Information Design Under Uncertainty for Vehicle-to-Vehicle Communication

Brendan T. Gould and Philip N. Brown

***Abstract*— The emerging technology of Vehicle-to-Vehicle (V2V) communication aims to improve road safety by allowing vehicles to share information about the world. However, information design is in general a non-trivial problem, and is only made more difficult by uncertainty about the world or agents. In this work, using an existing model of V2V communication with endogenous accident probability, we study an information designer's optimization problem under uncertainty about the "danger level" (the sensitivity of accident probability to agent behavior). First, we consider an information designer who does not know the danger level designing for agents who do; second, an informed designer designing for uninformed agents. In both cases, we present a simple characterization of the worst-case (i.e. largest accident probability) outcome that is possible under the uncertainty. When an information designer is uncertain about the world, the worst case occurs with the largest danger level. By contrast, when agents are uninformed, the worst case is caused by agents' beliefs being the *lowest* danger level. Both of these results simplify the optimization problem, allowing an optimal signaling policy to be more easily determined.**

I. Introduction

Technologies such as the Internet of Things (IoT) and Vehicle-to-Vehicle (V2V) communication are being rapidly developed and integrated into every aspect of human life. This integration has benefits, but also brings with it new challenges for the designers of these technologies. With the expectation of an engineered product being used in and for human society, engineers must carefully consider the effects that their technologies and human behavior will have on each other. These *socio-technical* systems take many forms, and recent work has identified situations where they may actually harm the population they are intended to help [1], [2].

An important context for this is road traffic, usually modelled as a congestion game [3], [4]. In this context, emergent outcomes resulting from selfish individual behavior can be far from socially optimal [5], [6], so engineers attempt to build systems to improve these outcomes. Attempted solutions include modifying the road network to alleviate congestion [7], incentivizing drivers to choose certain behaviors [8], [9], and strategically disclosing information to drivers to influence their decisions [10] (this is known as *information design*). However, in each of these cases, naively implementing the given solution can make the population worse off than before any intervention took place [11]–[13].

This problem is only made more complicated by considering scenarios where exact information may not be available. Particularly in information design, prior work assumes that underlying probability distributions are perfectly known [14]–[16]. However, it is also known that human behavior can be irrational and difficult to model [17]–[20], meaning technical systems must account for the fact that the behavior of their users cannot be perfectly predicted. Prior work in other domains has approached this problem by investigating control policies robust to unknown environments [21], [22], unknown agent utility functions [12], [23], [24], and even unknown agent decision making policies [25], but it is still largely an open question in the V2V domain.

In this paper, we use a model of V2V communication from our prior work [1] to ask how an information designer should design for a world and agents when only partial information is known. One unique aspect of our work is that road hazards occur endogenously; the behaviors chosen by drivers affect the probability of an accident occurring. This relationship is complex and difficult to predict (as shown by the paradoxes described in [1]), so it is likely unreasonable for information designers to have perfect information about it. In particular, there may be discrepancies between the accident-causing mechanism predicted by the information designer and the actual mechanism used by the world, or between the agents' belief about what will happen and the real world truth.

This work initiates a study of how uncertainty about endogenous accident causing mechanisms can affect the optimal signaling policy. We parameterize uncertainty using a new quantity we call the *danger level*, which describes how sensitive the induced accident probability is to driver behavior (higher danger levels imply higher accident probabilities for given driver behavior). In Theorem 3.1, we show that if a designer does not know the danger level of the world, then the worst-case accident probability is caused by the highest possible danger level. By contrast, we show in Proposition 3.2 that if a designer knows the world's danger level, but not the agents' beliefs about it, then the worst case is caused by the agents believing the lowest feasible value.

II. Model

A. Game Setup

We consider a non-atomic population of drivers interacting on a single road where accidents sometimes occur. Each driver may choose between careful (C) and reckless (R) driving behaviors. Reckless drivers reach their destination slightly faster, but will "pile on" to any existing accidents (A). Careful drivers avoid accidents, but incur a "regret" cost for the needless care if no accidents are encountered. These incentives are summarized in the following cost matrix:

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 material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-23-1-0171 and the National Science Foundation under award number ECCS-2013779.

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

where $r > 1$ is the cost of an accident.

Intuitively, driver behavior can affect the probability of an accident, and more drivers choosing reckless behaviors cause accidents to be more likely. Therefore, we let the class of functions $p_a(d)$ describe the probability of an accident when the average mass of reckless drivers is d. Here, a describes the sensitivity of accident probability to driver recklessness, and is called the *danger level*. The danger level parameter is a unique addition over similar models from prior work, and allows us to model uncertainty about how drivers' actions will affect the world. We assume that each $p_a(d)$ is continuous and increasing in d, non-decreasing in a, and $p_{a_1}(0) = p_{a_2}(0)$ for any a_1, a_2.

A fraction y of cars in the population are equipped with Vehicle-to-Vehicle communication (V2V) technology, which can autonomously detect accidents and broadcast warning signals about them to other V2V cars. If an accident exists, then with probability $t(y)$ it is detected and a warning signal is successfully broadcast. Otherwise, a "false-positive" warning signal is broadcast with probability $f(y)$. We assume that whenever a signal is broadcast it is received by all V2V cars, and that false positive signals are strictly less likely than true positives (i.e. $f(y) < t(y)$).

