Rationality and Behavior Feedback in a Model of Vehicle-to-Vehicle Communication

Brendan T. Gould and Philip N. Brown

***Abstract*— Vehicle-to-Vehicle (V2V) communication is intended to improve road safety through distributed information sharing; however, it is difficult to predict and optimize how human agents will respond to this information. In a *Bayesian game*, agents probabilistically adopt various types from a fixed, exogenous distribution. Agents in such models ostensibly perform Bayesian inference, which may not be a reasonable cognitive demand for most humans. To complicate matters, real-world information provided to agents is often implicitly dependent on agent behavior, meaning that the distribution of agent types is a function of the behavior of agents (i.e., the type distribution is *endogenous*). In this paper, we study an existing model of V2V communication, but relax it along two dimensions: first, we pose a behavior model which does not require human agents to perform Bayesian inference; second, an equilibrium model which avoids the challenging endogenous recursion. Surprisingly, we show that the simplified non-Bayesian behavior model yields the exact same equilibrium behavior as the original Bayesian model, which may lend credibility to Bayesian models. However, we also show that the endogenous type model is necessary to obtain certain informational paradoxes; these paradoxes do not appear in the simpler exogenous model. This suggests that standard Bayesian game models with fixed type distributions are not sufficient to express certain important phenomena.**

I. Introduction

As technology becomes more omnipresent in today's society, technological solutions are being developed for a broad range of applications. These applications increasingly include areas that have complex interactions with human society, such as the Internet of Things (IoT) and smart infrastructure concepts like vehicle-to-vehicle (V2V) communication. These interactions present a unique challenge to engineers, as prior work has shown that naively implemented solutions can unintentionally worsen the problems they were designed to solve [1], [2]. In particular, we consider the context of a traffic congestion game, where it is commonly known that selfish individual behavior is not socially optimal [3]–[5].

Prior work has considered various mechanisms to influence agents to choose socially optimal behaviors, such as financial incentives [6], [7]. *Bayesian persuasion* does this through information design, by strategically revealing or concealing information to change the posterior beliefs of these agents [8]–[10]. For example, one goal of V2V technology is to improve driver safety by broadcasting warning signals about road hazards. When a driver receives such a warning (even if it may be incorrect), they have a stronger belief that the road is unsafe, and are more likely to drive carefully. However, there are many limitations to information design.

First, Bayesian persuasion requires agents to be highly rational, but experiments have shown this may not be the case [11]–[17]. Furthermore, formal statistical statements, Bayes' Theorem in particular, are often misunderstood [18]. Assuming that human agents can quickly and accurately perform this calculation is likely unrealistic. Second, the optimal information sharing policy is often non-trivial. Prior work has shown that full information sharing may be suboptimal, or even worse than no information sharing [3], [9], [19]–[21]. Additionally, in many applications, the shared information is implicitly a function of agent behavior; adding complexity to the optimization problem.

Our work addresses both of these issues, using a congestion game where V2V cars can share information about accidents, introduced in [1]. We begin by posing a novel model of agent decision-making which does not require agents to perform Bayesian inference, re-framing the original Bayesian, incomplete information game to a non-Bayesian, imperfect information game. Surprisingly, we show that non-Bayesian equilibrium behavior exactly matches that of the original Bayesian model. This relaxes rationality expectations on human drivers, potentially improving the credibility of the Bayesian model. This is reminiscent of Harsanyi's classic work, but we believe that our characterization is not a direct consequence of [22] due to the addition of non-atomic agents and endogenous distribution of agent types.

Next, we investigate the relationship between model complexity and expressiveness by considering two classes of models. The simpler approach assumes that the probability of an accident (and thus the distribution of which agents receive which types of signals) is a constant model parameter, unaffected by emergent behavior (i.e. it is exogenous). Note that this is the standard approach in information design problems in the literature; road hazards or highway delay characteristics are almost universally assumed to be drawn from some fixed, known distribution [3]–[5], [23], [24].

However, our recent work [1] has considered models with an endogenous accident probability, expressing it as a function of equilibrium behavior. In this paper, we show that though the simpler exogenous models are easier to analyze, they are qualitatively different from the endogenous models. In particular, in endogenous models, a paradox can occur where information sharing worsens social cost. By contrast, this can never occur in the simpler (more standard) exogenous model, suggesting that the popular modeling framework of Bayesian games with fixed agent type distributions may sometimes be insufficient to express important phenomena.

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.

