Information Design for Vehicle-to-Vehicle Communication

Brendan T. Gould and Philip N. Brown

Abstract

The emerging technology of Vehicle-to-Vehicle (V2V) communication over vehicular *ad hoc* networks promises to improve road safety by allowing vehicles to autonomously warn each other of road hazards. However, research on other transportation information systems has shown that informing only a subset of drivers of road conditions may have a perverse effect of increasing congestion. In the context of a simple (yet novel) model of V2V hazard information sharing, we ask whether partial adoption of this technology can similarly lead to undesirable outcomes. In our model, drivers individually choose how recklessly to behave as a function of information received from other V2V-enabled cars, and the resulting aggregate behavior influences the likelihood of accidents (and thus the information propagated by the vehicular network). We fully characterize the game-theoretic equilibria of this model using our new equilibrium concept. Our model indicates that for a wide range of the parameter space, V2V information sharing surprisingly increases the equilibrium frequency of accidents relative to no V2V information sharing, and that it may increase equilibrium social cost as well.

I. Introduction

Technology is becoming increasingly intertwined with the society it serves, accelerated by emerging paradigms such as the internet of things (IoT) and various smart infrastructure concepts such as vehicle-to-vehicle communication (V2V). It is no longer appropriate to design the merely-technical aspects of systems in isolation; rather, engineers must explicitly consider the implicit feedback loop between designed autonomy and human decision-making. As a piece of this process, recent research has asked when new technological solutions may cause more harm than good [2].

A clear example of this is the area of equilibrium traffic congestion under selfish individual behavior. This topic has been well researched, and it is commonly understood that equilibria associated with this behavior may not be optimal at the system level [3]–[8]. Many proposed solutions to this problem focus on the effects of deploying smart infrastructure to alleviate congestion and safety issues [9], using incentive design [10]–[12] and information design [13], [14] to improve upon selfish network routing. However, this technology does not always have its intended effect; for example, self-driving cars can exacerbate equilibrium traffic congestion [15].

Bayesian persuasion describes the process of a sender disclosing or obfuscating information in an attempt to influence the actions of other strategic agents [16]–[18]. However, it is often difficult to anticipate the reactions of these agents to this information, giving rise to counter-intuitive results. It is known that merely making a subgroup of a population aware of a new road in a network can increase the equilibrium cost to that group, a phenomenon known as informational Braess' paradox [19], [20]. In information design problems in general, full disclosure of information is not always optimal [7], [17], [21]–[24].

This naturally gives rise to the question of "What is the optimal information sharing policy?" Prior research has posed this question in the context of congestion games where each driver's cost depends on the selected route and the total mass of drivers on that route [5], [25]. However, prior work has been highly limited by the context of congestion games; driver cost is usually just travel time. Driver safety is rarely considered, and if it appears at all, it is only indirectly through the effect it has on travel time [8], [11]. Since driver safety is certainly a serious concern in its own right, our work is designed to address this important gap in the literature.

In this paper, we initiate a study on the incentive effects of distributed hazard information sharing by V2V-equipped vehicles, and ask when maximal sharing optimizes driver safety. We pose a simple model of information sharing with partial V2V adoption; that is, some vehicles are unable to receive signals warning of road hazards. In contrast to existing literature, our model allows for *endogenous* road hazards where the likelihood of a road hazard is dependent on the aggregate recklessness of drivers.

After fully characterizing the emergent behavior in terms of a new equilibrium concept we call a *signaling* equilibrium (Theorem 3.1), our main result in Theorem 3.3 shows that there exist parameter regimes in which the optimal hazard signaling rate is 0; that is, sharing any road hazard information with only V2V-equipped vehicles leads to a higher frequency of accidents than sharing none. Further, we show that the optimal hazard signaling rate is *always* either 0 or 1, and provide a criterion to determine which of the two is correct for any parameter combination. This makes it very simple to implement the accident-minimizing policy predicted by our model using real world V2V technology.

A preliminary version of this work appeared in [1].

Brendan T. Gould and Philip N. Brown are with the department of Computer Science at the University of Colorado Colorado Springs, Colorado Springs, CO, USA {bgould2,pbrown2}@uccs.edu

This work was supported in part by the Undergraduate Research Academy at the University of Colorado Colorado Springs and in part by NSF Award #ECCS-2013779.

We then close with a consideration of the relationship between the *social cost* and hazard signaling rate. Social cost is the total expected cost experienced by all members of the population from any source. We show that minimizing accident probability is sometimes fundamentally opposed to minimizing social cost: a signaling policy which decreases the frequency of accidents may necessarily increase the social cost (and vice-versa).

II. Model and Performance Metrics

A. Game Setup

We adopt a nonatomic game formulation; i.e., we model a population of drivers as a continuum in which each of the infinitely-many drivers makes up an infinitesimally small portion of the population. This population makes decisions and interacts on a single road. Each driver can choose to drive carefully (C), or recklessly (R), and a traffic accident either occurs (A) or does not occur (¬A). This choice encapsulates many measures of driver behavior, such as acceleration, lane changing, following distance, etc. It is not meant to precisely model any measure individually, but rather to give a simplified idea of the overall decisions made and risks incurred by drivers. Careful drivers consistently choose the slower, safer behaviors, and reckless drivers make faster, riskier choices.

Because of this, careful drivers are able to detect and avoid any existing accidents, while reckless drivers will "pile on" and experience an expected accident cost of $r > 1$. However, if an accident is not present, careful drivers regret their caution (e.g., due to the longer trip time incurred) and experience a regret cost of 1. These costs are collected in this matrix:

	Accident (A)	No Accident (¬A)
Careful (C)	0	1
Reckless (R)	r	0

We write $d \in [0,1]$ to denote the overall fraction of drivers choosing to drive recklessly, and $p(d)$ to represent the resulting probability that an accident occurs. Throughout the manuscript, we assume that more reckless drivers make an accident strictly more likely, so that $p(d)$ is strictly increasing. Additionally, we assume that $p(d)$ is continuous.

We model partial V2V adoption, i.e. some fraction $y \in [0,1]$ of drivers have cars equipped with V2V technology. When these drivers encounter road hazards or traffic accidents, the technology may autonomously detect these hazards and broadcast warning signals. If an accident has occurred, we say that V2V technology will detect the accident with probability $t(y)$. When an accident is detected, the technology will broadcast a signal that is received by all other V2V cars. Furthermore, if no accident has occurred, we allow for the possibility that V2V technology incorrectly broadcasts a "false-positive" signal; this happens with probability $f(y)$. We assume that the technology broadcasts more true positives than false positives, i.e. $f(y) < t(y)$.

In many models of transportation information systems, it is known that distributing perfect information can actually make parts or all of the population worse off [19], [21]–[23]; that is, the *information design* problem is nontrivial: in some scenarios it may be optimal to withhold information from drivers. Accordingly, we wish to study the information design problem faced by the administrators of V2V technology. Therefore, let S be the event that a V2V car displays a warning to its driver, and B be the event that a warning signal has been broadcast. Then, we define a disclosure rate $\beta := \mathbb{P}(\mathrm{S}|\mathrm{B}) \in [0,1]$ as the probability that a broadcast signal is revealed to the driver.

This signaling scheme divides the population into three groups. We call a driver a *non-V2V driver* if their vehicle lacks V2V technology, and a V2V driver otherwise. We further differentiate V2V drivers by whether they have seen a warning signal, calling them *unsignaled V2V drivers* if they have not seen a warning and *signaled V2V drivers* if they have. We write x_{n}, x_{vu}, and x_{vs} where $x_{\mathrm{n}}, x_{\mathrm{vu}}, x_{\mathrm{vs}} \in [0,1]$ to represent the mass of reckless drivers in each group, respectively, and a behavior profile as $x = (x_{\mathrm{n}}, (x_{\mathrm{vu}}, x_{\mathrm{vs}}))$.

We assume that both non-V2V and V2V drivers have habitual behaviors x^*_{n} and x^*_{v}, respectively; these behaviors could come from repeated experiences driving on that road. Initially, each group of drivers makes their behavior decision according to their habits, and this behavior determines the probability of an accident. Non-V2V drivers receive no information that could lead them to change their behavior, and are thus effectively committed to their initial choice, i.e. $x_{\mathrm{n}} = x^*_{\mathrm{n}}$. However, V2V drivers are able to adjust their behavior based on whether or not they see a warning signal; choosing unsignaled behavior x_{vu} when they do not see a warning, and signaled behavior x_{vs} when they do. The habitual behavior of V2V drivers must be a weighted average of their behavior when they do and do not see warnings, i.e.

$$x^*_{\mathrm{v}} = \mathbb{P}(\neg\mathrm{S})x_{\mathrm{vu}} + \mathbb{P}(\mathrm{S})x_{\mathrm{vs}}. \tag{1}$$

Figure 1 displays the overall timeline of events and decisions in our model.

Define $P(x)$ to be the probability that an accident occurs, given some behavior profile x. Additionally define $Q(x)$ as the probability that a given V2V driver sees a warning light, given the same. When the dependence on x is clear from context, we will sometimes write simply P and Q. Then, $\mathbb{P}(\mathrm{A}) = P(x) = p(x^*_{\mathrm{n}} + x^*_{\mathrm{v}})$ and $\mathbb{P}(\mathrm{S}) = Q(x) = \beta(P(x)t(y) + (1 - P(x))f(y))$. Note that $P(x)$ is implicitly parameterized by β and y. Substituting this into (1) gives

$$P(x) = p(x_{\mathrm{n}} + (1 - Q(x))x_{\mathrm{vu}} + Q(x)x_{\mathrm{vs}}). \tag{2}$$

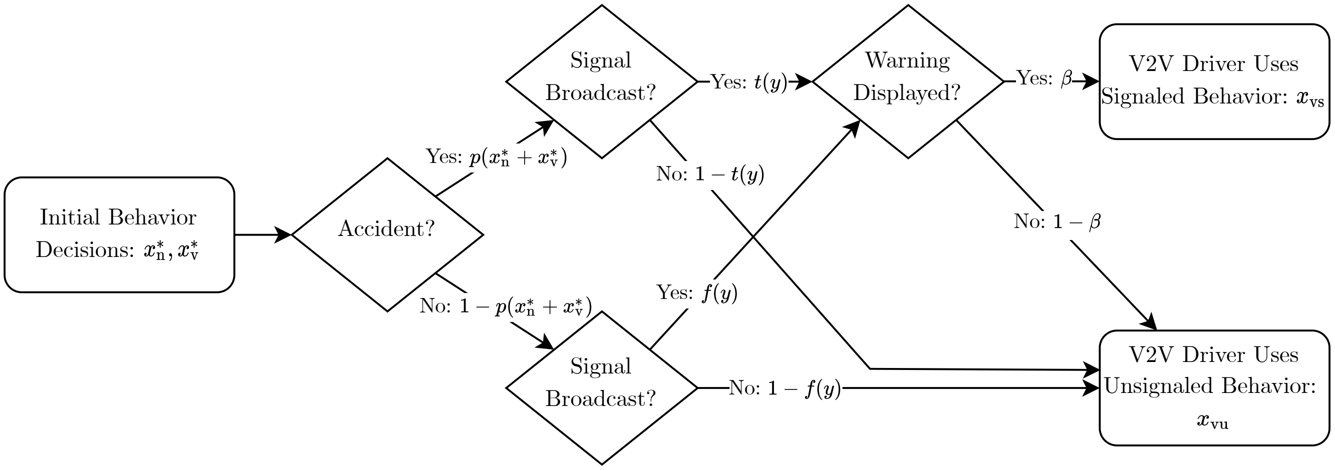


Fig. 1: Timeline of behavior decisions

Proposition 2.1: For all parameter combinations, (2) has at least one solution for $P(x)$.
See Appendix A for a proof of Proposition 2.1.

