

On Partial Adoption of Vehicle-to-Vehicle Communication: When Should Cars Warn Each Other of Hazards?

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. Our model indicates that for a wide range of our 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 [1].

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 [2]–[7]. Many proposed solutions to this problem focus on the effects of deploying smart infrastructure to alleviate congestion and safety issues, using incentive design [8]–[10] and information design [11], [12] 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 [13].

We study a common paradox of information design. *Bayesian persuasion* describes the process of a sender disclosing or obfuscating information in an attempt to change the actions of other strategic agents [14], [15]. However, it is

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.

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, known as informational Braess’ paradox [16], [17]. In information design problems in general, full disclosure of information is not always optimal [6], [15], [18]–[21].

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 [4], [7], [9], [22].

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 3.2 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. We then close with a discussion of the relationship between the *social cost* and the frequency of accidents, and show that these two objectives are sometimes fundamentally opposed to one another: 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. Each driver can choose to drive carefully (C), or recklessly (R), and a traffic accident either occurs (A) or does not occur ($\neg A$). Reckless drivers become involved in existing accidents and experience an expected cost of $r > 1$; however, careful drivers regret their caution if an accident is not present and experience a regret cost of 1. These costs are collected in this matrix:

	Accident (A)	No Accident ($\neg A$)
Careful (C)	0	1
Reckless (R)	r	0

We write d to denote the total mass 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.

We model partial V2V penetration, i.e. some fraction $y \in [0, 1]$ of drivers have cars equipped with V2V technology. If these drivers encounter a road hazard, V2V technology will autonomously detect this hazard and broadcast a warning signal with probability $q(y)$. For simplicity, we assume that if a signal is broadcast, an accident must have occurred, and that the signal is received by every V2V car.

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 [16], [18]–[20]. 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, given that it has received a signal, and let $\mathbb{P}(S) = \beta \in [0, 1]$. Note that $\mathbb{P}(A|S) = 1$, since we have assumed there are no false positive warnings.

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 *unsignedaled V2V drivers* and *signaled V2V drivers*. We write x_n , x_{vu} , and x_{vs} to represent the mass of reckless drivers in each group, respectively, and a behavior profile as $x = (x_n, (x_{vu}, x_{vs}))$.

We model that both non-V2V and V2V drivers have habitual behaviors x_n^* and x_v^* , respectively, and that these behaviors determine the probability of an accident. Every day, non-V2V cars follow this habitual behavior, i.e. $x_n = x_n^*$. However, V2V drivers are able to adjust their behavior based on whether or not they see a warning on a particular day. The habitual behavior of V2V drivers must be a weighted average of their behavior when they do and do not see warnings, i.e.

$$x_v^* = \mathbb{P}(\neg S)x_{vu} + \mathbb{P}(S)x_{vs}. \quad (1)$$

In Lemma 4.1, we show that at equilibrium, $x_{vs} = 0$. Therefore, in this manuscript we simplify (1) to

$$x_v^* = \mathbb{P}(\neg S)x_{vu}. \quad (2)$$

Define $P(x)$ as the probability that an accident occurs, given some behavior profile x . Additionally define $Q(x)$ as the probability that a specific 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, $\mathbb{P}(A) = P(x) = p(x_n^* + x_v^*)$ and $\mathbb{P}(S) = Q(x) = P(x)\beta q(y)$. Applying substitution with (2) gives that

$$P(x) = p(x_n + (1 - P(x)\beta q(y))x_{vu}). \quad (3)$$

This uniquely specifies $P(x)$, since the left side is strictly increasing in P and sweeps from $p(0)$ to $p(1)$, while the right side is strictly decreasing in P and bounded within the same range. Note that $P(x)$ is implicitly parameterized by β and y .

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

$$J_n(a; x) = \begin{cases} 1 - P(x) & \text{if } a = C, \\ rP(x) & \text{if } a = R, \end{cases} \quad (4)$$

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

$$J_{vs}(a; x) = \begin{cases} 0 & \text{if } a = C, \\ r & \text{if } a = R. \end{cases} \quad (6)$$

Note that (6) holds due to our assumption that $\mathbb{P}(A|S) = 1$. Finally, define a signaling game as the tuple $G = (\beta, y, r)$.