Finally, we define a game as the tuple of danger level, V2V mass, and crash cost: $G = (a, y, r)$.

B. Signaling Equilibria

Prior work has shown that full disclosure is not always optimal in information design problems [5], [6]. For example, V2V drivers could become accustomed to receiving a signal whenever there is an accident, and falsely assume that the lack of a signal implies the road is safe. This could increase driver recklessness and therefore accident probability. To that end, we consider the idea that V2V cars should sometimes *not* notify their drivers when a signal is received.

Call $\beta \in [0, 1]$ the *information quality*. When a V2V car receives a signal, it will display a warning to its driver with probability β. With probability $1 - \beta$, it will do nothing (and ignore all future signals from the same accident). The signaler aims to choose β to minimize accident probability.

If an agent drives a car with V2V technology, we call them a *V2V driver* (v), otherwise, a *non-V2V driver* (n). We further differentiate V2V drivers by whether their car has displayed a warning: if it has, they are *signaled* V2V drivers (vs), otherwise they are *unsignaled* (vu). We write $x_{\rm n}, x_{\rm vs}, x_{\rm vu}$ to denote the mass of drivers in each group choosing to behave recklessly, and a behavior profile as $x = (x_{\rm n}, (x_{\rm vu}, x_{\rm vs}))$.

Our model allows driver behavior to endogenously determine accident probability. However, drivers choose behaviors based on accident probability, creating a recursive relationship. This relationship is not present in standard Bayesian games and significantly complicates model analysis. Non-V2V drivers are unable to receive signals, and can choose only one set of behaviors ($x_{\rm n}$). V2V drivers can condition their behavior on the signal realization, meaning we must specify their choices in both cases to fully describe their behavior ($x_{\rm vu}$ and $x_{\rm vs}$). The probability of a signal is

$$\mathbb{P}[\mathrm{S}] = \beta(\mathbb{P}[\mathrm{A}]t(y) + (1 - \mathbb{P}[\mathrm{A}])f(y)), \tag{1}$$

and the *habitual* behavior of V2V drivers is the weighted average of both behaviors: $x_{\rm v} = \mathbb{P}[\neg \mathrm{S}]x_{\rm vu} + \mathbb{P}[\mathrm{S}]x_{\rm vs}$.

For convenience, we denote accident probability as $P(G, x, \beta) := \mathbb{P}[\mathrm{A}]$ and signal probability as $Q(G, x, \beta) := \mathbb{P}[\mathrm{S}]$. Then, accident probability is given by the choices of non-V2V drivers and the habitual behavior of V2V drivers:

$$P(G,x,\beta) = p_a(x_{\rm n} + (1 - Q(G,x,\beta))x_{\rm vu} + Q(G,x,\beta)x_{\rm vs}). \tag{2}$$

This accident probability determines the cost of both actions to a driver, given the information available to them:

$$J_{\rm n}(a;x) = \begin{cases} 1 - \mathbb{P}[\mathrm{A}] & \text{if } a = \mathrm{C}, \\ r\mathbb{P}[\mathrm{A}] & \text{if } a = \mathrm{R}, \end{cases} \tag{3a}$$

$$J_{\rm vu}(a;x) = \begin{cases} 1 - \mathbb{P}[\mathrm{A}|\neg \mathrm{S}] & \text{if } a = \mathrm{C}, \\ r\mathbb{P}[\mathrm{A}|\neg \mathrm{S}] & \text{if } a = \mathrm{R}, \end{cases} \tag{3b}$$

$$J_{\rm vs}(a;x) = \begin{cases} 1 - \mathbb{P}[\mathrm{A}|\mathrm{S}] & \text{if } a = \mathrm{C}, \\ r\mathbb{P}[\mathrm{A}|\mathrm{S}] & \text{if } a = \mathrm{R}. \end{cases} \tag{3c}$$

Our solution concept assumes that if any agent is choosing an action, then its cost to them is minimal. This is true when:

$$x_{\rm n} < 1 - y \implies J_{\rm n}(\mathrm{C};x) \le J_{\rm n}(\mathrm{R};x), \tag{4a}$$
$$x_{\rm n} > 0 \implies J_{\rm n}(\mathrm{R};x) \le J_{\rm n}(\mathrm{C};x), \tag{4b}$$
$$x_{\rm vu} < y \implies J_{\rm vu}(\mathrm{C};x) \le J_{\rm vu}(\mathrm{R};x), \tag{4c}$$
$$x_{\rm vu} > 0 \implies J_{\rm vu}(\mathrm{R};x) \le J_{\rm vu}(\mathrm{C};x), \tag{4d}$$
$$x_{\rm vs} < y \implies J_{\rm vs}(\mathrm{C};x) \le J_{\rm vs}(\mathrm{R};x), \tag{4e}$$
$$x_{\rm vs} > 0 \implies J_{\rm vs}(\mathrm{R};x) \le J_{\rm vs}(\mathrm{C};x). \tag{4f}$$

We define a *signaling equilibrium* as a behavior profile x satisfying both the incentive conditions in (4a)–(4f) and the consistency condition in (2). For any G and β, [1, Lemma 4.2] gives an equilibrium behavior profile; let $\mathcal{E}(G, \beta)$ represent this equilibrium.[1] We will sometimes write $P(G, \beta)$ to mean $P(G, x, \beta)$ for $x = \mathcal{E}(G, \beta)$.