II. Model

A. General Setup

Our model consists of a non-atomic, unit mass of agents (drivers) interacting on a single road. On this road, traffic accidents either occur (A) or do not occur (¬A). Throughout, we use $\mathbb{P}(E)$ to represent the probability of event E.

Drivers are able to choose to drive carefully (C) or recklessly (R). Careful drivers consistently choose slower, safer driving behaviors while reckless drivers choose faster, riskier behaviors. Reckless drivers "pile on" to existing accidents and experience an expected cost of $r > 1$; however, careful drivers regret their caution if an accident is not present and feel 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

Each driver has type $\tau \in \mathcal{T}$, and strategies S_τ. A *strategy* $s \in S_\tau$ for a driver is a procedure to choose which action to play given the information they know. Let x_τ^s be the mass of drivers of type τ choosing strategy $s \in S_\tau$, and describe all agents' behavior by the tuple $x = (x_\tau^s)_{\tau\in\mathcal{T}, s\in S_\tau}$. Let x_τ be the total mass of drivers of type τ, so that $\sum_{s\in S_\tau} x_\tau^s = x_\tau$.

A fraction y of agents drive cars with V2V technology. These cars can autonomously detect accidents and broadcast signals about them. If an accident occurs, a "true positive" signal is broadcast with probability $t(y)$; otherwise, a "false positive" is broadcast with probability $f(y) < t(y)$. Any broadcasted signals are received by all V2V drivers.

Counter-intuitively, sharing perfect information about the environment can make parts or all of the population worse off. Accordingly, [1] introduced the parameter β to describe the information quality of V2V technology. In the event that a warning signal is broadcast (B), a V2V car may not always display a warning signal (S) to its driver; it will do so with probability $\beta = \mathbb{P}(\mathrm{S}|\mathrm{B}) \in [0, 1]$. Therefore, we have that

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

The system planner performs *information design* on the value of β to minimize accident probability and social cost.

An attractively simple approach to analysis is to let the probability of an accident be a constant that is unaffected by social behavior; this approach has been used previously in related literature [5], [7]. We call this case an *exogenous* accident probability, and refer to it as $\mathbb{P}(\mathrm{A}) = \bar{P} \in [0, 1]$.

However, we expect that reckless driving habits would lead to more frequent accidents. Therefore, we allow accident probability to be endogenously affected by driver behavior. Write $d \in [0, 1]$ for the overall fraction of reckless drivers, and $p(d)$ for the resulting probability of an accident. We assume that p is a continuous, strictly increasing function, so more reckless drivers cause accidents to be more likely.

Intuitively, strategies where drivers are more reckless should cause more accidents; to measure this, let $\rho(\tau, s, P)$ be the probability that a driver of type $\tau \in \mathcal{T}$ choosing strategy $s \in S_\tau$ is reckless given accident probability P.

The endogenous accident probability resulting from a behavior tuple x for driver types $\mathcal{T}$ is described implicitly as a solution to the recursive relationship:

$$P_{\mathcal{T}}(x) = p\left(\sum_{\tau\in\mathcal{T}}\sum_{s\in S_\tau} \rho(\tau, s, P_{\mathcal{T}}(x))x_\tau^s\right). \tag{2}$$

By [1, Proposition 2.1], this relation always has a solution.

We consider two classes of games, distinguished by either exogenous or endogenous accident probabilities. We write the former as the tuple $\bar{G} = (\beta, y, r, \bar{P})$, and the latter as $G = (\beta, y, r)$. For both types of games, equilibrium conditions come from the standard Nash idea: a behavior tuple x is an equilibrium of G if for any type τ and any strategy $s \in S_\tau$,

$$x_\tau^s > 0 \implies J_\tau(s; x) = \min_{s\in S_\tau} J_\tau(s; x). \tag{3}$$

That is, if an agent is choosing a strategy, its cost to them is minimal.

Finally, we define social cost as the expected cost incurred by the entire population:

$$\mathcal{J}_{\mathcal{T}}(x) = \sum_{\tau\in\mathcal{T}}\sum_{s\in S_\tau} J_\tau(s, x)x_\tau^s. \tag{4}$$

We abuse notation and write $P_{\mathcal{T}}(G)$ to mean $P_{\mathcal{T}}(x)$ and $\mathcal{J}_{\mathcal{T}}(G)$ (or $\mathcal{J}_{\mathcal{T}}(\bar{G})$) to mean $\mathcal{J}_{\mathcal{T}}(x)$ where x is an equilibrium of the game G (or $\bar{G}$).[1]

B. Driver Decision Models

We consider two interpretations for the effect of V2V technology on the behavior of V2V drivers.