We write $J_{\mathrm{n}}(a;x)$ to denote the expected cost to a non-V2V driver choosing action $a \in \{\mathrm{C}, \mathrm{R}\}$, and similarly $J_{\mathrm{vu}}(a;x)$ and $J_{\mathrm{vs}}(a;x)$ for unsignaled and signaled V2V drivers, respectively. These costs are given by

$$J_{\mathrm{n}}(a;x) = \begin{cases} 1 - P(x) & \text{if } a = \mathrm{C}, \\ rP(x) & \text{if } a = \mathrm{R}, \end{cases} \tag{3}$$

$$J_{\mathrm{vu}}(a;x) = \begin{cases} 1 - \mathbb{P}(\mathrm{A}|\neg\mathrm{S}) & \text{if } a = \mathrm{C}, \\ r\mathbb{P}(\mathrm{A}|\neg\mathrm{S}) & \text{if } a = \mathrm{R}, \end{cases} \tag{4}$$

$$J_{\mathrm{vs}}(a;x) = \begin{cases} 1 - \mathbb{P}(\mathrm{A}|\mathrm{S}) & \text{if } a = \mathrm{C}, \\ r\mathbb{P}(\mathrm{A}|\mathrm{S}) & \text{if } a = \mathrm{R}. \end{cases} \tag{5}$$

Finally, we define a signaling game as the tuple $G = (\beta, y, r)$.

B. Signaling Equilibrium

We define a signaling equilibrium as a behavior profile $x^{\mathrm{ne}} = (x_{\mathrm{n}}^{\mathrm{ne}}, (x_{\mathrm{vu}}^{\mathrm{ne}}, x_{\mathrm{vs}}^{\mathrm{ne}}))$ with $0 \le x_{\mathrm{n}}^{\mathrm{ne}}, x_{\mathrm{vs}}^{\mathrm{ne}}, x_{\mathrm{vu}}^{\mathrm{ne}} \le 1$ satisfying the following:

$$x_{\mathrm{n}}^{\mathrm{ne}} < 1 - y \implies J_{\mathrm{n}}(\mathrm{C}; x^{\mathrm{ne}}) \le J_{\mathrm{n}}(\mathrm{R}; x^{\mathrm{ne}}), \tag{6}$$
$$x_{\mathrm{n}}^{\mathrm{ne}} > 0 \implies J_{\mathrm{n}}(\mathrm{R}; x^{\mathrm{ne}}) \le J_{\mathrm{n}}(\mathrm{C}; x^{\mathrm{ne}}), \tag{7}$$
$$x_{\mathrm{vu}}^{\mathrm{ne}} < y \implies J_{\mathrm{vu}}(\mathrm{C}; x^{\mathrm{ne}}) \le J_{\mathrm{vu}}(\mathrm{R}; x^{\mathrm{ne}}), \tag{8}$$
$$x_{\mathrm{vu}}^{\mathrm{ne}} > 0 \implies J_{\mathrm{vu}}(\mathrm{R}; x^{\mathrm{ne}}) \le J_{\mathrm{vu}}(\mathrm{C}; x^{\mathrm{ne}}), \tag{9}$$
$$x_{\mathrm{vs}}^{\mathrm{ne}} < y \implies J_{\mathrm{vs}}(\mathrm{C}; x^{\mathrm{ne}}) \le J_{\mathrm{vs}}(\mathrm{R}; x^{\mathrm{ne}}), \tag{10}$$
$$x_{\mathrm{vs}}^{\mathrm{ne}} > 0 \implies J_{\mathrm{vs}}(\mathrm{R}; x^{\mathrm{ne}}) \le J_{\mathrm{vs}}(\mathrm{C}; x^{\mathrm{ne}}). \tag{11}$$

Equations (6)-(11) enforce the standard conditions of a Nash equilibrium (i.e. if players are choosing any action, its cost to them is minimal). The novelty of this concept lies in the fact that we endogenously determine the likelihood of a signal, and therefore the mass of signaled and unsignaled V2V drivers, using accident probability at equilibrium. But accident probability is determined by driver behavior, which is in turn influenced by the probability of a signal. This creates a complex interdependence between driver behavior and signal probability, which is captured by our consistency equation (2).

Additionally, we define social cost as the expected individual cost given behavior:

$$\begin{aligned} S(x) &= J_{\mathrm{n}}(\mathrm{C}; x)(1 - y - x_{\mathrm{n}}) + J_{\mathrm{n}}(\mathrm{R}; x)x_{\mathrm{n}} \\ &\quad + (1 - Q)(J_{\mathrm{vu}}(\mathrm{C}; x)(y - x_{\mathrm{vu}}) + J_{\mathrm{vu}}(\mathrm{R}; x)x_{\mathrm{vu}}) \\ &\quad + Q(J_{\mathrm{vs}}(\mathrm{C}; x)(y - x_{\mathrm{vs}}) + J_{\mathrm{vs}}(\mathrm{R}; x)x_{\mathrm{vs}}). \end{aligned} \tag{12}$$

Proposition 2.2: For every signaling game G, there exists a signaling equilibrium $x^{\rm ne}$ and it is essentially unique. By this we mean that for any two signaling equilibria $x^{\rm ne}_1$ and $x^{\rm ne}_2$ of G, both of the following hold:

$$x^{\rm ne}_{\rm n1} + (1 - Q(x^{\rm ne}_1))x^{\rm ne}_{\rm vu1} + Q(x^{\rm ne}_1)x^{\rm ne}_{\rm vs1} = x^{\rm ne}_{\rm n2} + (1 - Q(x^{\rm ne}_2))x^{\rm ne}_{\rm vu2} + Q(x^{\rm ne}_2)x^{\rm ne}_{\rm vs2}, \tag{13}$$

$$P(x^{\rm ne}_1) = P(x^{\rm ne}_2). \tag{14}$$

We provide a proof of this fact in Lemma 4.2. A tedious series of arguments can show that (13) and (14) imply $S(x^{\rm ne}_1) = S(x^{\rm ne}_2)$, further supporting the notion that these equilibria are effectively the same. Throughout this paper, we use the terms "unique" and "essentially unique" interchangeably to mean (13) and (14) are satisfied.[1]

The notation we use in our model is summarized for convenience in Table I.

C. Research Objectives

Our first objective is to characterize the signaling equilibria of any game G. In Theorem 3.1, we show that every game G has an essentially unique signaling equilibrium. Additionally, we show that receiving a signal makes V2V drivers more cautious and not receiving a signal makes them more reckless at equilibrium, compared to non-V2V drivers.

Next, we seek to optimize accident probability and social cost by means of signal display probability. To that end, we abuse notation and write $P(G)$ to denote $P(x^{\rm ne})$ and $S(G)$ to denote $S(x^{\rm ne})$ where $x^{\rm ne}$ is a signaling equilibrium of game G. We then wish to find values for β_P and β_S such that

$$\beta_P \in \mathop{\arg\min}_{\beta \in [0,1]} P(G), \tag{15}$$

$$\beta_S \in \mathop{\arg\min}_{\beta \in [0,1]} S(G). \tag{16}$$

In Theorem 3.3, we provide an algorithm to determine a solution to (15), and show that there exist games where $\beta_P = 0$ is a solution, as depicted in Figure 2. Furthermore, in Proposition 3.8, we provide sufficient criteria for when $\beta_S = 1$ is a solution to (16), and show that there paradoxically exist regions of the parameter space where $\beta_S = 1$ is *not* a solution to (16). Furthermore, each of these paradoxes can still occur if $f(y) \equiv 0$, indicating that poor quality technology is not their sole cause. See Appendix B for a list of examples of the paradoxes in this case.

TABLE I: Table of key notation used in our model.

Symbol	Description
β	Probability that a V2V car that has received a warning signal will display a warning to its driver
y	Mass of the population driving V2V equipped vehicles
r	Expected cost of an accident
$t(y)$	Probability that a warning signal is broadcast to V2V cars, given that an accident has occurred
$f(y)$	Probability that a warning signal is broadcast to V2V cars, given that *no* accident has occurred
$x = (x_{\rm n}, (x_{\rm vu}, x_{\rm vs}))$	Behavior profile; mass of reckless drivers in each behavior group
$p(d)$	Probability of an accident as a function of the mass d of reckless drivers
$P(x)$	Probability of an accident as a function of a behavior profile
$Q(x)$	Probability that a warning signal is broadcast, as a function of a behavior profile
$S(x)$	Social cost as a function of a behavior profile

III. Our Contributions

A. Equilibrium Characterization

A signaling equilibrium takes the form of a tuple listing the mass of reckless drivers in each of our three population groups. These masses implicitly determine an equilibrium crash probability through (2). Though this relationship is complicated, our first theorem shows that an equilibrium is uniquely determined by any given parameter combination.

Theorem 3.1: For any V2V signaling game $G = (\beta, y, r)$, a signaling equilibrium exists and is essentially unique. In particular, the equilibrium $x^{\rm ne} = (x^{\rm ne}_{\rm n}, (x^{\rm ne}_{\rm vu}, x^{\rm ne}_{\rm vs}))$ can take one of the following 4 forms:

- $(0, (0, 0))$,
- $(0, (\chi_{\rm vu}, 0))$, for some $\chi_{\rm vu} \in [0, y]$,
- $(\chi_{\rm n}, (y, 0))$, for some $\chi_{\rm n} \in [0, 1 - y]$,
- $(1 - y, (y, \chi_{\rm vs}))$, for some $\chi_{\rm vs} \in [0, y]$.