C. Research Goal: Design Under Uncertainty

Our prior work has considered similar contexts where it is assumed that the value of a in the $p_a(d)$ function is known by both agents and the signal designer [1]. However, this assumption is likely unrealistic, as the exact mechanisms causing accidents are complex and difficult to measure. Therefore, we wish to study the effects of uncertainty on the information design problem. We formally state this question in the form of two optimization problems using (5) and (6).

We model uncertainty using the danger level parameter a. Using an uncertainty radius $\delta > 0$ and uncertainty

[1] We prove in [1] that these equilibria are *essentially unique*. That is, all equilibrium behavior tuples induce the same accident probability for any game G. Therefore, our results hold even if $\mathcal{E}(G, \beta)$ is an arbitrary equilibrium. We make this assumption to utilize the explicit form of these behavior profiles in our proofs.

center $a^* \in \mathbb{R}$, construct the range $[a^* - \delta, a^* + \delta]$. This range represents e.g. a confidence interval on the danger level that can be computed from historical data, providing a computationally feasible way to estimate a. In this work, we consider two different types of uncertainty.

First, we consider an information designer with uncertainty about the world. This designer knows the true danger level a lies in the interval $[a^* - \delta, a^* + \delta]$, but not its exact value. By contrast, agents do know the exact danger level. This case is motivated by the idea that drivers may have experiential knowledge of their local road that is not available to a signaler from out of town. We would like to find a signaling policy that minimizes the worst-case equilibrium crash probability. That is, we seek a value of β such that

$$\beta \in \arg\min_{\beta\in[0,1]} \max_{a\in[a^*-\delta,a^*+\delta]} P((a, y, r), \beta). \tag{5}$$

Second, we model a signaler who has uncertainty about the agents, and cannot perfectly predict agent behavior. Specifically, the signaler knows the true danger level a, while agents believe the danger level has some different value $\tilde{a}$. The signaler knows that $\tilde{a} \in [a^* - \delta, a^* + \delta]$, but not its exact value. Given the believed danger level $\tilde{a}$, agents choose behavior $\tilde{x} = \mathcal{E}((\tilde{a}, y, r), \beta)$. The signaler aims to minimize accident probability induced by the true danger level a and this behavior, again in the worst case:

$$\beta \in \arg\min_{\beta\in[0,1]} \max_{\tilde{a}\in[a^*-\delta,a^*+\delta]} P((a, y, r), \tilde{x}, \beta). \tag{6}$$

III. Results

Our results consist of an explicit solution to the optimization problem posed in (5), and a characterization of (6) allowing us to efficiently compute a nearly optimal signaling policy. Surprisingly, under both types of uncertainty, equilibrium accident probability is monotonic with respect to danger level, greatly simplifying the optimization problems.

In the case when the designer is uncertain about the world, we show that the worst-case accident probability is induced when the world has the largest possible danger level, allowing the signaler to design only for the game with this danger level. This simplifies the optimization problem in (5) to exactly the one described and solved in [1], immediately giving the optimal robust signaling policy.

Theorem 3.1: Let $G_{\max} := (a^* + \delta, y, r)$ denote the game with the largest feasible danger level. This game induces the worst-case crash probability under world uncertainty:

$$\arg\min_{\beta\in[0,1]} \max_{a\in[a^*-\delta,a^*+\delta]} P(G, \beta) = \arg\min_{\beta\in[0,1]} P(G_{\max}, \beta). \tag{7}$$

In other words, an information designer minimizing accident probability under uncertainty about the world may assume that the real danger level is the largest feasible value. See Appendix B for detailed proofs of Theorem 3.1.

Figure 1 shows the "cost of world uncertainty" C_{w}; that is, when the unknown danger level is actually at the center of the uncertainty window, the difference in crash probability induced by a signaling policy that optimizes the worst case

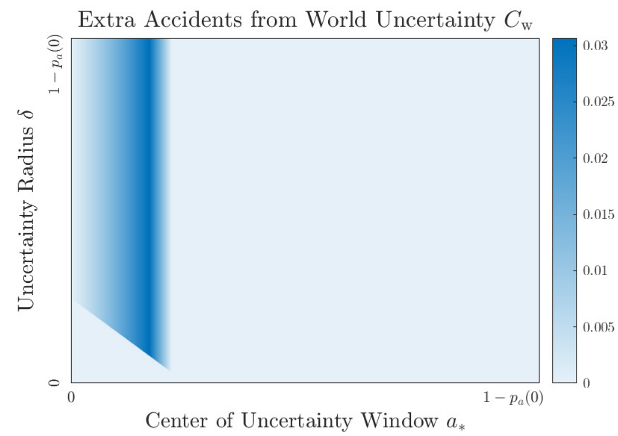

Fig. 1. Increase in accident probability due to uncertainty about the world. For sufficiently low danger levels, it is optimal to choose $\beta = 0$. Otherwise, it is optimal to choose $\beta = 1$. When the true danger level a is sufficiently small (so that a signaler with perfect information chooses $\beta = 0$), but the worst case danger level $a^* + \delta$ is large (so that a signaler with uncertainty chooses $\tilde{\beta} = 1$), uncertainty can cause extra accidents. In other words, uncertainty causes additional accidents by making signalers overuse V2V technology. The example depicted has $p_a(d) = 0.1 + ad$, $y = 0.9$, $r = 3$, $t(y) = 0.8y$, and $f(y) = 0.1y$.