1) Bayesian Agents: The traditional approach for a Bayesian game is studied in Section IV (generalizing the specific case considered in [1]). There are three types of agents: non-V2V drivers, V2V drivers who receive a signal, and V2V drivers who do not receive a signal. Non-V2V drivers must use the prior probability of an accident to calculate their expected costs, but V2V drivers can use the posterior probability after the signal realization.

We call this the *Bayesian* set of types, and write it $\mathcal{T}_B = \{\mathrm{n}, \mathrm{vu}, \mathrm{vs}\}$ for non-V2V, unsignaled V2V, and signaled V2V drivers respectively. It can be quickly seen that the mass of drivers in each group is $x_{\mathrm{n}} = 1 - y$, $x_{\mathrm{vu}} = \mathbb{P}(\neg\mathrm{S})y$, and $x_{\mathrm{vs}} = \mathbb{P}(\mathrm{S})y$. Each type has the strategies $S_{\mathrm{n}} = S_{\mathrm{vu}} = S_{\mathrm{vs}} = \{\mathrm{C}, \mathrm{R}\}$, with the accompanying cost functions:

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

$$J_{\mathrm{vu}}(s; x) = \begin{cases} 1 - \mathbb{P}(\mathrm{A}|\neg\mathrm{S}) & \text{if } s = \mathrm{C}, \\ r\mathbb{P}(\mathrm{A}|\neg\mathrm{S}) & \text{if } s = \mathrm{R}, \end{cases} \tag{6}$$

$$J_{\mathrm{vs}}(s; x) = \begin{cases} 1 - \mathbb{P}(\mathrm{A}|\mathrm{S}) & \text{if } s = \mathrm{C}, \\ r\mathbb{P}(\mathrm{A}|\mathrm{S}) & \text{if } s = \mathrm{R}. \end{cases} \tag{7}$$

[1] It can be shown that any two equilibria of the same game have equal accident probability and social cost. See the proof of Lemma 4.2 for a formal treatment of this idea in exogenous games, and Lemma 5.4 for the same in endogenous games.

2) Non-Bayesian Agents: Alternatively, Section V models V2V drivers who do not perform Bayesian updates. There are non-V2V and V2V drivers ($\mathcal{T}_I = \{\mathrm{n}, \mathrm{v}\}$), with $x_\mathrm{n} = 1 - y$ and $x_\mathrm{v} = y$. Instead of calculating a posterior, V2V drivers choose between trusting (T) the signal completely (assuming that the presence of a signal implies the existence of an accident, and vice versa), or ignoring it completely, using the prior to calculate an "assumed" cost. Drivers have strategies $S_\mathrm{n} = \{\mathrm{C}, \mathrm{R}\}$ and $S_\mathrm{v} = \{\mathrm{T}, \mathrm{C}, \mathrm{R}\}$, with cost functions:

$$J_\mathrm{n}(s; x) = \begin{cases} 1 - \mathbb{P}(\mathrm{A}) & \text{if } s = \mathrm{C}, \\ r\mathbb{P}(\mathrm{A}) & \text{if } s = \mathrm{R}. \end{cases} \tag{8}$$

$$J_\mathrm{v}(s; x) = \begin{cases} \mathbb{P}(\neg\mathrm{A} \cap \mathrm{S}) + r\mathbb{P}(\mathrm{A} \cap \neg\mathrm{S}) & \text{if } s = \mathrm{T}, \\ 1 - \mathbb{P}(\mathrm{A}) & \text{if } s = \mathrm{C}, \\ r\mathbb{P}(\mathrm{A}) & \text{if } s = \mathrm{R}. \end{cases} \tag{9}$$

The timeline of non-Bayesian games is shown in Figure 1. Note its differences from [1, Figure 1] (describing the behavior of Bayesian agents). Non-Bayesian V2V agents must choose a strategy *before* the signal realization, and can only use the signal if they commit to trusting it completely.

III. Main Results

Our first result is that Bayesian games of *incomplete* information can be reinterpreted as *imperfect* information games, reminiscent of [22]. This allows us to remove the heavy cognitive burden of Bayes' Theorem from agent calculations, producing a more credible model of human behavior.