B. Signaling Equilibrium

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

$$x_n^{\text{ne}} < 1 - y \implies J_n(C; x^{\text{ne}}) \leq J_n(R; x^{\text{ne}}), \quad (7)$$

$$x_n^{\text{ne}} > 0 \implies J_n(R; x^{\text{ne}}) \leq J_n(C; x^{\text{ne}}), \quad (8)$$

$$x_{vu}^{\text{ne}} < y \implies J_{vu}(C; x^{\text{ne}}) \leq J_{vu}(R; x^{\text{ne}}), \quad (9)$$

$$x_{vu}^{\text{ne}} > 0 \implies J_{vu}(R; x^{\text{ne}}) \leq J_{vu}(C; x^{\text{ne}}), \quad (10)$$

$$x_{vs}^{\text{ne}} < y \implies J_{vs}(C; x^{\text{ne}}) \leq J_{vs}(R; x^{\text{ne}}), \quad (11)$$

$$x_{vs}^{\text{ne}} > 0 \implies J_{vs}(R; x^{\text{ne}}) \leq J_{vs}(C; x^{\text{ne}}). \quad (12)$$

Equations (7)-(12) 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 unsignedaled V2V drivers, using driver behavior at equilibrium.

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

$$S(x) = J_n(C; x)(1 - y - x_n) + J_n(R; x)(x_n) + (1 - Q)(J_{vu}(C; x)(y - x_{vu}) + J_{vu}(R; x)(x_{vu})). \quad (13)$$

Note that the decision of signaled V2V drivers is not included in this equation, since they will never incur a cost.

Proposition 2.1: For every signaling game G , there exists a signaling equilibrium x^{ne} and it is essentially unique. By this we mean that for any two signaling equilibria x_1^{ne} and x_2^{ne} of G , both of the following hold:

$$x_{n1}^{\text{ne}} + (1 - Q(x_1^{\text{ne}}))x_{vu1}^{\text{ne}} = x_{n2}^{\text{ne}} + (1 - Q(x_2^{\text{ne}}))x_{vu2}^{\text{ne}}, \quad (14)$$

$$P(x_1^{\text{ne}}) = P(x_2^{\text{ne}}). \quad (15)$$

We provide a proof of this fact in Lemma 4.3. Throughout this paper, we use the terms “unique” and “essentially unique” interchangeably to mean (14) and (15) are satisfied.

C. Research Objectives

Our first objective is to characterize the signaling equilibrium 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 quality. To that end, we abuse notation and write $P(G)$ to denote $P(x^{\text{ne}})$ and $S(G)$ to denote $S(x^{\text{ne}})$ where x^{ne} is a signaling equilibrium of game G . We then wish to find values for β_P and β_S such that

$$\beta_P \in \arg \min_{\beta \in [0,1]} P(G), \quad (16)$$

$$\beta_S \in \arg \min_{\beta \in [0,1]} S(G). \quad (17)$$

In Theorem 3.2, we provide a constant time algorithm to determine a solution to (16), and show that there exist games where $\beta_P = 0$ is a solution, as depicted in Figure 1. Furthermore, in Proposition 3.5, we provide sufficient criteria for when $\beta_S = 1$ is a solution to (17), and show that there paradoxically exist regions of the parameter space where $\beta_S = 1$ is *not* a solution to (17).

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 (3). 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 $x^{\text{ne}} = (x_n^{\text{ne}}, (x_{\text{vu}}^{\text{ne}}, x_{\text{vs}}^{\text{ne}}))$ can take one of the following 3 forms:

- $(0, (0, 0))$,
- $(0, (\chi_{\text{vu}}, 0))$, for some $\chi_{\text{vu}} \in [0, y]$,
- $(\chi_n, (y, 0))$, for some $\chi_n \in [0, 1 - y]$.