and one that optimizes with full knowledge. If $\tilde{\beta}$ is a signaling policy minimizing accident probability under uncertainty (i.e. $\tilde{\beta} \in \arg\min_{\tilde{\beta}\in[1,1]} P((a^* + \delta, y, r), \tilde{\beta})$), and β is the same with certainty ($\beta \in \arg\min_{\beta\in[0,1]} P((a^*, y, r), \beta)$), then $C_{\mathrm{w}} = P((a^*, y, r), \tilde{\beta}) - P((a^*, y, r), \beta)$. Each point with a positive value in Figure 1 represents a parameter combination where $\tilde{\beta}$ differs from β.

Next, we consider the optimization problem in (6), and ask how an information designer can minimize accidents while uncertain about agents' belief about the world. We show that when the signaler is uncertain about agents, it is the *lowest* feasible danger level that induces the worst-case accident probability. Interestingly, this is exactly the opposite of the case when the signaler is uncertain about the world.

Proposition 3.2: Consider the game with the danger level believed by agents $\tilde{G} = (\tilde{a}, y, r)$ and the game with the smallest feasible danger level $G_{\min} = (a^* - \delta, y, r)$. Then, $G_{\min}$ induces the worst-case equilibrium accident probability under uncertainty about agent beliefs:

$$\max_{\tilde{a}\in[a^*-\delta,a^*+\delta]} P(G, \mathcal{E}(\tilde{G}, \beta), \beta) = P(G, \mathcal{E}(G_{\min}, \beta), \beta). \tag{8}$$

Intuitively, Proposition 3.2 shows that an information designer aiming to reduce accidents for a population of drivers who are uncertain about the world should assume the agents' belief about the danger level is minimal. Formal proofs of this statement are provided in Appendix C.

This simplifies the optimization problem from (6) to

$$\arg\min_{\beta\in[0,1]} P(G, \mathcal{E}(G_{\min}, \beta), \beta). \tag{9}$$

Unfortunately, this simplification is not in a form that directly gives an optimal signaling policy. Due to their uncertainty about the world, the behavior chosen by agents is not necessarily a signaling equilibrium for the real game, meaning we cannot reuse results from [1] in the same way as we did to solve (7). However, a nearly-optimal policy may be found by a grid search between $\beta = 0$ and $\beta = 1$, using techniques from [1] to efficiently compute $P(G, \mathcal{E}(G_{\min}, \beta), \beta)$. Proposition 3.2 is useful in this approach because it removes a potential dimension from the search by guaranteeing that the worst case along the $\tilde{a}$ axis is $a^* - \delta$, saving significant computational effort.

We use this technique to depict the "cost of agent uncertainty" C_{a} in Figure 2. Analogously to C_{w}, C_{a} is defined as the difference between the crash probability induced by a signaling policy minimizing for the worst case, and that of one created with full knowledge of the danger level: $C_{\mathrm{a}} = P(G^*, \mathcal{E}(G^*, \tilde{\beta}), \tilde{\beta}) - P(G^*, \mathcal{E}(G^*, \beta), \beta)$, with $G^* = (a^*, y, r)$, $\beta \in \arg\min_{\beta \in [0,1]} P(G^*, \mathcal{E}(G^*, \beta), \beta)$, and $\tilde{\beta} \in \arg\min_{\tilde{\beta} \in [0,1]} P((a^* - \delta, y, r), \mathcal{E}((a^* - \delta, y, r), \tilde{\beta}), \tilde{\beta})$.

IV. Conclusion

This work considered how V2V communication can be used to minimize accident probability under two different kinds of uncertainty. We showed that when an information designer is uncertain about the danger levels on a road, the worst-case accident probability is caused by the highest danger level. Furthermore, when this sensitivity is unknown to *agents* that are being designed for, the worst case is caused when agents believe that the sensitivity is the *lowest* possible. Building on techniques from prior work, each of these results provide criteria to efficiently determine the accident-minimizing signaling policy for use by a V2V communication system. This work could be extended in the future to include an analytical solution to the simplified optimization problem under agent uncertainty (presented in Proposition 3.2), contexts where both the information designer and agents are unsure of the world's danger level, or consider alternative ways to model uncertainty (e.g. explicitly assigning a probability distribution to belief).

Appendix

We now present formal proofs of our claimed results. First, in Appendix A, we state useful results from [1] that establish necessary characteristics of signaling equilibria. These characteristics are then used to prove the desired claims. In Appendix B, we prove Theorem 3.1, and in Appendix C we prove Proposition 3.2.

A. Calculating an Equilibrium

The following equations ((10)–(15)) are tools borrowed from [1], which we use to describe equilibrium behavior. We calculate two sets of thresholds ((10) and (12)), and determine their ordering. Each possible ordering uniquely selects a "family" of possible equilibria in (13), and all equilibria in a family share similar characteristics. For each family, (14) describes the equilibrium behavior tuple and (15) describes accident probability.