Theorem 3.1: Let $G = (\beta, y, r)$ be a game with endogenous accident probability, and $\bar{G} = (\beta, y, r, \bar{P})$ be an exogenous game. Then,

$$\mathcal{J}_{\mathcal{T}_B}(\bar{G}) = \mathcal{J}_{\mathcal{T}_I}(\bar{G}), \tag{10}$$

and

$$P_{\mathcal{T}_B}(G) = P_{\mathcal{T}_I}(G), \tag{11}$$

$$\mathcal{J}_{\mathcal{T}_B}(G) = \mathcal{J}_{\mathcal{T}_I}(G). \tag{12}$$

Proof: Equation (10), concerning the social cost of exogenous games, is proven by Lemma 5.3 in section V-B. Equations (11) and (12), analogous statements for endogenous games, are proven by Lemma 5.6 and Lemma 5.7 respectively in section V-C. ■

Theorem 3.1 shows that any outcome of a Bayesian signaling game can be captured by a decision model where drivers completely trust or ignore the signal's information. This gives us freedom to analyze assuming Bayesian or non-Bayesian drivers, whichever is more convenient. Additionally, since the social costs induced by both models are equivalent, we now write simply $\mathcal{J}(G)$ for the social cost of G, where G can be interpreted as either a Bayesian or non-Bayesian game.

Our second result concerns the modeling of accidents. Using an endogenous accident probability complicates model analysis, so it is natural to ask whether the same characteristics can be captured by an exogenous model. However, this is not the case. In particular, equilibrium social cost can paradoxically be increasing with information quality [1], but only when accident probability is endogenous.

Theorem 3.2: Fix $y \in [0, 1]$, $r > 1$, and $\bar{P} \in [p(0), p(1)]$. Consider $\beta_1, \beta_2 \in [0, 1]$ with $\beta_1 < \beta_2$. Let $\bar{G}_1 = (\beta_1, y, r, \bar{P})$ and $\bar{G}_2 = (\beta_2, y, r, \bar{P})$ be signaling games with exogenous accident probability. Then, it is always true that

$$\mathcal{J}(\bar{G}_1) \geq \mathcal{J}(\bar{G}_2). \tag{13}$$

Similarly, Let $G_1 = (\beta_1, y, r)$ and $G_2 = (\beta_2, y, r)$ be signaling games with endogenous accident probability. Counter-intuitively, there exist parameter combinations where

$$\mathcal{J}(G_1) < \mathcal{J}(G_2). \tag{14}$$

Proof: In games between Bayesian drivers, Lemma 4.2 proves (13), and Lemma 4.3 shows that (14) can be satisfied. Then, by Theorem 3.1, these results also hold for games between non-Bayesian drivers. This completes the proof. ■

IV. Bayesian Agents

We first consider a scenario where V2V drivers perform Bayesian updates on the probability of an accident using the warning signal realization. Analysis proceeds as follows: Lemma 4.1 establishes necessary conditions of any equilibrium (with either exogenous or endogenous accident probability) by describing behavior when drivers have a strict preference for one strategy. This result is sufficient to prove (13), which we do in section IV-B, Lemma 4.2. Finally, in section IV-C, Lemma 4.3 gives the counter-intuitive result that social cost may be increasing with information quality, but only if accident probability is endogenous.

A. Necessary Equilibrium Conditions

We divide V2V drivers by whether or not they have seen a warning signal, giving the set of driver types $\mathcal{T}_B = \{\mathrm{n}, \mathrm{vu}, \mathrm{vs}\}$. Using the calculated posterior probability, V2V drivers choose between the pure strategies of always driving carefully or always driving recklessly.

The following thresholds are useful in our analysis:

$$P_\mathrm{vs} := \frac{f(y)}{rt(y) + f(y)}, \tag{15}$$

$$P_\mathrm{n} := \frac{1}{1 + r}, \tag{16}$$

$$P_\mathrm{vu} := \frac{1 - \beta f(y)}{1 + r(1 - \beta t(y)) - \beta f(y)}, \tag{17}$$

where it holds that:

$$P_\mathrm{vs} < P_\mathrm{n} \leq P_\mathrm{vu}. \tag{18}$$

Note that (by Bayes' Theorem and (5)-(7)):

$$J_\mathrm{vu}(\mathrm{R}; x) \;\substack{\leq \\ = \\ >}\; J_\mathrm{vu}(\mathrm{C}; x) \iff \mathbb{P}(\mathrm{A}) \;\substack{\leq \\ = \\ >}\; P_\mathrm{vu}, \tag{19}$$

$$J_\mathrm{vs}(\mathrm{R}; x) \;\substack{\leq \\ = \\ >}\; J_\mathrm{vs}(\mathrm{C}; x) \iff \mathbb{P}(\mathrm{A}) \;\substack{\leq \\ = \\ >}\; P_\mathrm{vs} \tag{20}$$

We use this notation to mean that an ordering on the first expressions is equivalent to the same ordering on the later expressions. Equality and both inequalities are preserved.