[1] This definition is motivated by the degenerate cases where $\beta = 0$. If this is the case, then $Q = 0$, meaning V2V cars will never receive signals or display warnings to their drivers. Therefore, there is no real distinction between non-V2V drivers and V2V drivers, which causes many different behavior tuples to be effectively identical.

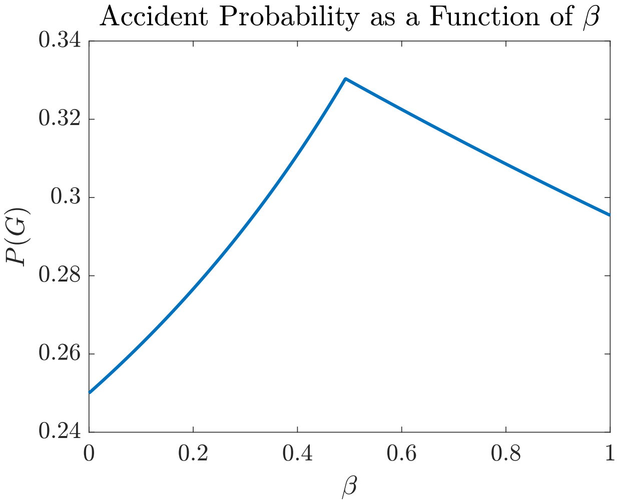


Fig. 2: Equilibrium accident frequency with respect to the signal display probability β. Note that when $\beta < 0.5$, displaying more warning signals steeply increases the frequency of accidents, and that even the maximum possible signal display probability $\beta = 1$ yields a higher accident frequency than if warning signals were never shown to drivers. The example depicted has accident probability characterized by $p(d) = 0.3d + 0.1$, signal probability characterized by $t(y) = 0.8y$ and $f(y) = 0.1y$, V2V penetration $y = 0.9$, and accident cost $r = 3$.

Theorem 3.1 captures several important characteristics of signaling equilibria. Chief among these is the fact that a signaling equilibrium exists and is essentially unique for every game $G = (\beta, y, r)$. Additionally, note that there is an “ordering” to recklessness: every unsignaled V2V driver must be reckless before any non-V2V driver can be, and all non-V2V drivers must be reckless before any signaled V2V driver can. This is because V2V technology has made signaled V2V drivers more confident that an accident has occurred, and therefore more likely to drive carefully than non-V2V drivers. Similarly, unsignaled V2V drivers have an extra measure of confidence that an accident has *not* occurred, and are therefore more likely to drive recklessly than non-V2V drivers.

The proof of Theorem 3.1 appears at length in Section IV. However, we provide here a key necessary condition of any signaling equilibrium; it serves as a cornerstone of the proof and will be instrumental in later results:

Lemma 3.2: For any signaling game $G = (\beta, y, r)$, a behavior profile $x^{\rm ne} = (x^{\rm ne}_{\rm n}, (x^{\rm ne}_{\rm vu}, x^{\rm ne}_{\rm vs}))$ is a signaling equilibrium if the following hold:

$$x^{\rm ne}_{\rm n} = \begin{cases} 0 & \text{if } P(x^{\rm ne}) > \frac{1}{1+r}, \\ p^{-1}(P(x^{\rm ne})) - (1 - Q(x^{\rm ne}))y & \text{if } P(x^{\rm ne}) = \frac{1}{1+r}, \\ 1 - y & \text{if } P(x^{\rm ne}) < \frac{1}{1+r}, \end{cases} \tag{17}$$

$$x^{\rm ne}_{\rm vu} = \begin{cases} 0 & \text{if } \mathbb{P}({\rm A}|\neg{\rm S}) > \frac{1}{1+r}, \\ \frac{p^{-1}(P(x^{\rm ne}))}{1-Q(x^{\rm ne})} & \text{if } \mathbb{P}({\rm A}|\neg{\rm S}) = \frac{1}{1+r}, \\ y & \text{if } \mathbb{P}({\rm A}|\neg{\rm S}) < \frac{1}{1+r}. \end{cases} \tag{18}$$

$$x^{\rm ne}_{\rm vs} = \begin{cases} 0 & \text{if } \mathbb{P}({\rm A}|{\rm S}) > \frac{1}{1+r}, \\ \frac{p^{-1}(P(x^{\rm ne}))-1+Q(x^{\rm ne})y}{Q(x^{\rm ne})} & \text{if } \mathbb{P}({\rm A}|{\rm S}) = \frac{1}{1+r}, \\ y & \text{if } \mathbb{P}({\rm A}|{\rm S}) < \frac{1}{1+r}. \end{cases} \tag{19}$$

Furthermore, this equilibrium is essentially unique (i.e all signaling equilibria that exist satisfy (13) and (14)).

The remainder of this section contains a brief overview of techniques and notation that provide useful understanding of the problem; a full proof of Lemma 3.2 and Theorem 3.1 is contained in Section IV.

It can be shown using Bayes' Theorem that

$$\mathbb{P}({\rm A}|\neg{\rm S}) \begin{array}{c}<\\=\\>\end{array} \frac{1}{1+r} \iff P(x) \begin{array}{c}<\\=\\>\end{array} \frac{1-\beta f(y)}{1+r(1-\beta t(y)) - \beta f(y)}, \tag{20}$$

$$\mathbb{P}({\rm A}|{\rm S}) \begin{array}{c}<\\=\\>\end{array} \frac{1}{1+r} \iff P(x) \begin{array}{c}<\\=\\>\end{array} \frac{f(y)}{rt(y) + f(y)}. \tag{21}$$

We use this notation to mean that any of the relationships between the first expressions is equivalent to the corresponding relationship between the second expressions. Equality and both inequalities are preserved.

Each of the above expressions acts as a threshold on the behavior of a particular group of drivers. For example, if $P(x) < \frac{1-\beta f(y)}{1+r(1-\beta t(y))-\beta f(y)}$, then by (20), $\mathbb{P}(\mathrm{A}|\neg\mathrm{S}) < \frac{1}{1+r}$. Then, by Lemma 3.2, $x_{\mathrm{vu}}^{\mathrm{ne}} = y$, meaning all unsignaled V2V cars are reckless. Similarly, if $P(x) > \frac{1-\beta f(y)}{1+r(1-\beta t(y))-\beta f(y)}$, (20) gives that $\mathbb{P}(\mathrm{A}|\neg\mathrm{S}) > \frac{1}{1+r}$, so Lemma 3.2 guarantees that all unsignaled V2V cars are careful.

In the same way, the expression $\frac{f(y)}{rt(y)+f(y)}$ acts as a threshold on the behavior of signaled V2V drivers, and $\frac{1}{1+r}$ does for non-V2V drivers. For convenience, we use the following shorthand to reference these thresholds:

$$P_{\mathrm{vs}} := \frac{f(y)}{rt(y)+f(y)}, \quad P_{\mathrm{n}} := \frac{1}{1+r}, \quad P_{\mathrm{vu}} := \frac{1-\beta f(y)}{1+r(1-\beta t(y))-\beta f(y)}, \tag{22}$$

where it holds that:

$$P_{\mathrm{vs}} < P_{\mathrm{n}} \leq P_{\mathrm{vu}}. \tag{23}$$

Intuitively, this ordering corresponds to the fact that each group of drivers has different information about the world. Unsignaled drivers are more confident that an accident has not occurred (because they receive a signal some of the time that accidents do occur), and are therefore willing to "risk" driving recklessly at higher inherent accident probabilities than non-V2V drivers are. The opposite is true for signaled V2V drivers.

Based on these thresholds, we can determine the essentially unique equilibrium behavior tuple for any accident probability P. The sets below correspond to regions of parameter space that constrain the equilibrium accident probability with respect to our behavior thresholds (Lemma 4.1), and therefore allow us to compute behavior directly from the model parameters (Lemma 4.2).

$$E_1 := \{(\beta, y, r) : P_{\mathrm{vu}} < p(0)\} \tag{24}$$

$$E_2 := \{(\beta, y, r) : p(0) \leq P_{\mathrm{vu}} \leq p((1-\beta P_{\mathrm{vu}}(t(y)-f(y))-\beta f(y))y)\} \tag{25}$$

$$E_3 := \{(\beta, y, r) : p((1-\beta P_{\mathrm{vu}}(t(y)-f(y))-\beta f(y))y) < P_{\mathrm{vu}} \wedge P_{\mathrm{n}} < p((1-\beta P_{\mathrm{n}}(t(y)-f(y))-\beta f(y))y)\} \tag{26}$$

$$E_4 := \{(\beta, y, r) : p((1-\beta P_{\mathrm{n}}(t(y)-f(y))-\beta f(y))y) \leq P_{\mathrm{n}} \leq p(1-(\beta P_{\mathrm{n}}(t(y)-f(y))+\beta f(y))y)\} \tag{27}$$

$$E_5 := \{(\beta, y, r) : p(1-(\beta P_{\mathrm{n}}(t(y)-f(y))+\beta f(y))y) < P_{\mathrm{n}} \wedge P_{\mathrm{vs}} < p(1-(\beta P_{\mathrm{vs}}(t(y)-f(y))+\beta f(y))y)\} \tag{28}$$

$$E_6 := \{(\beta, y, r) : p(1-(\beta P_{\mathrm{vs}}(t(y)-f(y))+\beta f(y))y) \leq P_{\mathrm{vs}} \leq p(1)\} \tag{29}$$

$$E_7 := \{(\beta, y, r) : p(1) < P_{\mathrm{vs}}\} \tag{30}$$

These sets are visualized for particular slices of parameter space in Figure 3.

Finally, the equilibrium behavior within each range is consistent with the form claimed in Theorem 3.1. This completes our conceptual overview of the characterization result; again, see Section IV for the complete proof.

B. Information Design for Minimizing Accident Probability

Though common intuition would suggest that increasing the quality of information given to drivers would allow them to make more informed decisions and arrive at less costly outcomes, prior research has shown that this is not always the case [19], [20]. Because of this, it is a non-trivial question for V2V administrators to decide the optimal quality of information to distribute. This quality will be certainly be bound by technical limitations, but administrators can freely manipulate it within the feasible range.

When a car with V2V technology receives a warning signal, it does not necessarily have to display a warning light to its driver. Paradoxically, we show that ignoring accidents in this way can decrease accident probability at equilibrium.