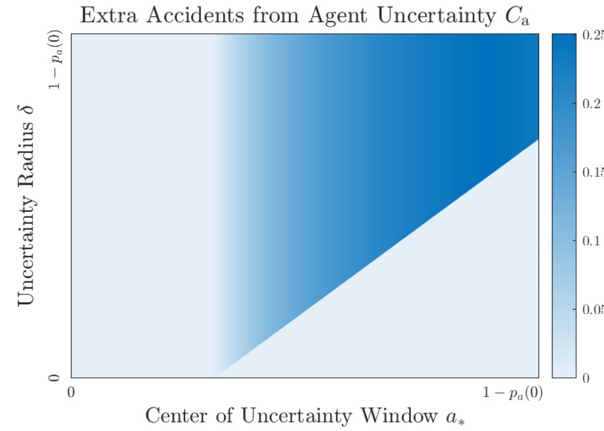

Fig. 2. Increase in accident probability due to uncertainty about agent beliefs. This occurs when agents believe that the danger level is large, and the uncertainty radius is at least as large as a minimum threshold. The example depicted has $p_a(d) = 0.1 + ad$, $y = 0.9$, $r = 3$, $t(y) = 0.8y$, and $f(y) = 0.1y$.

The first set of thresholds is:

$$P_{\mathrm{vs}} := \frac{f(y)}{rt(y) + f(y)}, \tag{10a}$$

$$P_{\mathrm{n}} := \frac{1}{1 + r}, \tag{10b}$$

$$P_{\mathrm{vu}} := \frac{1 - \beta f(y)}{1 + r(1 - \beta t(y)) - \beta f(y)}, \tag{10c}$$

where it holds that

$$P_{\mathrm{vs}} < P_{\mathrm{n}} \le P_{\mathrm{vu}}. \tag{11}$$

Each of these is a prior accident probability that causes a group of drivers to be indifferent between careful and reckless behavior. By Bayes' Theorem, $P(G, x, \beta) = P_{\mathrm{vs}} \implies J_{\mathrm{vs}}(\mathrm{C}; x) = J_{\mathrm{vs}}(\mathrm{R}; x)$, and similarly for the other two groups of drivers. For each accident probability threshold, let Q_{vs}, Q_{n}, and Q_{vu} be the corresponding signal probabilities.

The second set of thresholds is:

$$E_{1U}(a) := p_a(0)\,, \tag{12a}$$

$$E_{2U}(a) := p_a((1 - \beta P_{\mathrm{vu}}(t(y) - f(y)) - \beta f(y))y), \tag{12b}$$

$$E_{3U}(a) := p_a((1 - \beta P_{\mathrm{n}}(t(y) - f(y)) - \beta f(y))y)\,, \tag{12c}$$

$$E_{4U}(a) := p_a(1 - (\beta P_{\mathrm{n}}(t(y) - f(y)) + \beta f(y))y)\,, \tag{12d}$$

$$E_{5U}(a) := p_a(1 - (\beta P_{\mathrm{vs}}(t(y) - f(y)) + \beta f(y))y)\,, \tag{12e}$$

$$E_{6U}(a) := p_a(1)\,. \tag{12f}$$

Equilibrium families are defined in the following way:

$$E_1 := \{(a, y, r) : P_{\mathrm{vu}} < E_{1U}(a)\}, \tag{13a}$$

$$E_2 := \{(a, y, r) : E_{1U}(a) \le P_{\mathrm{vu}} \le E_{2U}(a)\}, \tag{13b}$$

$$E_3 := \{(a, y, r) : E_{2U}(a) < P_{\mathrm{vu}} \wedge P_{\mathrm{n}} < E_{3U}(a)\}, \tag{13c}$$

$$E_4 := \{(a, y, r) : E_{3U}(a) \le P_{\mathrm{n}} \le E_{4U}(a)\}, \tag{13d}$$

$$E_5 := \{(a, y, r) : E_{4U}(a) < P_{\mathrm{n}} \wedge P_{\mathrm{vs}} < E_{5U}(a)\}, \tag{13e}$$

$$E_6 := \{(a, y, r) : E_{5U}(a) \le P_{\mathrm{vs}} \le E_{6U}(a)\}, \tag{13f}$$

$$E_7 := \{(a, y, r) : E_{6U}(a) < P_{\mathrm{vs}}\}. \tag{13g}$$

Now, we describe equilibrium behavior and accident probability in each of the "families" defined in (13). This simplifies much of the analysis to determining which family a parameter combination belongs to. With uncertainty about the world, any particular value of the danger level will immediately determine agent behavior through (14) and accident probability using (15). Even with uncertainty about agent beliefs, where the chosen behavior may not be an equilibrium, (14) and a value for the danger level believed by agents will determine the behaviors they choose.