This captures several important characteristics of signaling equilibria. Chief among these is the fact that a signaling equilibrium exists and is essentially unique for all games $G = (\beta, y, r)$. Additionally, note that signaled V2V drivers are certain that an accident has occurred, and will therefore never choose to be reckless at equilibrium. Finally, 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 proceeds as follows: Lemma 4.1 provides sufficient conditions for an essentially unique signaling equilibria. Next, we divide our parameter space into five regions that can be analyzed individually, and Lemma 4.2 derives the possible values of P in each region. Finally, Lemma 4.3 uses these results to show that exactly one signaling equilibrium exists for every parameter combination. We provide the full proofs for Theorem 3.1 in Section IV.

B. Information Design for Minimizing Accident Probability

Though it seems intuitively obvious that increasing the quality of information given to drivers would allow them to make more informed decisions and arrive at less costly

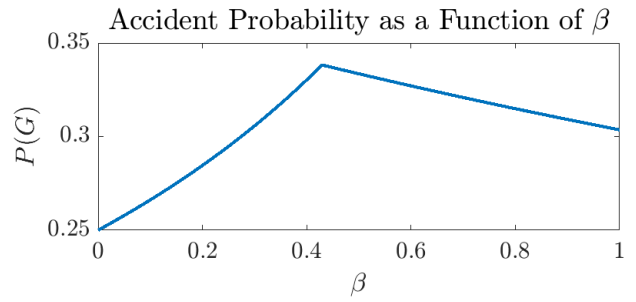


Fig. 1. Equilibrium accident frequency with respect to the signal quality β . Note that when $\beta < 0.43$, displaying more warning signals steeply increases the frequency of accidents, and that even the maximum possible signal quality $\beta = 1$ yields a higher accident frequency than if warning signals were never shown to drivers. The example depicted has accident probability characterized by $p(d) = 0.3d + 0.1$, signal probability characterized by $q(y) = 0.9y$, V2V penetration $y = 0.9$, and accident cost $r = 3$.

outcomes, prior research has shown that this is not always the case [16], [17]. 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 bound by technical limitations, but within that range, administrators can freely manipulate it by purposefully not displaying warnings to some drivers that receive signals. Paradoxically, we show that ignoring accidents in this way can decrease accident probability at equilibrium.

Theorem 3.2: For any signaling game $G = (\beta, y, r)$, we must have that either:

$$0 \in \arg \min_{\beta \in [0,1]} P(G) \text{ or } 1 \in \arg \min_{\beta \in [0,1]} P(G). \quad (18)$$

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 proceeds as follows: first, we 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, which is the desired result.

We will now provide a full definition of $P(G)$, the function optimized by (16). Note that in general, $P(G) = P(x^{\text{ne}})$, but with the following regions, we can be more specific:

$$p \left(1 - \frac{\beta q(y)}{1+r} y \right) < \frac{1}{1+r} \quad (19)$$

$$p \left(\left(1 - \frac{\beta q(y)}{1+r} \right) y \right) \leq \frac{1}{1+r} \leq p \left(1 - \frac{\beta q(y)}{1+r} y \right) \quad (20)$$

$$p(0) > \frac{1}{1+r(1-\beta q(y))} \quad (21)$$

$$p(0) \leq \frac{1}{1+r(1-\beta q(y))} \leq p \left(\left(1 - \frac{\beta q(y)}{1+r(1-\beta q(y))} \right) y \right) \quad (22)$$

$$p\left(\left(1 - \frac{\beta q(y)}{1+r(1-\beta q(y))}\right)y\right) < \frac{1}{1+r(1-\beta q(y))}$$

$$\text{and } \frac{1}{1+r} < p\left(\left(1 - \frac{\beta q(y)}{1+r}\right)y\right) \quad (23)$$

For any signaling game $G = (\beta, y, r)$, the probability of an accident at its signaling equilibrium is given by

$$P(G) = \begin{cases} p(0) & \text{if (21),} \\ \frac{1}{1+r(1-\beta q(y))} & \text{if (22),} \\ P(x^{\text{ne}}) & \text{if (23),} \\ \frac{1}{1+r} & \text{if (20),} \\ P(x^{\text{ne}}) & \text{if (19),} \end{cases} \quad (24)$$

where x^{ne} is the signaling equilibrium satisfying Lemma 4.3. A proof of this simplification can be found in Lemma 4.2. Since there exists a signaling equilibrium for any game G , this function is defined for all parameter combinations.