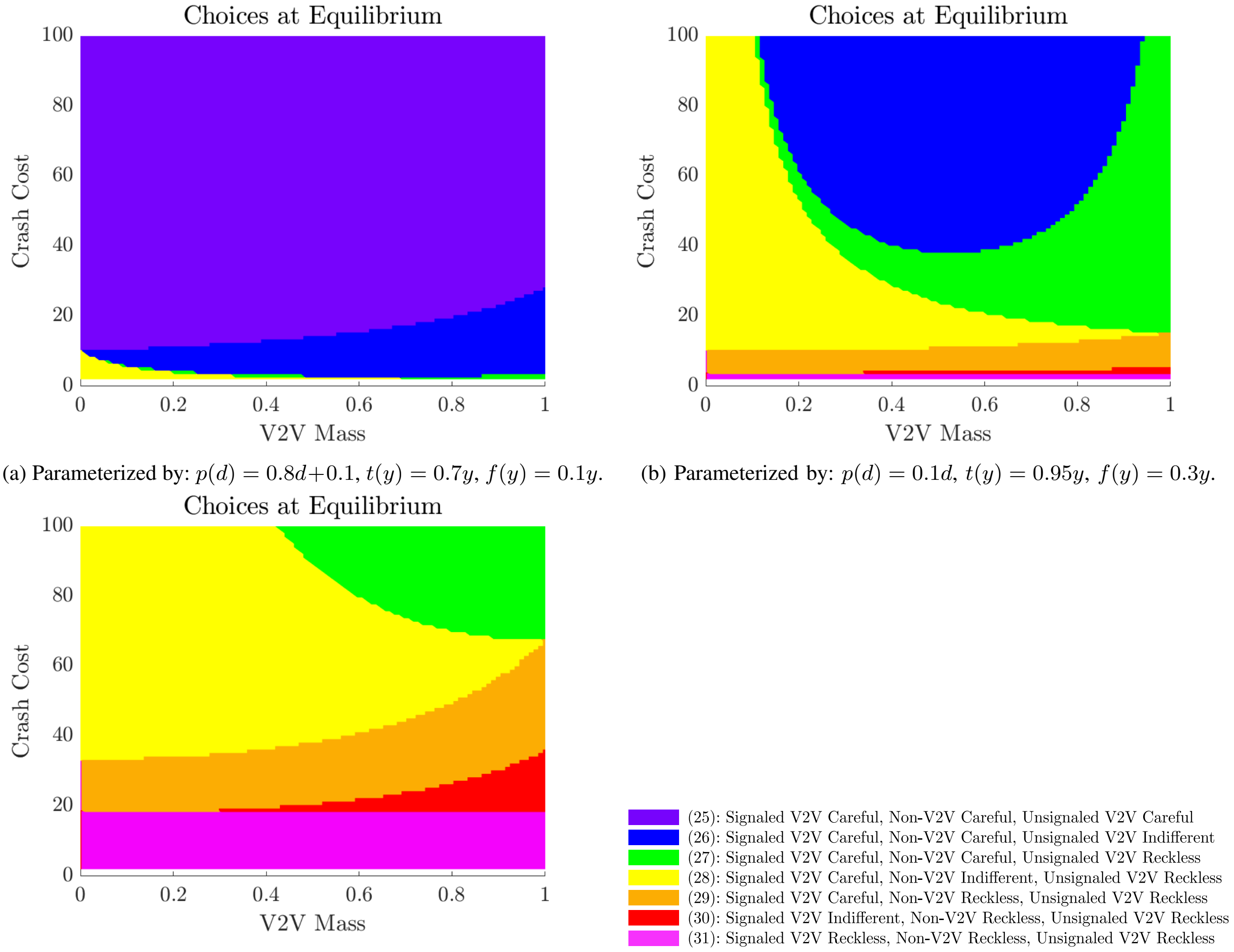
Theorem 3.3: For any signaling game $G = (\beta, y, r)$, we must have that either:

$$0 \in \arg\min_{\beta \in [0,1]} P(G) \text{ or } 1 \in \arg\min_{\beta \in [0,1]} P(G). \tag{31}$$

Furthermore, there exist signaling games where $\beta = 1$ does not minimize accident probability.

In other words, the minimum accident probability is guaranteed to be caused by never displaying warnings, or by displaying them as often as technologically possible. The proof of Theorem 3.3 proceeds as follows:

- First, Lemma 3.4 shows that the feasible equilibrium ranges in any game must satisfy a particular ordering with respect to β.
- Next, Lemma 3.5 uses this ordering to show that the probability of an accident is weakly increasing for low values of β, and weakly decreasing otherwise. Then, the smallest possible value of β will always be a minimum within the increasing range, and the largest value of β must be a minimum in the decreasing range.
- Therefore, one of the two must be a global minimum, completing the proof of Theorem 3.3.



(a) Parameterized by: $p(d) = 0.8d+0.1$, $t(y) = 0.7y$, $f(y) = 0.1y$.

(b) Parameterized by: $p(d) = 0.1d$, $t(y) = 0.95y$, $f(y) = 0.3y$.

(c) Parameterized by: $p(d) = 0.03d$, $t(y) = 0.95y$, $f(y) = 0.5y$.

Fig. 3: Equilibrium ranges. With $\beta = 1$ fixed, these plots show the equilibrium range that is valid for a combination of the parameters y and r. Each colored range corresponds to one of E_1 to E_7.

Lemma 3.4: For any combination of the parameters y and r, let $\beta_1, \beta_2 \in [0, 1]$ and $\beta_1 < \beta_2$. Additionally, let $G_1 = (\beta_1, y, r)$ and $G_2 = (\beta_2, y, r)$. Then the equilibrium type that is valid for G_1 and G_2 is ordered by β. Specifically, for any integer $i \in [1, 7]$,

$$G_1 \in E_i \implies G_2 \in E_j \tag{32}$$

for some integer $j \in [i, 7]$.

Proof: Let $P_{\text{vu1}} = \frac{1-\beta_1 f(y)}{1+r(1-\beta_1 t(y))-\beta_1 f(y)}$ and $P_{\text{vu2}} = \frac{1-\beta_2 f(y)}{1+r(1-\beta_2 t(y))-\beta_2 f(y)}$, and note that $P_{\text{vu1}} \leq P_{\text{vu2}}$.

If $G_1 \in E_2$, then $G_2 \notin E_1$. If it was, then we would have

$$P_{\text{vu2}} < p(0) \leq P_{\text{vu1}},$$

which is a clear contradiction. Therefore, by Theorem 3.1, $G_2 \in \bigcup_{i=2}^{7} E_i$, which is the desired result.

Similarly, if $G_1 \in E_3$, then $G_2 \notin E_1 \cup E_2$. If $G_2 \in E_1$, we would have that

$$P_{\text{vu2}} < p(0) \leq p((1 - \beta_1 P_{\text{vu1}}(t(y) - f(y)) - \beta_1 f(y))y) < P_{\text{vu1}},$$

and if $G_2 \in E_2$,

$$P_{\text{vu2}} \leq p((1 - \beta_2 P_{\text{vu2}}(t(y) - f(y)) - \beta_2 f(y))y) \leq p((1 - \beta_1 P_{\text{vu1}}(t(y) - f(y)) - \beta_1 f(y))y) < P_{\text{vu1}}.$$

(The second inequality holds since $\beta_1 < \beta_2$, P_{vu} is increasing in β, and p is an increasing function.) In any case, we have a contradiction. Again by Theorem 3.1, $G_2 \in \bigcup_{i=3}^{7} E_i$, completing the proof in this case.

This technique can be extended using (23) to cover the cases where $G_1 \in E_4$, $G_1 \in E_5$, $G_1 \in E_6$, or $G_1 \in E_7$. In any case, it is a matter of assuming G_2 is not contained in the claimed sets, finding a string of inequalities that leads to a contradiction, and applying Theorem 3.1 to achieve the desired result. ■

We will now provide a full definition of $P(G)$, the function optimized by (15). It is always true that $P(G) = P(x^{\mathrm{ne}})$, but using the ranges defined by (24)-(30), we can be more specific. For any signaling game $G = (\beta, y, r)$, the probability of an accident at its unique signaling equilibrium x^{ne} is given by

$$P(G) = \begin{cases} p(0) & \text{if } G \in E_1, \\ P_{\mathrm{vu}} & \text{if } G \in E_2, \\ P(x^{\mathrm{ne}}) & \text{if } G \in E_3, \\ P_{\mathrm{n}} & \text{if } G \in E_4, \\ P(x^{\mathrm{ne}}) & \text{if } G \in E_5, \\ P_{\mathrm{vs}} & \text{if } G \in E_6, \\ p(1) & \text{if } G \in E_7. \end{cases} \tag{33}$$

This simplification is due to Lemma 4.1. Since there exists a signaling equilibrium for any game G, this function is defined for all parameter combinations. We now present a piecewise monotonicity result on $P(G)$ with respect to β.

Lemma 3.5: There exists a $\bar{\beta}$ such that for $\beta \le \bar{\beta}$, $P((\beta, y, r))$ is weakly increasing in β, and for $\bar{\beta} < \beta$, $P((\beta, y, r))$ is weakly decreasing in β. Unfortunately, the problem does not permit a closed form expression where this $\bar{\beta}$ is isolated; however, $\bar{\beta}$ is the quantity that satisfies

$$\frac{1-\bar{\beta}f(y)}{1+r(1-\bar{\beta}t(y))-\bar{\beta}f(y)} = p\left(\left(1-\bar{\beta}\frac{1-\bar{\beta}f(y)}{1+r(1-\bar{\beta}t(y))-\bar{\beta}f(y)}(t(y)-f(y))-\bar{\beta}f(y)\right)y\right).$$

Proof: For any combination of the parameters y and r, let $\beta_1, \beta_2 \in [0,1]$ and $\beta_1 < \beta_2$. Let $G_1 = (\beta_1, y, r)$ and $G_2 = (\beta_2, y, r)$. Our approach is a exhaustive comparison of crash probabilities within and between the cases of (33). If

$$\frac{1-\beta_2 f(y)}{1+r(1-\beta_2 t(y))-\beta_2 f(y)} \le p\left(\left(1-\beta_2\frac{1-\beta_2 f(y)}{1+r(1-\beta_2 t(y))-\beta_2 f(y)}(t(y)-f(y))-\beta_2 f(y)\right)y\right),$$

then accident probability is weakly increasing. (Note that this condition is merely a special case of the rightmost inequality described in (25).)