By [1, Lemma 4.2], we know the form of equilibrium behavior tuples. For any game G, any of its signaling equilibria is essentially identical to the following behavior tuple x:

$$G \in E_1 \implies x = (0,(0,0)), \tag{14a}$$

$$G \in E_2 \implies x = \left(0, \left(\frac{p_a^{-1}(P_{\mathrm{vu}})}{1-Q_{\mathrm{vu}}}, 0\right)\right), \tag{14b}$$

$$G \in E_3 \implies x = (0,(y,0)), \tag{14c}$$

$$G \in E_4 \implies x = \left(p_a^{-1}(P_{\mathrm{n}})) - (1-Q_{\mathrm{n}})y, (y,0)\right), \tag{14d}$$

$$G \in E_5 \implies x = (1-y,(y,0)), \tag{14e}$$

$$G \in E_6 \implies x = \left(1-y, \left(y, \frac{p_a^{-1}(P_{\mathrm{vs}}) - 1 + Q_{\mathrm{vs}}y}{Q_{\mathrm{vs}}}\right)\right), \tag{14f}$$

$$G \in E_7 \implies x = (1-y,(y,y)). \tag{14g}$$

Finally, by [1, Lemma 4.1], we know that each equilibrium family restricts the accident probability in a specific way:

$$G \in E_1 \implies P(G,\beta) = p_a(0), \tag{15a}$$

$$G \in E_2 \implies P(G,\beta) = P_{\mathrm{vu}}, \tag{15b}$$

$$G \in E_3 \implies P_{\mathrm{n}} < P(G,\beta) < P_{\mathrm{vu}}, \tag{15c}$$

$$G \in E_4 \implies P(G,\beta) = P_{\mathrm{n}}, \tag{15d}$$

$$G \in E_5 \implies P_{\mathrm{vs}} < P(G,\beta) < P_{\mathrm{n}}, \tag{15e}$$

$$G \in E_6 \implies P(G,\beta) = P_{\mathrm{vs}}, \tag{15f}$$

$$G \in E_7 \implies P(G,\beta) = p_a(1). \tag{15g}$$

B. Proofs of Theorem 3.1

An outline of the proof of Theorem 3.1 is as follows: First, we show that the equilibrium families are ordered by danger level — if two games are identical except for their values of a, the one with the larger danger level can only have equilibria with a lower mass of reckless drivers (Lemma 1.1). This makes intuitive sense: as accidents become more likely, we expect that more agents will want to drive carefully in response. Next, we use the ordering on equilibrium families to establish the monotonicity of accident probability with respect to danger level (Lemma 1.2). Showing that equilibrium accident probability is non-decreasing in a immediately implies the desired result.

Lemma 1.1: Let $a_1 < a_2$. Further, let $G_1 = (a_1, y, r)$, and $G_2 = (a_2, y, r)$. For any $i \in \{1, \dots, 7\}$, $G_1 \in E_i \implies G_2 \in E_j$ for some $j \le i$.

Proof: We prove this in cases. First note that if $i = 7$, the statement is vacuously true since every game belongs to some equilibrium family by [1, Theorem 3.1].

Consider the case where $i = 1$, and assume by way of contradiction that $G_2 \notin E_1$. Since $G_1 \in E_1$, we have that $P_{\mathrm{vu}} < E_{1U}(a_1)$ by (13a). Additionally, $E_{1U}(a_2) \le P_{\mathrm{vu}}$, lest $G_2 \in E_1$. But then since $E_{1U}(a_1) = p_{a_1}(0) = p_{a_2}(0) = E_{1U}(a_2)$, this implies $P_{\mathrm{vu}} < E_{1U} \le P_{\mathrm{vu}}$, which is a clear contradiction.

Now, let $i = 2$ and assume by contradiction that $G_2 \notin E_1 \cup E_2$. In the same way as above, we immediately have that $E_{2U}(a_2) < P_{\mathrm{vu}}$ and $P_{\mathrm{vu}} < E_{2U}(a_1)$. Note that $p_a(d)$ is non-decreasing in a, so $E_{2U}(a_1) < E_{2U}(a_2)$ by (12b). But again, this implies that $P_{\mathrm{vu}} < P_{\mathrm{vu}}$, a clear contradiction.

The remaining cases can be shown in a very similar manner, completing the proof. ■

Now, we show that accident probability is increasing with respect to danger level.

Lemma 1.2: Let $a_1 < a_2$, and define $G_1 = (a_1, y, r)$ and $G_2 = (a_2, y, r)$. For any signaling policy β, $P(G_1, \beta) \le P(G_2, \beta)$.

Proof: We prove this in cases. First, assume that $G_1 \in E_7$. By Lemma 1.1, $G_2 \in \bigcup_{i=1}^{7} E_i$.

If $G_2 \in E_6$, then $P(G_2, x, \beta) = P_{\mathrm{vs}}$ by (15f). Similarly using (15g), since $G_1 \in E_7$, $P(G_2, x, \beta) < P_{\mathrm{vs}}$. But then $P(G_1, x, \beta) < P(G_2, x, \beta)$, which is the desired result.

If $G_2 \in \bigcup_{i=2}^{6} E_i$, then a very similar technique suffices. Furthermore, it is impossible that $G_2 \in E_1$. Assume by way of contradiction that it is. Then, by (13a) we have that $P_{\mathrm{vu}} < E_{1U}(a_2) = p_{a_2}(0)$. Similarly, by (13g), $E_{6U}(a_1) = p_{a_1}(1) < P_{\mathrm{vs}}$. But then we have

$$P_{\mathrm{vu}} < p_{a_2}(0) = p_{a_1}(0) < p_{a_1}(1) < P_{\mathrm{vs}}, \tag{16}$$

which clearly contradicts (11). Thus, we are finished in the case that $G_1 \in E_7$.