Using (5)-(7), we are now equipped to state the necessary conditions of any equilibrium:

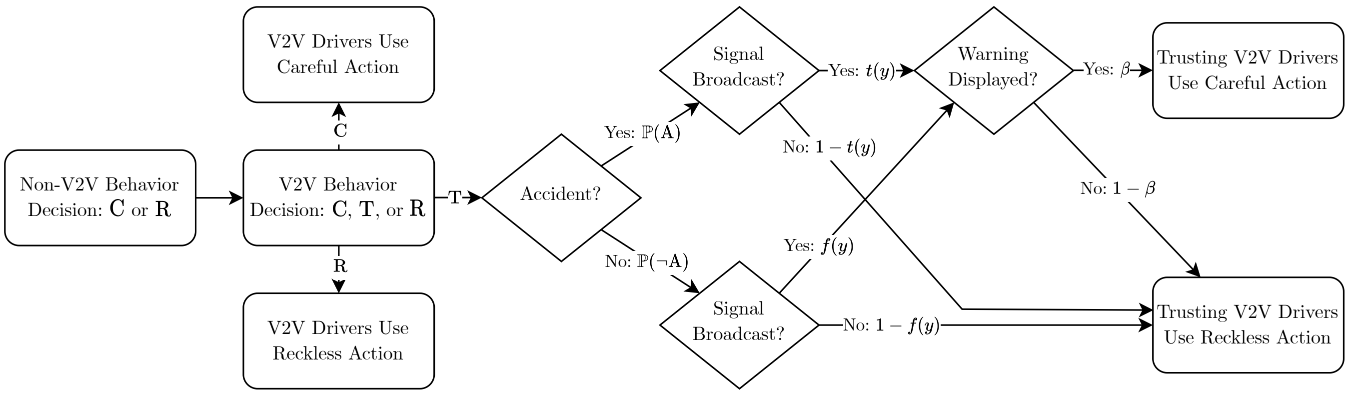

Fig. 1. Timeline of decisions and events for non-Bayesian drivers.

Lemma 4.1: If x is a Bayesian signaling equilibrium, then

$$x_{\rm n}^{\rm R} = \begin{cases} 0 & \text{if } \mathbb{P}({\rm A}) > P_{\rm n}, \\ 1-y & \text{if } \mathbb{P}({\rm A}) < P_{\rm n}, \end{cases} \tag{21}$$

$$x_{\rm vu}^{\rm R} = \begin{cases} 0 & \text{if } \mathbb{P}({\rm A}) > P_{\rm vu}, \\ \mathbb{P}(\neg{\rm S})y & \text{if } \mathbb{P}({\rm A}) < P_{\rm vu}, \end{cases} \tag{22}$$

$$x_{\rm vs}^{\rm R} = \begin{cases} 0 & \text{if } \mathbb{P}({\rm A}) > P_{\rm vs}, \\ \mathbb{P}({\rm S})y & \text{if } \mathbb{P}({\rm A}) < P_{\rm vs}. \end{cases} \tag{23}$$

Proof: These all follow from (3), (5)-(7), and (19)-(20). Assume by way of contradiction that $\mathbb{P}({\rm A}) > P_{\rm vu}$, but $x_{\rm vu}^{\rm R} > 0$. By (19), $\mathbb{P}({\rm A}|\neg{\rm S}) > \frac{1}{1+r}$. Then, (6) gives that

$$J_{\rm vu}({\rm R}, x) = r\mathbb{P}({\rm A}|\neg{\rm S}) > 1 - \mathbb{P}({\rm A}|\neg{\rm S}) = J_{\rm vu}({\rm C}, x).$$

But this clearly contradicts (3), since $x_{\rm vu}^{\rm R} > 0$. The proof of the remaining cases is similar. ∎

B. Exogenous Crash Probability

Let accident probability be an exogenous constant $\bar{P}$. Then, social cost is decreasing with information quality β.

Lemma 4.2: Let $\beta_1 < \beta_2 \in [0,1]$, and $\bar{G}_1 = (\beta_1, y, r, \bar{P})$ and $\bar{G}_2 = (\beta_2, y, r, \bar{P})$ between exogenous, Bayesian signaling games. Then, $\mathcal{J}_{\mathcal{T}_B}(\bar{G}_1) \geq \mathcal{J}_{\mathcal{T}_B}(\bar{G}_2)$.

Proof: Let $P_{\rm vu1}$, $P_{\rm vu2}$ be the threshold $P_{\rm vu}$ for β_1 and β_2, respectively. Note that $P_{\rm vu1} < P_{\rm vu2}$ since $f(y) < t(y)$.