If $p(0) \le \frac{1-\beta_2 f(y)}{1+r(1-\beta_2 t(y))-\beta_2 f(y)}$, $G_2 \in E_2$, and otherwise $G_2 \in E_1$. By Lemma 3.4, if $G_2 \in E_2$, then either $G_1 \in E_1$ or $G_1 \in E_2$. In either case, by (33), $P(G_1) \le P(G_2)$. Otherwise, $G_2 \in E_1$, so $G_1 \in E_1$ as well (again by Lemma 3.4). Clearly, $p(0) \le p(0)$, so in any case, we have that $P(G_1) \le P(G_2)$, the desired result.

Now, consider sufficiently large values of β, i.e. assume

$$\frac{1-\beta_1 f(y)}{1+r(1-\beta_1 t(y))-\beta_1 f(y)} > p\left(\left(1-\beta_1\frac{1-\beta_1 f(y)}{1+r(1-\beta_1 t(y))-\beta_1 f(y)}(t(y)-f(y))-\beta_1 f(y)\right)y\right).$$

We shall show that accident probability is weakly decreasing. (This condition is simply a special case of the leftmost inequality in (26).)

Note that $G_1 \notin E_1$ and $G_1 \notin E_2$, so $G_1 \in \bigcup_{i=3}^{7} E_i$ by Theorem 3.1. Very similar techniques to the above suffice in every case except when $G_1, G_2 \in E_3$ or $G_1, G_2 \in E_5$.

If $G_1 \in E_3$ and $G_2 \in E_3$, then Lemma 4.2 guarantees that $(0, (y, 0))$ is an equilibrium for both games. Then, by (2),

$$P(x) = p((1-Q(x))y) = p((1-\beta P(x)(t(y)-f(y))-\beta f(y))y).$$

Let P_1 and P_2 be the quantities that satisfy $P_1 = p((1-\beta P_1(t(y)-f(y))-\beta f(y))y)$ and $P_2 = p((1-\beta P_2(t(y)-f(y))-\beta f(y))y)$, respectively. By (2), $P(G_1) = P_1$ and $P(G_2) = P_2$. Assume by way of contradiction that $P_1 \le P_2$. We use algebraic manipulations to work "up" one level of recursion, starting with the definition of β_1 and β_2. This gives that

$$(1-\beta_1(P_1(t(y)-f(y))+f(y)))y > (1-\beta_2(P_2(t(y)-f(y))+f(y)))y.$$

Since $p(d)$ is increasing, it preserves the inequality, so

$$p((1-\beta_1 P_1(t(y)-f(y))-\beta_1 f(y)))y) > p((1-\beta_2 P_2(t(y)-f(y))-\beta_2 f(y)))y).$$

But then by definition of P_1 and P_2, we can substitute to obtain $P_1 > P_2$, contradicting our hypothesis. Therefore, we must have that $P(G_1) = P_1 > P_2 = P(G_2)$, the desired conclusion. If $G_1 \in E_5$ and $G_2 \in E_5$, a very similar technique can be used. Thus, P is decreasing with β. ■

This result immediately gives a minimizing signal display probability in each range. We use this result to prove Theorem 3.3.

Proof of Theorem 3.3: Immediately from Lemma 3.5, we have that either the smallest or largest value of β must minimize P. Therefore, either $0 \in \arg\min_{\beta\in[0,1]} P(G)$, or $1 \in \arg\min_{\beta\in[0,1]} P(G)$, as desired.

It remains to show that there actually exist signaling games such that $\beta_P = 1$ does not minimize accident probability. To that end, let $p(d) = 0.8d + 0.1$, $t(y) = 0.8y$, $f(y) = 0.1y$, $y = 0.9$, and $r = 20$. Then $G_0 = (0, y, r) \in E_1$, and $G_1 = (1, y, r) \in E_2$. Therefore, by (33), $P(G_0) = p(0) = 0.1$, and $P(G_1) = \frac{1-(1)f(y)}{1+r(1-(1)t(y))-(1)f(y)} \approx 0.1398$. Since $P(G_0) < P(G_1)$, $\beta_P = 1$ cannot be a solution to (15). ■

C. Sensitivity Analysis on Accident Cost

Theorem 3.3 guarantees the existence of signaling equilibria where no signaling is optimal; however, it does not claim that these equilibria occur for realistic parameter values nor does it consider the sensitivity of these equilibria to those values. To investigate this further, we now conduct a rudimentary sensitivity analysis of these equilibria.

To that end, fix y and r and define the perversity index

$$\delta := \frac{P((1, y, r))}{P((0, y, r))}. \tag{34}$$

Intuitively, this makes a good measure of the paradox severity; values of $\delta > 1$ mean the paradox occurs, and large values of δ mean that many extra accidents are caused by naively always signaling.

Figure 4 shows plots of δ for various model parameters. From these plots, we see that there are regions where δ is rather sensitive to r, particularly when r is small. However, after r reaches a large enough value, δ appears to change only gradually with r. Further, we note that it is possible for very large values of r to cause $\delta > 1$, indicating that our paradox occurs even for large accident costs. We take these facts as an indication of the robustness of the signaling paradox.

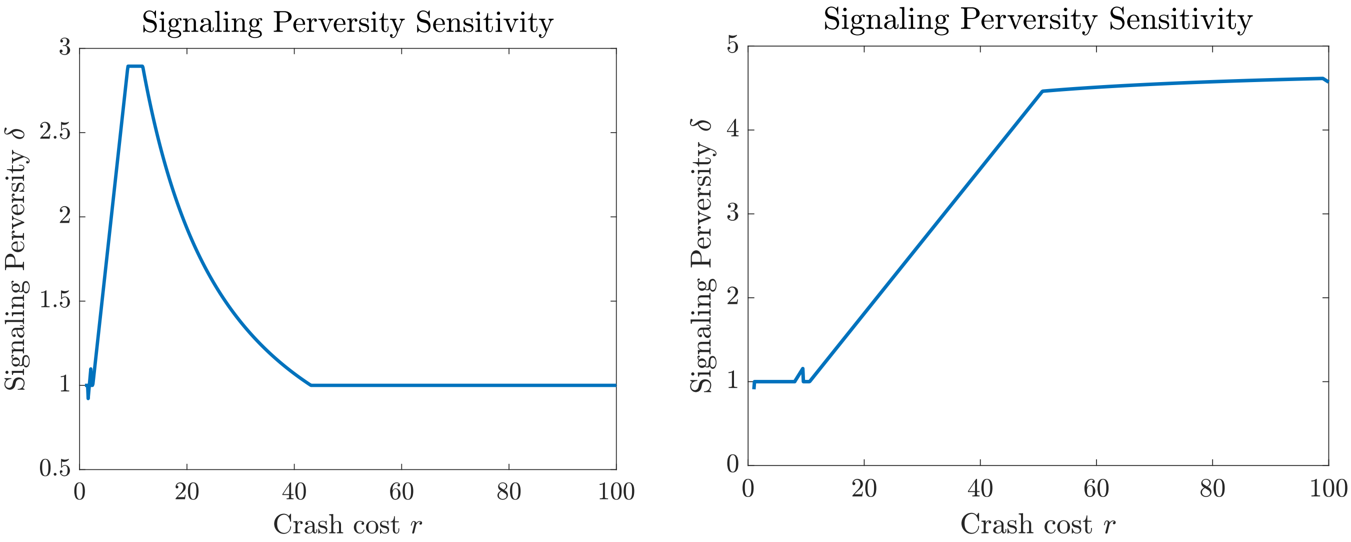
D. Information Design for Minimizing Social Cost

It is also useful to consider how to minimize social cost at equilibrium. Again, intuition suggests that the social cost minimizing value of β should be 1, but this is not always the case. We present the counter-intuitive result that there exist games where full information sharing among V2V drivers does not optimize social cost.

We provide examples to illustrate that β_S need not be 1.

Example 3.6: Let $p(d) = d^{\frac{1}{4}}$, $t(y) = 0.9y$, $f(y) = 0.1y$, $y = 0.066$, and $r = 1.001$. Additionally, let $\beta_1 = 0.4204$, $\beta_2 = 1$, $G_1 = (\beta_1, y, r)$, and $G_2 = (\beta_2, y, r)$. Then, $G_1 \in E_3$ and $G_2 \in E_3$. Numerical solvers give that $S(G_1) \approx 0.4949$, while $S(G_2) \approx 0.4960$. Thus, $S(G_2) > S(G_1)$. This parameter set is visualized in Figure 5a.

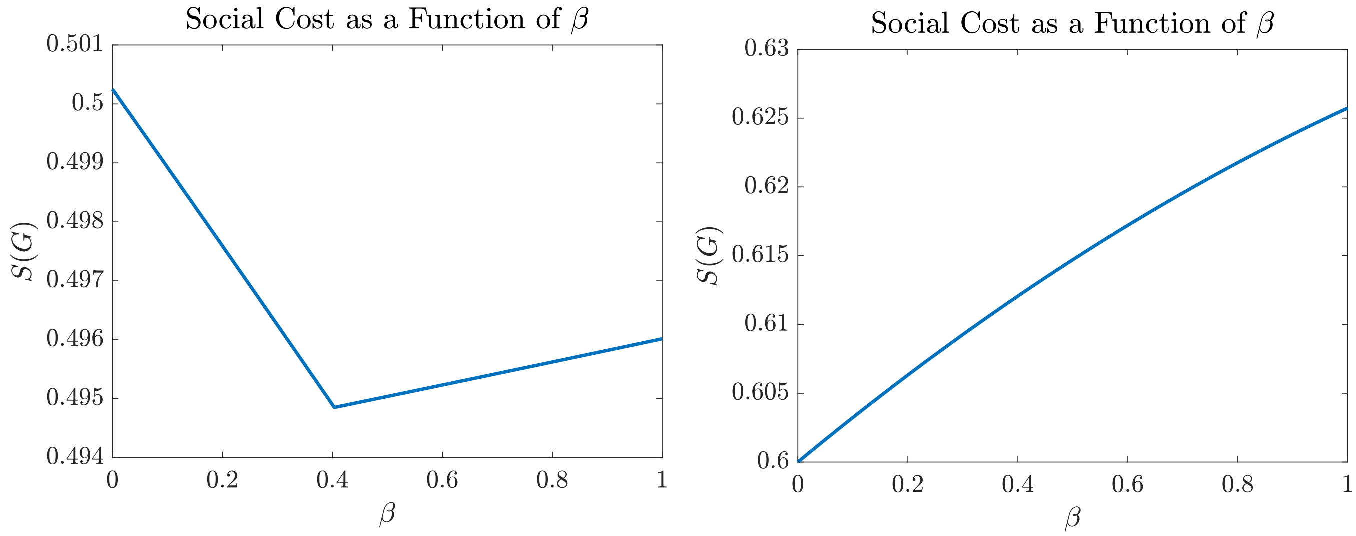
Example 3.7: Let $p(d) = 0.03d$, $t(y) = 0.95y$, $f(y) = 0.5y$, $y = 0.4$, and $r = 20$. Additionally, let $\beta_1 = 0$, $\beta_2 = 1$, $G_1 = (\beta_1, y, r)$, and $G_2 = (\beta_2, y, r)$. Then, $G_1 \in E_5$ and $G_2 \in E_5$. Numerical solvers give that $S(G_1) \approx 0.6$, while $S(G_2) \approx 0.6257$. Thus, $S(G_2) > S(G_1)$. This parameter set is visualized in Figure 5b.