The above techniques suffice in any of the remaining cases where $G_1 \in \bigcup_{i=1}^{6} E_i$, completing the proof. ■

This lets us immediately derive the signaling policy minimizing crash probability in the worst case. Since increasing danger level cannot decrease accident probability, the game with the largest danger level must also have the largest accident probability.

Proof of Theorem 3.1: Immediately from Lemma 1.2, we have that

$$\max_{a \in [a^* - \delta, a^* + \delta]} P(G, \beta) = P(G_{\max}, \beta). \tag{17}$$

Since the optimization problems are identical, the set of optimizers must also be, completing the proof. ■

C. Proofs of Proposition 3.2

We now provide a complete proof of Proposition 3.2. To do this, we first show that agents who believe that accident probability is less sensitive to the mass of reckless drivers (i.e. that a is lower) are necessarily more reckless (Lemma 1.3). Next, in Lemma 1.4, we show that a larger mass of reckless drivers causes more accidents at equilibrium with all else held constant. Together, these imply that if agents believe the true danger level is lower, then more accidents will occur at equilibrium, as claimed.

Lemma 1.3: Consider any $a_1 < a_2$. Let $G_1 = (a_1, y, r)$ and $G_2 = (a_2, y, r)$. Further let $x_1 = \mathcal{E}(G_1, \beta)$ and $x_2 = \mathcal{E}(G_2, \beta)$ for any β. Then, x_1 is *more reckless* than x_2, i.e. $x_{\mathrm{n}1} \geq x_{\mathrm{n}2}$, $x_{\mathrm{vu}1} \geq x_{\mathrm{vu}2}$, and $x_{\mathrm{vs}1} \geq x_{\mathrm{vs}2}$.

Proof: We prove this in cases. First, assume that $G_1 \in E_7$. By Lemma 1.1, $G_2 \in \bigcup_{i=1}^{7} E_i$. Then, the desired result is obvious from equations (14a)–(14g).

Now, assume $G_2 \in E_6$. Lemma 1.1 implies that $G_2 \in \bigcup_{i=1}^{6} E_i$. Unless $G_2 \in E_6$, (14a)–(14g) are again sufficient for the claimed result. If it is, then it remains to show that $\frac{p_a^{-1}(P_{\mathrm{vs}})-1+Q_{\mathrm{vs}}y}{Q_{\mathrm{vs}}}$ is decreasing with a.

By definition of p, we have that $p_{a_1}(0) = p_{a_2}(0) \leq p_{a_1}(x) \leq p_{a_2}(x)$ for any x. Since p is continuous, there exists some $z \in [0, x]$ such that $p_{a_2}(z) = p_{a_1}(x)$; call this shared quantity n. But then $p_{a_1}^{-1}(n) = x$, $p_{a_2}^{-1}(n) = z$, and $z \leq x$. Therefore, $p_{a_1}^{-1}(x) \geq p_{a_2}^{-1}(x)$ for any x. Since every other term of $\frac{p_a^{-1}(P_{\mathrm{vs}})-1+Q_{\mathrm{vs}}y}{Q_{\mathrm{vs}}}$ is constant with respect to a, this is the desired result.

The remaining cases can be shown in a similar manner, completing the proof. ■

It remains to show how the increased recklessness established in the above result affects equilibrium accident probability.

Lemma 1.4: Consider two behavior tuples $x_1 = (x_{\mathrm{n}1}, (x_{\mathrm{vu}1}, x_{\mathrm{vs}1}))$ and $x_2 = (x_{\mathrm{n}2}, (x_{\mathrm{vu}2}, x_{\mathrm{vs}2}))$ with $x_{\mathrm{n}1} \leq x_{\mathrm{n}2}$, $x_{\mathrm{vu}1} \leq x_{\mathrm{vu}2}$, and $x_{\mathrm{vs}1} \leq x_{\mathrm{vs}2}$. For constant G and β, we have that $P(G, x_1, \beta) \leq P(G, x_2, \beta)$.

Proof: Assume by way of contradiction that $P(G, x_1, \beta) > P(G, x_2, \beta)$. This immediately implies that $Q(G, x_1, \beta) > Q(G, x_2, \beta)$ by (1), and therefore $(1 - Q(G, x_1, \beta))x_1 \leq (1 - Q(G, x_2, \beta))x_2$.

We consider two cases. First, assume that $x_{\mathrm{vu}1} < x_{\mathrm{vu}2}$, meaning $x_{\mathrm{vu}1} < y$. By (14a) – (14g), we have $x_{\mathrm{n}1} = x_{\mathrm{vs}1} = 0$. Therefore, $0 = Q(G, x_1, \beta)x_{\mathrm{vs}1} \leq Q(G, x_2, \beta)x_{\mathrm{vs}2}$. Additionally, $x_{\mathrm{n}1} \leq x_{\mathrm{n}2}$ by assumption. Therefore, by equation (2), $P(G, x_1, \beta) \leq P(G, x_2, \beta)$, which is an obvious contradiction, so we are finished in this case.