Lemma 3.3: For sufficiently low values of β , $P((\beta, y, r))$ is weakly increasing; otherwise it is weakly decreasing.

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 an exhaustive comparison of crash probabilities within and between the cases of (24). If

$$\frac{1}{1+r(1-\beta_2 q(y))} \leq p\left(\left(1 - \frac{\beta_2 q(y)}{1+r(1-\beta_2 q(y))}\right)y\right),$$

then G_1 and G_2 satisfy either (21) or (22). In any case, algebraic manipulations on (24) give that $P(G_1) < P(G_2)$.

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

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

Then, G_1 and G_2 each satisfy (23), (20), or (19). We claim that P is decreasing with respect to β . Simple algebra and Lemma 4.2 suffice to obtain this result in every case except the following. If G_1 and G_2 satisfy (19), then Lemma 4.3 guarantees that $(1-y, y, 0)$ is an equilibrium. Then, by (3),

$$P(x) = p(1-y + (1-P(x)\beta q(y))y) = p(1-P(x)\beta q(y)y).$$

Let P_1 and P_2 be the quantities that satisfy $P_1 = p(1 - P_1\beta_1 q(y)y)$ and $P_2 = p(1 - P_2\beta_2 q(y)y)$, respectively. By (3) and (24), $P(G_1) = P_1$ and $P(G_2) = P_2$. Assume by way of contradiction that $P_1 \leq P_2$. We use algebraic manipulations to work ‘‘up’’ one level of recursion, starting with the definition of β_1 and β_2 . This gives that

$$1 - P_1\beta_1 q(y)y > 1 - P_2\beta_2 q(y)y.$$

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

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

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 both games satisfy (23), a very similar technique can be used. Thus, P is decreasing with β . ■

This result immediately gives a minimizing signal quality in each range. We use this result to prove Theorem 3.2.

Proof of Theorem 3.2: Immediately from Lemma 3.3, 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 $p(d) = 0.8d + 0.1$, $q(y) = 0.9y$, $y = 0.7$, and $r = 20$. Then $G_0 = (0, y, r)$ satisfies (21), and $G_1 = (1, y, r)$ satisfies (22). Therefore, by (24), $P(G_0) = p(0) = 0.1$, and $P(G_1) = \frac{1}{1+r(1-(1)q(y))} \approx 0.1190$. Since $P(G_0) < P(G_1)$, $\beta_P = 1$ cannot be a solution. ■

C. 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. We present the counter-intuitive result that there exist games where full information sharing among V2V drivers does not optimize social cost.

We provide an example to illustrate that β_S need not be 1.

Example 3.4: Let $p(d) = d^{0.25}$, $q(y) = 0.9y$, $y = 0.07$, 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 and G_2 both satisfy (23). 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.

Proposition 3.5: For any signaling game $G = (\beta, y, r)$, $\beta_S = 1$ is a solution to (17) unless G satisfies (23).

Proof Sketch: For any of our equilibrium ranges, applying Lemmas 4.1 and 4.3 gives an algebraic expression for $S(G)$. In all ranges except (23), the expressions can be easily shown to be decreasing in β using Lemma 3.3. Thus, $\beta_S = 1$ is a solution to (17). ■

Note that if G satisfies (23), $S(G)$ can be increasing with respect to β , but $P(G)$ is guaranteed to be decreasing by Lemma 3.3. Additionally, if G satisfies (22), 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.