(a) Signaling perversity under high accident probability. Accident probability characterized by $p(d) = 0.3d+0.1$, signal probability characterized by $t(y) = 0.9y$ and $f(y) = 0.1y$, and V2V penetration $y = 0.9$.

(b) Signaling perversity under low accident probability. Accident probability characterized by $p(d) = 0.1d + 0.01$, signal probability characterized by $t(y) = 0.9y$ and $f(y) = 0.1y$, and V2V penetration $y = 0.9$.

Fig. 4: Signaling perversity δ as a function of r. The general smoothness of the curves indicates a robustness of the signaling paradox with respect to accident probability.



(a) Social cost in E_3. Accident probability characterized by $p(d) = d^{\frac{1}{4}}$, signal probability characterized by $t(y) = 0.9y$ and $f(y) = 0.1y$, V2V penetration $y = 0.066$, and accident cost $r = 1.001$.

(b) Social cost in E_5. Accident probability characterized by $p(d) = 0.03d$, signal probability characterized by $t(y) = 0.95y$ and $f(y) = 0.5y$, V2V penetration $y = 0.4$, and accident cost $r = 20$.

Fig. 5: Equilibrium social cost with respect to the signal display probability β. Note the paradoxical result that there exist parameters where displaying more warning signals increases the social cost at equilibrium.

In general, $S(G)$ is decreasing with respect to β, with two exceptions. Within E_3 and E_5, social cost may sometimes be increasing.

Proposition 3.8: For any signaling game $G = (\beta, y, r)$, $S(G)$ is decreasing with respect to β unless $G \in E_3$ or $G \in E_5$. There exist games in both E_3 and E_5 where $S(G)$ is increasing with respect to β.

Proof: First, consider the case where $G \in E_2$. By Lemma 4.1, $P(G) = P_{\mathrm{vu}}$. Furthermore, by Lemma 4.2, $x^{\mathrm{ne}} = \left(0, \left(\frac{p^{-1}(P_{\mathrm{vu}})}{1-Q}, 0\right)\right)$ is the essentially unique signaling equilibrium for G. Then, (12) simplifies to

$$S(G) = r\frac{1 - \beta t(y)}{1 + r(1 - \beta t(y)) - \beta f(y)},$$

which is decreasing in β. (This can be seen by simply taking the partial derivative with respect to β, and noting that it is negative.) A similar technique suffices for the same result in the cases where $G \in E_1$, $G \in E_4$, $G \in E_6$, or $G \in E_7$.

Now, consider the case where $G \in E_3$ or $G \in E_5$. Example 3.6 describes a game in E_3 where $S(G)$ is increasing with β. Similarly, Example 3.7 describes a game in E_5 where $S(G)$ is increasing with β. Thus, social cost need not be decreasing with β in these ranges. ■

Note that if $G \in E_3$, $S(G)$ can be increasing with respect to β, but $P(G)$ is guaranteed to be decreasing by Lemma 3.5. Additionally, if $G \in E_2$, then $P(G)$ is increasing and $S(G)$ is decreasing. This implies that V2V administrators face an inherent trade-off in their optimization decision. To minimize accident probability, they must sometimes accept a higher than optimal social cost, and vice versa. This conflict is depicted in Figure 6.

IV. Proofs of Theorem 3.1

Proof of Lemma 3.2: Assume that a behavior tuple $x = (x_{\mathrm{n}}, (x_{\mathrm{vu}}, x_{\mathrm{vs}}))$ satisfies equations (17)-(19). We shall show that x is a signaling equilibrium.

First, assume that $P(x) > \frac{1}{1+r}$. By basic algebra, $1 - P < rP$, or equivalently by (17), $J_{\mathrm{n}}(\mathrm{C}; x) < J_{\mathrm{n}}(\mathrm{R}; x)$. Then, simply note that $x_{\mathrm{n}} = 0$ satisfies (6) and (7). Similarly, if $P = \frac{1}{1+r}$ we have that $J_{\mathrm{n}}(\mathrm{C}; x) = J_{\mathrm{n}}(\mathrm{R}; x)$, and if $P < \frac{1}{1+r}$, then $J_{\mathrm{n}}(\mathrm{C}; x) > J_{\mathrm{n}}(\mathrm{R}; x)$. In any case, (17) forces x_{n} to satisfy (6) and (7). An identical method shows that the conditions enforced by (18) satisfy (8) and (9), and that (19) satisfies (10) and (11). Since (6)-(11) are all satisfied, x must be a signaling equilibrium.

We will now show that any signaling equilibrium $x' = (x'_{\mathrm{n}}, (x'_{\mathrm{vu}}, x'_{\mathrm{vs}}))$ is essentially identical to x. Consider (17) in the case where $P(x') > \frac{1}{1+r}$. Assume by way of contradiction that $x'_{\mathrm{n}} \neq x_{\mathrm{n}}$, which immediately forces that $x'_{\mathrm{n}} > 0$. Since x' is a signaling equilibrium, by (7), we have that

$$J_{\mathrm{n}}(\mathrm{R}; x') \leq J_{\mathrm{n}}(\mathrm{C}; x').$$

But we showed above that because $P(x') > \frac{1}{1+r}$,

$$J_{\mathrm{n}}(\mathrm{C}; x') < J_{\mathrm{n}}(\mathrm{R}; x'),$$

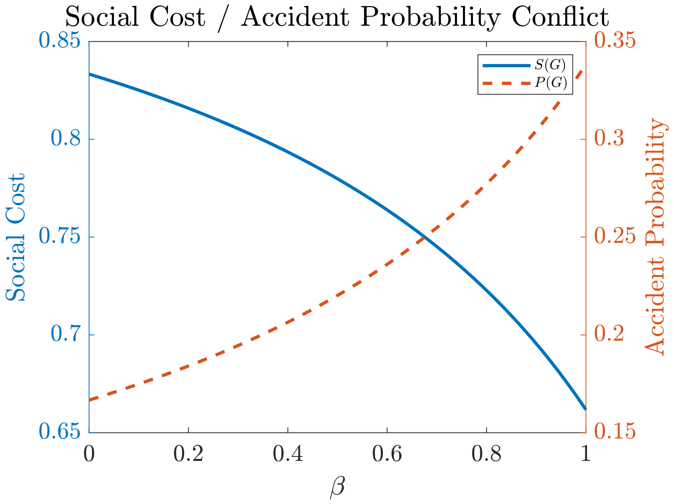


Fig. 6: Equilibrium accident probability and social cost with respect to the signal display probability β. Note that the social cost minimizing value of β is 1, while the accident probability minimizing value of β is 0. The example depicted has accident probability characterized by $p(d) = 0.8d + 0.1$, signal probability characterized by $t(y) = 0.9y$ and $f(y) = 0.1y$, V2V penetration $y = 0.8$, and accident cost $r = 5$.

a clear contradiction. Therefore, $x'_{\text{n}} = x_{\text{n}}$. A very similar technique using (6) shows that if $P(x') < \frac{1}{1+r}$, then $x'_{\text{n}} = 1 - y = x_{\text{n}}$. This technique can be further reused together with (8)-(11) to show that $x'_{\text{vu}} = x_{\text{vu}}$ in the first and third cases of (18) and $x'_{\text{vs}} = x_{\text{vs}}$ in the first and third cases of (19). In any of these cases, we then have that $x' = x$, so clearly x' is essentially identical to x.

It remains to show that x' is essentially identical to x in the second cases of (17)-(19). If $\mathbb{P}(\text{A}|\text{S}) = \frac{1}{1+r}$, then (21) and (23) imply that $P(x') = P_{\text{vs}} < P_{\text{n}} \leq P_{\text{vu}}$. But then by (20), $P(x') < \frac{1}{1+r}$ and $\mathbb{P}(\text{A}|\neg\text{S}) < \frac{1}{1+r}$. Therefore, by the above, $x'_{\text{n}} = 1 - y$ and $x'_{\text{vu}} = y$, so by (2),

$$P(x') = p(1 - y + y - Qy + Qx_{\text{vs}}) = p(1 - Q(y - x_{\text{vs}})).$$

But then simple algebra gives that $x'_{\text{vs}} = \frac{p^{-1}(P(x'))-1+Q(x')y}{Q(x')}$, which proves (19).

If $\beta t(y) > 0$, then the inequality described in (23) becomes strict, and a very similar technique suffices to show (17) and (18). Otherwise, we must have that $\beta t(y) = 0$. In the second case of either (17) or (18), (2) simplifies to

$$P(x') = p(x'_{\text{n}} + x'_{\text{vu}}) = \frac{1}{1+r},$$

which implies that $p^{-1}\left(\frac{1}{1+r}\right) = x'_{\text{n}} + (1-Q)x'_{\text{vu}} + Qx'_{\text{vs}}$. Since the same must be true for x, (13) and (14) are satisfied, so the equilibria are essentially identical. Therefore, the tuple satisfying (17)-(19) is a signaling equilibrium, and is essentially unique. ■

Recall that (24)-(30) divide our parameter space into seven regions. Lemma 4.1 shows that each of these regions restricts the possible values of P.

Lemma 4.1: For any signaling game $G = (\beta, y, r)$,

$$G \in \bigcup_{i=1}^{7} E_i, \tag{35}$$

and each of these ranges restricts the possible values of $P(G)$ according to the following table.

