipped with ABS maintained shorter headways than those without ABS (Sagberg, Fosser, & Saetemo, 1997).

The third example of behavioral adaptation comes from a study by Kallberg (1993) on the effects of providing better nighttime lane delineation. A common-sense prediction would be that better visibility through better delineation could have only positive effects on safety. However, that is not what Kallberg found. Kallberg monitored speeds and accidents on sections of roads that were either equipped or not equipped with post-mounted delineators (retroreflectors raised substantially above the road surface) for increased visibility. The results showed that on lower-speed, curvy roadways, increased visibility led to faster speeds and more accidents—a net negative effect.

4.2. Selective degradation

Leibowitz and Owens (1977) have developed a theory that predicts that not all improvements in visibility will lead to increased safety. Their theory is based on the existence of two visual systems, focal and ambient. The focal system serves the central area of vision, is used for object detection, its usage in night driving is not frequent, and thus drivers' awareness of their own detection ability is low. On the other hand, the ambient system serves the visual periphery, is used for lane keeping, its usage in night driving is continuous, and thus the driver's awareness of performance is high. Importantly, the focal system is very sensitive to decreased illumination, while the ambient system is not.

A critical implication of the selective-degradation theory is that a countermeasure that improves ambient vision without improving focal vision might have a negative net effect on safety. For example, it follows from this theory that, in poor visibility conditions, we should not make lane tracking too easy by making the lane markings too reflective or by developing special lamps that would improve only lane visibility. This theory argues that such common-sense countermeasures could be potentially harmful, because the improvements in ambient vision would not be accompanied by improvements in focal vision, and the improved ambient vision would lead to an unjustifiable level of confidence concerning focal capabilities.

The selective degradation theory can be recast in terms of bounded rationality. A driver has continuous feedback from ambient vision (How well am I staying on the road?), while the feedback from focal vision (If there was one present, how well could I see a dark-clad pedestrian?) is infrequent. Within the concept of bounded rationality, this state of affairs leads to reliance on the feedback that is readily available (about the performance of ambient vision). But such behavior can have major negative consequences. Because we feel that lane maintenance is easy even at low light levels, we do not slow down at night. Consequently, we often overdrive our headlamps.

4.3. Licensing of older drivers

As we get older, our vision worsens, our reaction time lengthens, and we have more problems in complex decision-making situations. According to common sense, because these skills are related
to driving, older drivers should be worse drivers. Thus, there should be stricter licensing requirements for older drivers. It turns out that the two conclusions do not necessarily follow. There are two reasons for this. One reason is that, because there is no single measure of safety, the overall safety pattern is more complex (and more interesting). The other reason is that older drivers voluntarily reduce their driving in the most risky environments (for example, at night, in bad weather, on unknown roads, and routes requiring a left turn across other traffic), and they reduce their overall amount of driving as well.

Recent US data from Massie, Campbell, and Williams (1995) illustrate the complex nature of the safety of older drivers. These data show that the effect of driver age on the fatality rate per unit of distance is a U-shaped function. This way of looking at the situation, older drivers are less safe than middle-aged drivers, and the eldest are less safe than the youngest. (The youngest drivers are another story, and I will briefly come back to them later.) However, older drivers drive substantially less than middle-aged drivers. If one combines the increased risk per distance traveled with the decreased quantity and improved selection of exposure, the result is that the risk per year stays approximately the same between the ages of about 40 and 70.

I mentioned earlier that avoidance of left turns across traffic is an example of a compensatory action taken by older drivers. There is a good reason for this behavior. When making a decision about the safety of a gap in traffic, it is not only the distance but also the speed of the oncoming vehicle that is relevant. In other words, what is critical is not the gap in terms of distance but the gap in terms of time. In this context, Staplin, Lococo, and Sim (1993) have provided another example of bounded rationality at play. They found that in making judgments about the safety of left turns across traffic, older drivers relied primarily on distance in gaps. This fits well into the concept of bounded rationality: older drivers do not use all of the relevant information (distance and speed), but only the most salient information (distance).

Another interesting twist to this issue was provided by the results of a study by Hakamies-Blomqvist, Johansson, and Lundberg (1996). They compared accident involvement of older persons in Sweden and Finland—two countries that are comparable in many respects, but have substantially different approaches to licensing of older drivers. Sweden has very liberal laws, with no age-related screening connected to the license renewal. The renewal is only a formality. On the other hand, in Finland the licensing requirements get progressively more stringent with increased age. The results obtained by Hakamies-Blomqvist et al. were fascinating. The age-related patterns of accident involvement were similar in the two countries. However, fatalities among unprotected road users (pedestrians and bicyclists) increased more sharply with age in Finland than in Sweden. Hakamies-Blomqvist et al. concluded that the stricter age-related licensing in Finland resulted in shifts to even more risky modes of transportation.