IV. PROOFS OF THEOREM 3.1

Lemma 4.1: For any signaling game $G = (\beta, y, r)$, a behavior profile $x^{\text{ne}} = (x_n^{\text{ne}}, (x_{\text{vu}}^{\text{ne}}, x_{\text{vs}}^{\text{ne}}))$ is a signaling equilibrium if $x_{\text{vs}}^{\text{ne}} = 0$, and the following hold:

$$x_n^{\text{ne}} = \begin{cases} 0 & \text{if } P(x^{\text{ne}}) > \frac{1}{1+r}, \\ p^{-1}(P(x^{\text{ne}})) - (1 - Q(x^{\text{ne}}))y & \text{if } P(x^{\text{ne}}) = \frac{1}{1+r}, \\ 1 - y & \text{if } P(x^{\text{ne}}) < \frac{1}{1+r}, \end{cases} \quad (25)$$

$$x_{\text{vu}}^{\text{ne}} = \begin{cases} 0 & \text{if } \mathbb{P}(A|\neg S) > \frac{1}{1+r}, \\ \frac{p^{-1}(P(x^{\text{ne}}))}{1-Q(x^{\text{ne}})} & \text{if } \mathbb{P}(A|\neg S) = \frac{1}{1+r}, \\ y & \text{if } \mathbb{P}(A|\neg S) < \frac{1}{1+r}. \end{cases} \quad (26)$$

If $\beta q(y) > 0$, then this is the only signaling equilibrium. Otherwise, all signaling equilibria that exist satisfy (14) and (15). That is, this equilibrium is essentially unique.

The proof of Lemma 4.1 is simple, but notationally intensive; for brevity, we provide a proof sketch here. A relationship which is useful in this section is:

$$\frac{1}{1+r} \leq \frac{1}{1+r(1-\beta q(y))}. \quad (27)$$

Additionally, it can be shown using Bayes' Theorem that,

$$\mathbb{P}(A|\neg S) \stackrel{\leq}{>} \frac{1}{1+r} \iff P(x) \stackrel{\leq}{>} \frac{1}{1+r(1-\beta q(y))}. \quad (28)$$

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.

Proof Sketch: In the forward direction, simple algebra largely suffices. For each case of (25) and (26), the necessary values of P and $\mathbb{P}(A|\neg S)$ can be plugged into (4) and (5) to obtain a relationship on the expected costs to each group to drive carefully or recklessly. This relationship will always satisfy (7)-(10) for the corresponding value of x_n^{ne} or x_{vu}^{ne} .

Additionally, we always have that $J_{vs}(C) < J_{vs}(R)$, meaning that $x_{vs}^{\text{ne}} = 0$ will always satisfy (11) and (12). Therefore, the tuple described is a signaling equilibrium.

The reverse direction is slightly more nuanced, but similar. The first and third cases of (25) and (26), as well as the condition that $x_{vs}^{\text{ne}} = 0$ are simple. The described values of P and $\mathbb{P}(A|\neg S)$ imply a relationship between the cost of each behavior to non-V2V and unsignaled V2V drivers, respectively. After applying (7)-(10), this relationship contradicts all values of x_n^{ne} and x_{vu}^{ne} except for those claimed in the Lemma. Similarly, (6) immediately gives a relationship on costs that contradicts (11)-(12) for any values except $x_{vs}^{\text{ne}} = 0$.

It remains to show the second cases of (25) and (26). If $\beta q(y) > 0$, the the inequality described in (27) becomes strict. Consider (25) and assume $P = \frac{1}{1+r}$. Then, by (28) and the above, we know that $x_{vu}^{\text{ne}} = y$. By (3), this provides an algebraic expression for P as a function of solely x_n^{ne} , which can be solved to produce the desired result. A proof of (26) is nearly identical.

Otherwise, we must have that $\beta q(y) = 0$. In the second case of either (25) or (26), (3) simplifies to

$$P(x) = p(x_n^{\text{ne}} + x_{vu}^{\text{ne}}) = \frac{1}{1+r}.$$

This implies that all signaling equilibria that exist satisfy (14) and (15), so they are essentially identical. Therefore, the tuple satisfying (25) and (26) is a signaling equilibrium, and is essentially unique. ■

Recall that (19)-(23) divide our parameter space into five regions. Lemma 4.2 shows that each of these regions restricts the possible values of P .