Otherwise, we must have $x_{\mathrm{vu}1} = x_{\mathrm{vu}2}$. Simple algebra starting from (2) gives

$$P(G, x, \beta) = p_a(x_{\mathrm{n}} + x_{\mathrm{vu}} - Q(G, x, \beta)(x_{\mathrm{vu}} - x_{\mathrm{vs}})). \quad (18)$$

By assumption, $x_{\mathrm{n}1} \leq x_{\mathrm{n}2}$ and $x_{\mathrm{vs}1} \leq x_{\mathrm{vs}2}$. But then $(x_{\mathrm{vu}1} - x_{\mathrm{vs}1}) \geq (x_{\mathrm{vu}2} - x_{\mathrm{vs}2})$, meaning $-Q(G, x_1, \beta)(x_{\mathrm{vu}1} - x_{\mathrm{vs}1}) < -Q(G, x_2, \beta)(x_{\mathrm{vu}2} - x_{\mathrm{vs}2})$. But by equation (18), this immediately implies that $P(G, x_1, \beta) \leq P(G, x_2, \beta)$, a clear contradiction, completing the proof. ■

Finally, we are equipped to prove Proposition 3.2.

Proof of Proposition 3.2: By definition of the uncertainty window, $a^* - \delta \leq \tilde{a}$. But then, by Lemma 1.3, for any $x = \mathcal{E}(G_{\min}, \beta)$ and $x' = \mathcal{E}(\tilde{G}, \beta)$,

$$x_{\mathrm{n}1} \geq x_{\mathrm{n}2}', x_{\mathrm{vu}1} \geq x_{\mathrm{vu}2}', \text{ and } x_{\mathrm{vs}1} \geq x_{\mathrm{vs}2}'.$$

This result allows us to apply Lemma 1.4, forcing that $P(G, x', \beta) \leq P(G, x, \beta)$. But since $\tilde{a}$ was arbitrary, this means $a^* - \delta$ must maximize accident probability, which is the desired result. ■

References

[1] B. T. Gould and P. N. Brown, "Information design for Vehicle-to-Vehicle communication," *Transportation Research Part C: Emerging Technologies*, vol. 150, p. 104084, May 2023.

[2] B. T. Gould and P. N. Brown, "Rationality and Behavior Feedback in a Model of Vehicle-to-Vehicle Communication," in *IEEE Conference on Decision and Control*, (Marina Bay Sands, Singapore), Dec. 2023. To be published.

[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] 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.

[6] 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.

[7] R. Steinberg and W. I. Zangwill, "The Prevalence of Braess' Paradox," *Transportation Science*, vol. 17, p. 301, Aug. 1983.

[8] 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.

[9] E. Bıyık, D. A. Lazar, R. Pedarsani, and D. Sadigh, "Incentivizing Efficient Equilibria in Traffic Networks With Mixed Autonomy," *IEEE TCNS*, vol. 8, pp. 1717–1729, Dec. 2021.

[10] Y. Zhu and K. Savla, "Information Design in Non-atomic Routing Games with Partial Participation: Computation and Properties," 2020.

[11] C. Zhao, B. Fu, and T. Wang, "Braess paradox and robustness of traffic networks under stochastic user equilibrium," *Transportation Research Part E: Logistics and Transportation Review*, vol. 61, pp. 135–141, Jan. 2014.

[12] 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.

[13] O. Massicot and C. Langbort, "Competitive Comparisons of Strategic Information Provision Policies in Network Routing Games," *IEEE Transactions on Control of Network Systems*, vol. 9, pp. 1589–1599, Dec. 2022.

[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] 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.

[17] A. Tversky and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[18] V. Hebbar and C. Langbort, "On The Role of Social Identity in the Market for (Mis)information," Apr. 2022. arXiv:2203.16660 [cs, eess].

[19] D. Kahneman and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.

[20] R. J. Aumann, "Rationality and Bounded Rationality," *Games and Economic Behavior*, May 1997.

[21] A. K. Chen, B. L. Ferguson, D. Shishika, M. Dorothy, J. R. Marden, G. J. Pappas, and V. Kumar, "Path Defense in Dynamic Defender-Attacker Blotto Games (dDAB) with Limited Information," in *2023 American Control Conference (ACC)*, pp. 447–453, May 2023.

[22] C. K. Verginis, "Funnel Control for Uncertain Nonlinear Systems via Zeroing Control Barrier Functions," *IEEE Control Systems Letters*, vol. 7, pp. 853–858, 2023.

[23] F. Sezer and C. Eksin, "Robust Social Welfare Maximization via Information Design in Linear-Quadratic-Gaussian Games," *IEEE Control Systems Letters*, vol. 7, pp. 3096–3101, 2023.

[24] B. L. Ferguson and J. R. Marden, "Robust Utility Design in Distributed Resource Allocation Problems with Defective Agents," *Dynamic Games and Applications*, Aug. 2022.

[25] O. Massicot and C. Langbort, "Almost-Bayesian Quadratic Persuasion (Extended Version)," Dec. 2022. arXiv:2212.13619 [cs].