First, assume that $\bar{P} > P_{\rm vu2} > P_{\rm vu1}$. By (18) and Lemma 4.1, $\bar{P} > P_{\rm n} > P_{\rm vs}$ so $x_{\rm n}^{\rm R} = 0$, $x_{\rm vu}^{\rm R} = 0$, and $x_{\rm vs}^{\rm R} = 0$ for any equilibrium of $\bar{G}_1$ or $\bar{G}_2$. Thus, (4)-(7) give

$$\mathcal{J}_{\mathcal{T}_B}(\bar{G}_1) = 1 - \bar{P} \geq \mathcal{J}_{\mathcal{T}_B}(\bar{G}_2),$$

and we are finished in this case.

Now, assume $\bar{P} = P_{\rm vu2}$ (implying $\bar{P} > P_{\rm vu1}$). By (18), $P_{\rm vs} < P_{\rm n} < \bar{P}$, so by Lemma 4.1, $x_{\rm n}^{\rm R} = 0$ and $x_{\rm vs}^{\rm R} = 0$ in any equilibrium of $\bar{G}_1$ or $\bar{G}_2$. Let x_1 be an equilibrium of $\bar{G}_1$ and x_2 one of $\bar{G}_2$. By the above, $\mathcal{J}_{\mathcal{T}_B}(\bar{G}_1) = 1 - \bar{P}$. By (7), $J_{\rm vu}({\rm C}, x_2) = J_{\rm vu}({\rm R}, x_2)$. Thus, $\sum_{s\in S_{\rm vu}} J_{\rm vu}(s, x_2) x_{\rm vu2}^s = J_{\rm vu}({\rm R}, x)\mathbb{P}(\neg{\rm S})y$, so (4) simplifies to

$$\mathcal{J}_{\mathcal{T}_B}(\bar{G}_1) = 1 - \bar{P} \geq 1 - P_{\rm vu2} = \mathcal{J}_{\mathcal{T}_B}(\bar{G}_2),$$

and we are again finished. A similar technique is used when $P_{\rm vu1} = \bar{P} < P_{\rm vu2}$, $P_{\rm n} < \bar{P} < P_{\rm vu1}$, $\bar{P} = P_{\rm n}$, $P_{\rm vs} < \bar{P} < P_{\rm n}$, $\bar{P} = P_{\rm vs}$, or $\bar{P} < P_{\rm vs}$, completing the proof. ∎

Lemma 4.2 shows that higher quality information will never increase social cost for exogenous accident probabilities, which is perhaps the expected result. However, this is not necessarily true for endogenous accident probabilities.

C. Endogenous Crash Probability

Now consider Bayesian signaling games with endogenous accident probability. Recall the agent types: non-V2V drivers, unsignaled V2V drivers, and signaled V2V drivers. Drivers of each type $\tau \in \mathcal{T}_B$ have only the pure strategies of driving carefully or recklessly, where $\rho(\tau, {\rm R}, P_{\mathcal{T}_B}(x)) = 1$ and $\rho(\tau, {\rm C}, P_{\mathcal{T}_B}(x)) = 0$. Therefore, (2) simplifies to

$$\mathbb{P}({\rm A}) = P_{\mathcal{T}_B}(x) = p\left(x_{\rm n}^{\rm R} + x_{\rm vu}^{\rm R} + x_{\rm vs}^{\rm R}\right). \tag{24}$$

In this case, the model is identical to the one described in [1]; we simply reference the previous result.

Lemma 4.3 ([1, Proposition 3.8]): There exist games between Bayesian drivers with endogenous accident probability such that social cost at equilibrium is increasing with β.

V. Non-Bayesian Agents

We now discuss games with agents who do not perform explicit Bayesian updates. Analysis generally follows the same path as section IV. Lemma 5.1 describes the behavior of agents who strictly prefer one strategy, giving necessary conditions for an equilibrium of any such game.

We then characterize the equilibria of these games. Lemma 5.2 describes social cost with exogenous accident probability, and Lemmas 5.4 and 5.5 relate an endogenous accident probability to game parameters. Finally, we compare these quantities to those induced by Bayesian agents, and show they are identical (Lemmas 5.3, 5.6, and 5.7).