Lemma 4.2: For any signaling game $G = (\beta, y, r)$, at least one of (19)-(23) is true, and

$$(19) \implies P < \frac{1}{1+r}, \quad (29)$$

$$(20) \implies P = \frac{1}{1+r}, \quad (30)$$

$$(21) \implies P = p(0), \quad (31)$$

$$(22) \implies P = \frac{1}{1+r(1-\beta q(y))}, \quad (32)$$

$$(23) \implies \frac{1}{1+r} < P < \frac{1}{1+r(1-\beta q(y))}. \quad (33)$$

The five claims are proved via contradiction. Applying Lemma 4.1 to the contradiction hypothesis gives the only possible equilibrium tuple x^{ne} . Next, using this tuple and (3), 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: First note that (23) is simply the compliment of the other conditions, meaning that at least one must be true for any game G .

Since the remaining technique is identical across all five claims, we will prove only (30) for brevity. Consider the case where G is in the range defined by defined by (20), i.e.

$$p\left(\left(1 - \frac{\beta q(y)}{1+r}\right)y\right) \leq \frac{1}{1+r} \leq p\left(1 - \frac{\beta q(y)}{1+r}y\right).$$

If we assume by way of contradiction that $P(x^{\text{ne}}) < \frac{1}{1+r}$, by (27) and (28), $\mathbb{P}(A|\neg S) < \frac{1}{1+r}$. Therefore, by Lemma 4.1, $x_n^{\text{ne}} = 1 - y$, $x_{vu}^{\text{ne}} = y$, and $x_{vs}^{\text{ne}} = 0$. We then substitute into (3) to obtain $P(x^{\text{ne}}) = p(1 - y + (1 - P\beta q(y))y + 0) = p(1 - P\beta q(y)y)$.

Then, starting with our contradiction hypothesis, we perform algebraic operations to take P "up" one level of this recursive definition. This gives that

$$1 - \frac{\beta q(y)}{1+r}y < 1 - P\beta q(y)y.$$

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

$$p\left(1 - \frac{\beta q(y)}{1+r}y\right) < p(1 - P\beta q(y)y).$$

Therefore, applying (20), we have that

$$\frac{1}{1+r} \leq p\left(1 - \frac{\beta q(y)}{1+r}y\right) < p(1 - P\beta q(y)y) = P < \frac{1}{1+r},$$

an obvious contradiction. This technique can also be used to show that assuming $P > \frac{1}{1+r}$ forces a contradiction. Therefore, we must have that $P(x^{\text{ne}}) = \frac{1}{1+r}$.

A proof of the remaining claims can be accomplished in an identical manner. ■

From Lemma 4.2 we now know the possible long term accident probabilities in any region of parameter space. Based on these values, using Lemma 4.1, we can derive what a signaling equilibrium in each range must look like.

Lemma 4.3: For any signaling game $G = (\beta, y, r)$, a unique signaling equilibrium x_n^{ne} exists and takes the following form:

$$(19) \implies x^{\text{ne}} = (1 - y, y, 0), \quad (34)$$

$$(20) \implies x^{\text{ne}} = \left(p^{-1}\left(\frac{1}{1+r}\right) - \left(1 - \frac{\beta q(y)}{1+r}\right)y, y, 0\right), \quad (35)$$

$$(21) \implies x^{\text{ne}} = (0, 0, 0), \quad (36)$$

$$(22) \implies x^{\text{ne}} = \left(0, \frac{p^{-1} \left(\frac{1}{1+r(1-\beta q(y))} \right)}{1 - \frac{\beta q(y)}{1+r(1-\beta q(y))}}, 0 \right), \quad (37)$$

$$(23) \implies x^{\text{ne}} = (0, y, 0). \quad (38)$$

Proof: We prove this in cases. First, assume that G is in the range defined by (19) i.e.

$$p \left(1 - \frac{\beta q(y)}{1+r} y \right) < \frac{1}{1+r}.$$

By Lemma 4.2, $P < \frac{1}{1+r}$. Then, by (27) and (28), $\mathbb{P}(A|\neg S) < \frac{1}{1+r}$. Finally, by Lemma 4.1, $(1-y, y, 0)$ is a signaling equilibrium and essentially unique.