$G \in E_1$	$G \in E_2$	$G \in E_3$	$G \in E_4$	$G \in E_5$	$G \in E_6$	$G \in E_7$
$P(G) = p(0)$	$P(G) = P_{\text{vu}}$	$P_{\text{n}} < P(G) < P_{\text{vu}}$	$P(G) = P_{\text{n}}$	$P_{\text{vs}} < P(G) < P_{\text{n}}$	$P(G) = P_{\text{vs}}$	$P(G) = p(1)$

Each of these claims is proved via contradiction. Applying Lemma 3.2 to the contradiction hypothesis places restrictions on the values of x_{n}^{ne}, $x_{\text{vu}}^{\text{ne}}$, and $x_{\text{vs}}^{\text{ne}}$. Next, using these values and (2), we perform algebraic operations to take P "up" one level of its recursive definition. Finally, we show that this new expression for P forces a contradiction.

Proof: Consider any game $G = (\beta, y, r)$. If $G \notin E_1$ and $G \notin E_2$, then we must have that

$$p((1 - \beta P_{\text{vu}}(t(y) - f(y)) - \beta f(y))y) < P_{\text{vu}}.$$

Similarly, if $G \notin E_7$ and $G \notin E_6$, then we must have that

$$p((1-\beta P_{\mathrm{vu}}(t(y)-f(y))-\beta f(y))y) < P_{\mathrm{vu}}.$$

Therefore, if $G \notin E_3$ and $G \notin E_5$,

$$p((1-\beta P_{\mathrm{n}}(t(y)-f(y))-\beta f(y))y) \leq P_{\mathrm{n}} \leq p(1-(\beta P_{\mathrm{n}}(t(y)-f(y))+\beta f(y))y).$$

But this implies that $G \in E_4$. Therefore, $G \in \bigcup_{i=1}^{7} E_i$, as desired.

Since the remaining technique is sufficient for all seven claims, we will prove only the case where $G \in E_3$, i.e.

$$p((1-\beta P_{\mathrm{vu}}(t(y)-f(y))-\beta f(y))y) < P_{\mathrm{vu}} \text{ and } P_{\mathrm{n}} < p((1-\beta P_{\mathrm{n}}(t(y)-f(y))-\beta f(y))y).$$

Note that if $\beta t(y) = 0$, $P_{\mathrm{n}} = P_{\mathrm{vu}}$, and the above conditions simplify to $p(y) < P_{\mathrm{vu}}$ and $P_{\mathrm{n}} < p(y)$, respectively. But this is clearly a contradiction, since it implies that $P_{\mathrm{n}} < p(y) < P_{\mathrm{vu}} = P_{\mathrm{n}}$. Therefore, $\beta t(y) > 0$, and both inequalities described in (23) are strict (i.e. $P_{\mathrm{n}} < P_{\mathrm{vu}}$).

If we assume by way of contradiction that $P(x^{\mathrm{ne}}) \leq \frac{1}{1+r}$, by (20) and because $P_{\mathrm{n}} < P_{\mathrm{vu}}$, $\mathbb{P}(\mathrm{A}|\neg\mathrm{S}) < \frac{1}{1+r}$. Therefore, by Lemma 3.2, $x^{\mathrm{ne}}_{\mathrm{vu}} = y$. Furthermore, it is always true that $x^{\mathrm{ne}}_{\mathrm{n}} \geq 0$ and $x^{\mathrm{ne}}_{\mathrm{vs}} \geq 0$. Then, starting with our contradiction hypothesis, we perform algebraic operations to take P "up" one level of its recursive definition in (2). This gives that

$$x^{\mathrm{ne}}_{\mathrm{n}} + x^{\mathrm{ne}}_{\mathrm{vu}} - \left(\beta P(x^{\mathrm{ne}})(t(y)-f(y)) + \beta f(y)\right)\left(x^{\mathrm{ne}}_{\mathrm{vu}} - x^{\mathrm{ne}}_{\mathrm{vs}}\right) \geq 0 + y - \left(\frac{\beta(t(y)-f(y))}{1+r} + \beta f(y)\right)(y-0).$$

Since $p(d)$ is strictly increasing, it preserves the inequality, giving

$$p(x^{\mathrm{ne}}_{\mathrm{n}} + x^{\mathrm{ne}}_{\mathrm{vu}} - \left(\beta P(x^{\mathrm{ne}})(t(y)-f(y)) + \beta f(y)\right)\left(x^{\mathrm{ne}}_{\mathrm{vu}} - x^{\mathrm{ne}}_{\mathrm{vs}}\right)) = P(x^{\mathrm{ne}}) \geq p((1-\beta P_{\mathrm{n}}(t(y)-f(y))-\beta f(y))y).$$

Therefore, applying (26), we have that

$$p((1-\beta P_{\mathrm{n}}(t(y)-f(y))-\beta f(y))y) \leq P(x^{\mathrm{ne}}) \leq P_{\mathrm{n}} < p((1-\beta P_{\mathrm{n}}(t(y)-f(y))-\beta f(y))y),$$

an obvious contradiction. Therefore, we must have that $\frac{1}{1+r} = P_{\mathrm{n}} < P(x^{\mathrm{ne}})$. This technique can also be used to show that $P(x^{\mathrm{ne}}) < P_{\mathrm{vu}} = \frac{1-\beta f(y)}{1+r(1-\beta t(y))-\beta f(y)}$, completing the proof in this case.

A proof of the remaining claims can be accomplished in a similar manner. ■

From Lemma 4.1 we now know the possible equilibrium accident probabilities in any region of parameter space. Using these values and Lemma 3.2, we can derive what a signaling equilibrium in each range must look like.

Lemma 4.2: For any signaling game $G = (\beta, y, r)$, a unique signaling equilibrium $x^{\mathrm{ne}}_{\mathrm{n}}$ exists and takes the following form:

$$G \in E_1 \implies x^{\mathrm{ne}} = (0,(0,0)), \tag{36}$$

$$G \in E_2 \implies x^{\mathrm{ne}} = \left(0, \left(\frac{p^{-1}(P_{\mathrm{vu}})}{1-Q}, 0\right)\right), \tag{37}$$

$$G \in E_3 \implies x^{\mathrm{ne}} = (0,(y,0)), \tag{38}$$

$$G \in E_4 \implies x^{\mathrm{ne}} = \left(p^{-1}(P_{\mathrm{n}})) - (1-Q)y, (y,0)\right), \tag{39}$$

$$G \in E_5 \implies x^{\mathrm{ne}} = (1-y,(y,0)), \tag{40}$$

$$G \in E_6 \implies x^{\mathrm{ne}} = \left(1-y, \left(y, \frac{p^{-1}(P_{\mathrm{vs}}) - 1 + Qy}{Q}\right)\right), \tag{41}$$

$$G \in E_7 \implies x^{\mathrm{ne}} = (1-y,(y,y)). \tag{42}$$

For each of the five claims, we reuse the following proof method:

1) For each region, apply Lemma 4.1 to obtain a condition on $P(x)$
2) Use (20), (21), and (23), and the condition on $P(x)$ as needed to derive similar conditions on $\mathbb{P}(\mathrm{A}|\neg\mathrm{S})$ and $\mathbb{P}(\mathrm{A}|\mathrm{S})$
3) Apply Lemma 3.2 using these conditions to derive which type of equilibrium exists in that region

Proof: We prove this in cases. First, assume that $G \in E_3$, i.e.

$$p((1-\beta P_{\mathrm{vu}}(t(y)-f(y))-\beta f(y))y) < P_{\mathrm{vu}} \text{ and } P_{\mathrm{n}} < p((1-\beta P_{\mathrm{n}}(t(y)-f(y))-\beta f(y))y).$$

By Lemma 4.1, $P_{\mathrm{n}} < P < P_{\mathrm{vu}}$. Then, by (20), (21), and (23), $\mathbb{P}(\mathrm{A}|\neg\mathrm{S}) < \frac{1}{1+r}$ and $\mathbb{P}(\mathrm{A}|\mathrm{S}) > \frac{1}{1+r}$. Finally, by Lemma 3.2, $(0,(y,0))$ is a signaling equilibrium and essentially unique.

An identical method can be used to show that $(0,(0,0))$, $(1-y,(y,0))$, and $(1-y,(y,y))$ are essentially unique signaling equilibria if $G \in E_1$, $G \in E_5$, or $G \in E_7$, respectively.

Now, assume that $G \in E_2$. By Lemma 4.1, $P(x^{\mathrm{ne}}) = P_{\mathrm{vu}}$, and (20) implies that $\mathbb{P}(\mathrm{A}|\neg\mathrm{S}) = \frac{1}{1+r}$

If $\beta t(y) > 0$, then $P_{\mathrm{n}} < P_{\mathrm{vu}}$, so by (20), (21), and (23), $P(x^{\mathrm{ne}}) > \frac{1}{1+r}$ and $\mathbb{P}(\mathrm{A}|\mathrm{S}) > \frac{1}{1+r}$. By Lemma 3.2,

$$\left(0, \left(\frac{p^{-1}(P_{\mathrm{vu}})}{1-Q}, 0\right)\right)$$

is then an essentially unique signaling equilibrium.

Otherwise, $\beta t(y) = 0$, so $Q(x^{\mathrm{ne}}) = 0$. Note that by (23), $P(x^{\mathrm{ne}}) = P_{\mathrm{vu}} = P_{\mathrm{n}} > P_{\mathrm{vs}}$. By (20), $\mathbb{P}(\mathrm{A}|\neg\mathrm{S}) = \frac{1}{1+r}$, and by (21), $\mathbb{P}(\mathrm{A}|\mathrm{S}) > \frac{1}{1+r}$ Therefore by Lemma 3.2, $x_{\mathrm{n}}^{\mathrm{ne}} = p^{-1}(P(x^{\mathrm{ne}})) - (1 - Q(x^{\mathrm{ne}}))y = p^{-1}(\frac{1}{1+r}) - y$, $x_{\mathrm{vu}}^{\mathrm{ne}} = \frac{p^{-1}(P(x^{\mathrm{ne}}))}{1-Q(x^{\mathrm{ne}})} = p^{-1}(\frac{1}{1+r})$, and $x_{\mathrm{vs}}^{\mathrm{ne}} = 0$. By (2), this gives that

$$\frac{1}{1+r} = p\left(p^{-1}\left(\frac{1}{1+r}\right) - y + p^{-1}\left(\frac{1}{1+r}\right)\right),$$

forcing $p^{-1}(\frac{1}{1+r}) = y$. Therefore, by substitution, $(0, (y, 0))$ must be an essentially unique signaling equilibrium (note that this is a special case of the more general form given above). A similar technique can be used to show that signaling equilibria of the forms claimed are forced when $G \in E_4$ and $G \in E_6$.