A. Necessary Equilibrium Conditions

To model the behavior of non-Bayesian agents, we use the set of driver types $\mathcal{T}_I = \{{\rm n}, {\rm v}\}$, representing non-V2V drivers and V2V drivers, respectively. All drivers have the pure strategies of always being careful or always being reckless (using the prior probability of an accident to compute expected cost), but V2V drivers additionally have the option to fully "trust" the signal, believing the presence of a warning light implies the existence of an accident and the inverse.

Again, we use thresholds to describe behavior. By (9),

$$J_{\mathrm{v}}(\mathrm{T};x) \lesseqgtr J_{\mathrm{v}}(\mathrm{C};x) \iff \mathbb{P}(\mathrm{A}) \lesseqgtr P_{\mathrm{vu}}, \quad (25)$$

$$J_{\mathrm{v}}(\mathrm{R};x) \lesseqgtr J_{\mathrm{v}}(\mathrm{T};x) \iff \mathbb{P}(\mathrm{A}) \lesseqgtr P_{\mathrm{vs}} \quad (26)$$

Crucially, Bayes' Theorem was never necessary for these calculations, yet they give the *same* behavior thresholds as (19) and (20). This gives a compelling argument that Bayesian and non-Bayesian behaviors are equivalent. Since both models are making the same decisions, it is unsurprising that the resulting equilibria are identical. We now make an analogous statement to Lemma 4.1 using (25) and (26).

Lemma 5.1: For any equilibrium of x of a signaling game,

$$x_{\mathrm{n}}^{\mathrm{R}} = \begin{cases} 0 & \text{if } \mathbb{P}(\mathrm{A}) > P_{\mathrm{n}}, \\ 1-y & \text{if } \mathbb{P}(\mathrm{A}) < P_{\mathrm{n}}, \end{cases} \quad (27)$$

$$x_{\mathrm{v}}^{\mathrm{R}} = \begin{cases} 0 & \text{if } \mathbb{P}(\mathrm{A}) > P_{\mathrm{vs}}, \\ y & \text{if } \mathbb{P}(\mathrm{A}) < P_{\mathrm{vs}}, \end{cases} \quad (28)$$

$$x_{\mathrm{v}}^{\mathrm{T}} = \begin{cases} 0 & \text{if } \mathbb{P}(\mathrm{A}) < P_{\mathrm{vs}}, \\ y & \text{if } P_{\mathrm{vs}} < \mathbb{P}(\mathrm{A}) < P_{\mathrm{vu}}, \\ 0 & \text{if } \mathbb{P}(\mathrm{A}) > P_{\mathrm{vu}} \end{cases} \quad (29)$$

Proof: This follows directly from (3), (8)-(9), and (25)-(26). Assume by way of contradiction that $P_{\mathrm{vs}} < \mathbb{P}(\mathrm{A}) < P_{\mathrm{vu}}$, but $x_{\mathrm{v}}^{\mathrm{T}} < y$. By (25), $J_{\mathrm{v}}(\mathrm{T};x) < J_{\mathrm{v}}(\mathrm{C};x)$, and by (26), $J_{\mathrm{v}}(\mathrm{T};x) < J_{\mathrm{v}}(\mathrm{R};x)$. Therefore, by (3), $x_{\mathrm{v}}^{\mathrm{C}} = x_{\mathrm{v}}^{\mathrm{R}} = 0$. But since $x_{\mathrm{v}} = y$, this implies $x_{\mathrm{v}}^{\mathrm{T}} = y$, a contradiction. The proof of the remaining cases is similar. ■

B. Exogenous Crash Probability

We again begin by assuming an exogenous accident probability $\bar{P}$. Again, we see that models with this assumption are qualitatively different from those assuming an interdependence between driver behavior and accident probability.

Lemma 5.2: Let $\beta_1 < \beta_2 \in [0,1]$, and $\bar{G}_1 = (\beta_1, y, r, \bar{P})$ and $\bar{G}_2 = (\beta_2, y, r, \bar{P})$ be exogenous, non-Bayesian signaling games. Then, $\mathcal{J}_{\mathcal{T}_I}(\bar{G}_1) \geq \mathcal{J}_{\mathcal{T}_I}(\bar{G}_2)$.