4.4. Licensing of young drivers

As mentioned earlier, the youngest drivers have the highest fatality rates per distance traveled. To address this issue, graduated licensing has been introduced in several jurisdictions. Graduated licensing provides young drivers with a more extended period of supervised practice, and a more gradual immersion into full driving privileges. Several recent studies (e.g., Langley, Wagenaar, & Begg, 1996; Shope, Molnar, Elliott, & Waller, 2001) have shown that graduated licensing programs are, indeed, successful in reducing accident rates of young drivers.
Is the overinvolvement of young drivers in traffic accidents a consequence of chronological age or years of having a driver's license? In highly motorized countries, age and driving experience are highly correlated. Consequently, several studies have tried to untangle the relative contributions of these two factors by examining accident involvement of novice drivers of varying age. It turns out that both factors have an effect. For example, Cooper, Pinili, and Chen (1995) showed that accident involvement drops sharply and continuously for drivers between 16 and 25 years of age, but also that novice drivers of any age are more likely to be involved in accidents during the first year of driving than thereafter.

A noncommon-sense implication of these findings is that graduated licensing should be applied not only to young novice drivers, but also to novice drivers of any age. Indeed, such graduated licensing has been introduced in some jurisdictions (e.g., Nova Scotia). It recognizes the noncommon-sense fact that driving is a highly complex skill requiring extended practice to master regardless of the age when one starts to drive. Such a graduated licensing system has been shown to reduce accident rates for all novice drivers (Mayhew, Simpson, de Grooseilliers, & Williams, 2001).

5. Need for science-based safety countermeasures

As the level of motorization increases, the need for a science-based approach to addressing traffic safety problems increases as well. With increasing motorization (as measured by the number of vehicles per population), fatalities per vehicle drop exponentially. This relationship is often referred to as Smeed's law (Smeed, 1949). The magnitude of the effect of motorization was illustrated by Adams (1985) in his analysis of the 1980 data for 62 countries. The two extreme countries in his sample, in terms of motorization, were the US and Liberia. As Adams pointed out, if the US had had Liberia's fatality rate per vehicle, more than 6 million people would have been killed in 1980 (113 times the actual number).

What is it about motorization that brings down the fatality rate per vehicle? Motorization is a proxy for economic development represented by measures such as gross national product and income per capita. Economic development, in turn, provides the luxury of being able to pay attention to road safety problems. In addressing safety problems, the first step has been to pay attention to the most obvious problems, for which common-sense countermeasures are often sufficient. Higher economic development provides the means for having better-educated drivers, safer and better-maintained vehicles, more developed and safer road infrastructure, speedier and better medical care, more traffic-safety research, etc. It is likely that common-sense countermeasures, made possible by economic development, are a substantial reason for the fatality drop associated with Smeed's law.

In highly motorized countries, such as the US, however, the party is over. The easy problems have been addressed. The remaining problems (for example, driver aggression and driver fatigue) are more complex than the ones dealt with in the early stages of motorization. Furthermore, new highly complex issues are continuously being introduced by bringing new technology into the vehicle—technology designed ostensibly to help the driver (for example, mobile phones and smart cruise controls). Examples of these new issues include increased driver distraction and decreased driver vigilance, the latter related to relinquishing some of the control over the driving task to
technology. Thus, although past improvements were heavily based on common sense, future improvements will not be. Further progress will necessitate rigorous scientific inquiry.

6. Conclusions

Bounded rationality is a generally adaptive behavior characterized by a rapid decision-making process that involves only the most salient aspects of the problem. In the driving context, bounded rationality is evident on a macro (societal) level as well as on a micro (individual driver) level. On a macro level, bounded rationality underlined the development of many of the early, common-sense countermeasures for traffic safety problems. On a micro level, bounded rationality can lead to unexpected driver behaviors and thus to unexpected effects of some common-sense countermeasures.

Despite the fact that bounded rationality on the individual level partially negated the predicted benefits of some common-sense countermeasures, as motorization progressed, the application of common-sense countermeasures contributed to major improvements in road safety as measured by fatalities per distance traveled. However, because the obvious, common-sense countermeasures have largely been already identified (although not always implemented), further improvements will require insights rooted in comprehensive theories of driving behavior, and formal study of the implications of these insights. Further improvements will require real science; common sense will not do. Only the unbounded rationality of real science can take into account the bounded rationality of the driver.

References