An identical method can be used to show that $(0, 0, 0)$ and $(0, y, 0)$ are essentially unique signaling equilibria if G is in the range of (21) or (23), respectively.

Now, assume that G is in the range defined by (20). By Lemma 4.2, $P(x^{\text{ne}}) = \frac{1}{1+r}$. If $\beta q(y) > 0$, then $\frac{1}{1+r} < \frac{1}{1+r(1-\beta q(y))}$, so by (28), $\mathbb{P}(A|\neg S) < \frac{1}{1+r}$. By Lemma 4.1,

$$\left(p^{-1} \left(\frac{1}{1+r} \right) - \left(1 - \frac{\beta q(y)}{1+r} \right) y, y, 0 \right)$$

is then an essentially unique signaling equilibrium. Otherwise, $\beta q(y) = 0$, so $Q(x^{\text{ne}}) = 0$ and (28) implies that $\mathbb{P}(A|\neg S) = \frac{1}{1+r}$. Therefore by Lemma 4.1, $x_n^{\text{ne}} = p^{-1}(P(x^{\text{ne}})) - (1 - Q(x^{\text{ne}}))y = p^{-1}(\frac{1}{1+r}) - y$ and $x_{\text{vu}}^{\text{ne}} = \frac{p^{-1}(P(x^{\text{ne}}))}{1-Q(x^{\text{ne}})} = p^{-1}(\frac{1}{1+r})$. By (3), 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 (22) forces a signaling equilibrium of the form claimed.

By Lemma 4.2, 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 4.3 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. We have shown 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. For future work, it will be interesting to situate these models in the context of network routing problems or to consider more expressive signaling policies.

- [1] 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.
- [2] M. Gairing, B. Monien, and K. Tiemann, "Selfish Routing with Incomplete Information," *Theory of Computing Systems*, vol. 42, pp. 91–130, Jan. 2008.
- [3] 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. Publisher: INFORMS.
- [4] S. Dafermos and A. Nagurney, "On some traffic equilibrium theory paradoxes," *Transportation Research Part B: Methodological*, vol. 18, pp. 101–110, Apr. 1984.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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. 1–1, Feb 2021.
- [9] D. A. Lazar, E. Biyik, 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.
- [10] E. Biyik, 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. Conference Name: IEEE Transactions on Control of Network Systems.
- [11] Y. Zhu and K. Savla, "Information Design in Non-atomic Routing Games with Partial Participation: Computation and Properties," *arXiv:2005.03000*, 2020.
- [12] 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.
- [13] 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.
- [14] E. Kamenica and M. Gentzkow, "Bayesian Persuasion," *American Economic Review*, vol. 101, pp. 2590–2615, Oct. 2011.
- [15] D. Bergemann and S. Morris, "Information Design: A Unified Perspective," *Journal of Economic Literature*, vol. 57, pp. 44–95, Mar. 2019.
- [16] 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.
- [17] C. Roman and P. Turrini, "How does information affect asymmetric congestion games?," *arXiv:1902.07083 [cs, math]*, Feb. 2019. arXiv: 1902.07083.
- [18] J. Liu, S. Amin, and G. Schwartz, "Effects of Information Heterogeneity in Bayesian Routing Games," *arXiv:1603.08853 [cs]*, Mar. 2016. arXiv: 1603.08853.
- [19] 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.
- [20] H. Tavaafoghi and D. Teneketzis, "Informational incentives for congestion games," in *55th Annual Allerton Conference on Communication, Control, and Computing*, (Allerton), p. 35, 2017.
- [21] 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.
- [22] 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.