By Lemma 4.1, any game G must satisfy at least one of the above conditions, and therefore has an essentially unique signaling equilibrium. ■

Finally, we are equipped to prove Theorem 3.1.

Proof of Theorem 3.1: Lemma 4.2 demonstrated existence and essential uniqueness of a signaling equilibrium for all signaling games G. Note that each of these equilibria are consistent with the forms claimed, completing the proof. ■

V. Conclusion

This paper has posed and analyzed a simple model of self-interested driver behavior in the presence of road hazard signals. Our characterization result in Theorem 3.1 describes the necessary qualities of any equilibrium, namely that unsignaled V2V drivers are more reckless and signaled V2V drivers are more cautious, relative to those without V2V technology. In Theorem 3.3, we describe how careful information design of β, the disclosure rate of warning signals, can reduce accident probability at equilibrium compared to the naive approach. In particular, our main result is that warning a subset of drivers more often about traffic accidents can paradoxically lead to an increased probability of accidents occurring, relative to leaving all drivers uninformed. Finally, we close with a similar paradox in terms of social cost in Proposition 3.8.

Our work is by no means exhaustive, and there is still much room for future research in this area. One main limitation is that we do not include network routing effects in our model. Prior research has shown that the structure of a network can have complex and paradoxical influences on the equilibrium behavior [19], [20]. Since our model considers only a single road, future results could achieve a result more directly applicable to the real world by integrating network effects.

Furthermore, we use a binary signaling policy, which is rather simplistic. V2V technology could potentially communicate more information about road hazards, such as the number of cars detecting the hazard or reported severity. Increased detail in the signal would allow for both a larger degree of control in information design and a more nuanced model of the driver decision making process.

Other possibilities for future work include using a more realistic model of human behavior. One way to do this would be to relax the assumption that human drivers behave as *homo economicus* and perform Bayesian inference to decide their behavior while driving. Alternative approaches could consider an environment with competing information providers, or explore "trust" and how drivers interact with a system they know could be withholding information from them.

Appendix

A. *Proof of Proposition 2.1*

Equation (2) defines equilibrium crash probability $P(x)$ through a rather complicated recursive relationship. However, we show that this relationship must always have a solution.

Proof of Proposition 2.1

Note that by rearranging the right side of (2), we have that

$$P(x) = p(x_{\mathrm{n}} + x_{\mathrm{vu}} - \beta(P(x)(t(y) - f(y)) + f(y))(x_{\mathrm{vu}} - x_{\mathrm{vs}})).$$

Note that $P(x)$ can take on any value in the range $[p(0), p(1)]$. Therefore, consider the function $g(c) : [p(0), p(1)] \to [p(0), p(1)]$ by $g(c) := c$. Clearly, $g(c)$ is continuous. Next, consider the function $h(c) : [p(0), p(1)] \to [p(0), p(1)]$ by $h(c) := p(x_{\mathrm{n}} + x_{\mathrm{vu}} - \beta(c(t(y) - f(y)) + f(y))(x_{\mathrm{vu}} - x_{\mathrm{vs}}))$. Since compositions of continuous functions are continuous, and $p(d)$ is continuous, $h(c)$ must also be. But then $g(c)$ and $h(c)$ are continuous functions bounded within the same range, so there must exist at least one $\bar{c} \in [p(0), p(1)]$ such that $g(\bar{c}) = h(\bar{c})$. That is, (2) must have at least one solution for $P(x)$ for all parameter combinations. ■

B. Counter-Examples for when $f(y) \equiv 0$

A potential critique of the paradoxes prevented in our paper is that false positive signals excessively deceive drivers, causing the strange equilibrium behavior. However, this is not the case. Even without any false positive signals, our two main paradoxes are still possible.

Example 1.1: Accident probability can be increasing with β even if $f(y) \equiv 0$. Let $p(d) = 0.3d + 0.1$, $t(y) = 0.9y$, $f(y) \equiv 0$, $y = 0.9$, and $r = 3$. Additionally, let $\beta_1 = 0$, $\beta_2 = 0.4004$, $G_1 = (\beta_1, y, r)$, and $G_2 = (\beta_2, y, r)$. Then, $G_1 \in E_2$ and $G_2 \in E_2$. Therefore, $P(G_1) \approx 0.25$ and $P(G_2) \approx 0.3304$ Thus, increasing the quality of V2V information can increase the accident probability at equilibrium.

Example 1.2: Social cost can be increasing with β even if $f(y) \equiv 0$. Let $p(d) = d^{\frac{1}{4}}$, $t(y) = 0.9y$, $f(y) \equiv 0$, $y = 0.07$, and $r = 1.001$. Additionally, let $\beta_1 = 0.9$, $\beta_2 = 1$, $G_1 = (\beta_1, y, r)$, and $G_2 = (\beta_2, y, r)$. Then, $G_1 \in E_3$ and $G_2 \in E_3$. Numerical solvers give that $S(G_1) \approx 0.4889$, while $S(G_2) \approx 0.4890$. Thus, increasing the quality of V2V information can increase the expected cost to drivers.

Recall that in general, it is possible for social cost to be increasing with β when $G \in E_5$ (see Example 3.7). Interestingly, this is no longer possible if $f(y) \equiv 0$.

C. Code Availability

All code used for this project is available at `https://github.com/descon-uccs/gould-trptc-2022`.

References

[1] B. Gould and P. Brown, "On Partial Adoption of Vehicle-to-Vehicle Communication: When Should Cars Warn Each Other of Hazards?," in *2022 American Control Conference*, pp. 627–632, 2022.

[2] P. N. Brown, "When Altruism is Worse than Anarchy in Nonatomic Congestion Games," in *2021 American Control Conference (ACC)*, pp. 4503–4508, IEEE, May 2021.

[3] M. Gairing, B. Monien, and K. Tiemann, "Selfish Routing with Incomplete Information," *Theory of Computing Systems*, vol. 42, pp. 91–130, Jan. 2008.

[4] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "Selfish Routing in Capacitated Networks," *Mathematics of Operations Research*, vol. 29, pp. 961–976, Nov. 2004.

[5] S. Dafermos and A. Nagurney, "On some traffic equilibrium theory paradoxes," *Transportation Research Part B: Methodological*, vol. 18, pp. 101–110, Apr. 1984.

[6] J. G. Wardrop, "ROAD PAPER. SOME THEORETICAL ASPECTS OF ROAD TRAFFIC RESEARCH.," *Proceedings of the Institution of Civil Engineers*, vol. 1, pp. 325–362, May 1952.

[7] O. Massicot and C. Langbort, "Public Signals and Persuasion for Road Network Congestion Games under Vagaries," *IFAC-PapersOnLine*, vol. 51, pp. 124–130, Jan. 2019.

[8] M. Wu and S. Amin, "Information Design for Regulating Traffic Flows under Uncertain Network State," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing*, (Allerton), pp. 671–678, Sept. 2019.

[9] J. Nie, J. Zhang, W. Ding, X. Wan, X. Chen, and B. Ran, "Decentralized Cooperative Lane-Changing Decision-Making for Connected Autonomous Vehicles*," *IEEE Access*, vol. 4, pp. 9413–9420, 2016.

[10] B. L. Ferguson, P. N. Brown, and J. R. Marden, "The Effectiveness of Subsidies and Tolls in Congestion Games," *IEEE Transactions on Automatic Control*, pp. 2729–2742, Feb 2021.

[11] D. A. Lazar, E. Bıyık, D. Sadigh, and R. Pedarsani, "Learning how to dynamically route autonomous vehicles on shared roads," *Transportation Research Part C: Emerging Technologies*, vol. 130, pp. 1–16, Sept. 2021.

[12] E. Bıyık, D. A. Lazar, R. Pedarsani, and D. Sadigh, "Incentivizing Efficient Equilibria in Traffic Networks With Mixed Autonomy," *IEEE Transactions on Control of Network Systems*, vol. 8, pp. 1717–1729, Dec. 2021.

[13] Y. Zhu and K. Savla, "Information design in nonatomic routing games with partial participation: Computation and properties," *IEEE Transactions on Control of Network Systems*, vol. 9, no. 2, pp. 613–624, 2022.

[14] O. Massicot and C. Langbort, "Competitive comparisons of strategic information provision policies in network routing games," *IEEE Transactions on Control of Network Systems*, pp. 1–1, 2021.

[15] N. Mehr and R. Horowitz, "How Will the Presence of Autonomous Vehicles Affect the Equilibrium State of Traffic Networks?," *IEEE Transactions on Control of Network Systems*, vol. 7, pp. 96–105, Mar 2020.

[16] E. Kamenica and M. Gentzkow, "Bayesian Persuasion," *American Economic Review*, vol. 101, pp. 2590–2615, Oct. 2011.

[17] D. Bergemann and S. Morris, "Information Design: A Unified Perspective," *Journal of Economic Literature*, vol. 57, pp. 44–95, Mar. 2019.

[18] E. Akyol, C. Langbort, and T. Başar, "Information-Theoretic Approach to Strategic Communication as a Hierarchical Game," *Proceedings of the IEEE*, vol. 105, pp. 205–218, Feb. 2017.

[19] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Informational Braess' Paradox: The Effect of Information on Traffic Congestion," *Operations Research*, vol. 66, pp. 893–917, Aug. 2018.

[20] C. Roman and P. Turrini, "How does information affect asymmetric congestion games?," *arXiv:1902.07083 [cs, math]*, Feb. 2019.

[21] J. Liu, S. Amin, and G. Schwartz, "Effects of Information Heterogeneity in Bayesian Routing Games," *arXiv:1603.08853 [cs]*, Mar. 2016.

[22] M. O. Sayin, E. Akyol, and T. Başar, "Hierarchical multistage Gaussian signaling games in noncooperative communication and control systems," *Automatica*, vol. 107, pp. 9–20, Sept. 2019.

[23] H. Tavafoghi and D. Teneketzis, "Informational incentives for congestion games," in *55th Annual Allerton Conference on Communication, Control, and Computing*, (Allerton), p. 35, 2017.

[24] R. Balakrishna, M. Ben-Akiva, J. Bottom, and S. Gao, "Information Impacts on Traveler Behavior and Network Performance: State of Knowledge and Future Directions," in *Advances in Dynamic Network Modeling in Complex Transportation Systems* (S. V. Ukkusuri and K. Ozbay, eds.), vol. 2, pp. 193–224, New York, NY: Springer New York, 2013.

[25] M. Ben-Akiva, A. De Palma, and K. Isam, "Dynamic network models and driver information systems," *Transportation Research Part A: General*, vol. 25, pp. 251–266, Sept. 1991.