Proof: This can be shown using a technique identical to that of the proof of Lemma 4.2. ■

That is, social cost is non-increasing with information quality β for these types of games. Furthermore, that social cost is unaffected by the change in agent decision model:

Lemma 5.3: For any game $\bar{G} = (\beta, y, r, \bar{P})$ with exogenous accident probability $\bar{P}$,

$$\mathcal{J}_{\mathcal{T}_B}(\bar{G}) = \mathcal{J}_{\mathcal{T}_I}(\bar{G}). \quad (30)$$

Proof: We again prove this in cases. First, assume that $P_{\mathrm{n}} < \bar{P} < P_{\mathrm{vu}}$. By Lemma 4.1, $x_{\mathrm{n}}^{\mathrm{R}} = 0$, $x_{\mathrm{vu}}^{\mathrm{R}} = \mathbb{P}(\neg\mathrm{S})y$, and $x_{\mathrm{vs}}^{\mathrm{R}} = 0$. Similarly by Lemma 5.1, $x_{\mathrm{n}}^{\mathrm{R}} = 0$, $x_{\mathrm{v}}^{\mathrm{R}} = 0$, and $x_{\mathrm{v}}^{\mathrm{T}} = y$. Then, (4) simplifies to give

$$\mathcal{J}_{\mathcal{T}_B}(\bar{G}) = \mathcal{J}_{\mathcal{T}_I}(\bar{G}) = (1-\bar{P})(1-y) + r\bar{P}(1-\beta t(y))y + ((1-\bar{P})\beta f(y))y,$$

completing this case. The same technique gives the desired result if $\bar{P} < P_{\mathrm{vs}}$, $P_{\mathrm{vs}} < \bar{P} < P_{\mathrm{n}}$, or $P_{\mathrm{vu}} < \bar{P}$.

Now, assume $\bar{P} = P_{\mathrm{n}}$. By Lemma 4.1, $x_{\mathrm{vs}}^{\mathrm{R}} = 0$, and by Lemma 5.1, $x_{\mathrm{v}}^{\mathrm{R}} = 0$. By (5) (or equivalently (8)), $J_{\mathrm{n}}(\mathrm{R};x) = J_{\mathrm{n}}(\mathrm{C};x)$, so $\sum_{s\in S_{\mathrm{n}}} J_{\mathrm{n}}(s,x)x_{\mathrm{n}}^{s} = J_{\mathrm{n}}(\mathrm{R},x)(1-y)$.

If $\bar{P} = P_{\mathrm{vu}}$, then a very similar argument implies that $\sum_{s\in S_{\mathrm{vu}}} J_{\mathrm{vu}}(s,x)x_{\mathrm{vu}}^{s} = J_{\mathrm{vu}}(\mathrm{R},x)(1-\mathbb{P}(\mathrm{S}))y$, and $\sum_{s\in S_{\mathrm{v}}} J_{\mathrm{v}}(s,x)x_{\mathrm{v}}^{s} = J_{\mathrm{v}}(\mathrm{T},x)y$. Otherwise, by (18), $P_{\mathrm{vs}} < \bar{P} < P_{\mathrm{vu}}$, meaning $x_{\mathrm{vu}}^{\mathrm{R}} = (1-\mathbb{P}(\mathrm{S}))y$ by Lemma 4.1 and $x_{\mathrm{v}}^{\mathrm{T}} = y$ by Lemma 5.1. In any case, (4) again simplifies to

$$\mathcal{J}_{\mathcal{T}_B}(\bar{G}) = \mathcal{J}_{\mathcal{T}_I}(\bar{G}) = (1-\bar{P})(1-y) + r\bar{P}(1-\beta t(y))y + ((1-\bar{P})\beta f(y))y,$$

the desired result. This idea suffices in every case. ■

C. Endogenous Crash Probability

Finally, we use the non-Bayesian decision model with an endogenous accident probability. This creates the strategy spaces $S_{\mathrm{n}} = \{\mathrm{C}, \mathrm{R}\}$ and $S_{\mathrm{v}} = \{\mathrm{T}, \mathrm{C}, \mathrm{R}\}$, where $\rho(\tau, \mathrm{R}, P_{\mathcal{T}_I}(x)) = 1$ and $\rho(\tau, \mathrm{C}, P_{\mathcal{T}_I}(x)) = 0$ for each $\tau \in \mathcal{T}_I$, and $\rho(x_{\mathrm{v}}, \mathrm{T}, P_{\mathcal{T}_I}(x)) = \mathbb{P}(\neg\mathrm{S})$. Then, (2) gives

$$\mathbb{P}(\mathrm{A}) = P_{\mathcal{T}_I}(x) = p(x_{\mathrm{n}}^{\mathrm{R}} + x_{\mathrm{v}}^{\mathrm{R}} + \mathbb{P}(\neg\mathrm{S})x_{\mathrm{v}}^{\mathrm{T}}). \quad (31)$$

Recall that Lemma 5.1 gives equilibrium behavior when agents strictly prefer one strategy. We now complete this result for agents who are indifferent between two strategies.

Lemma 5.4: For any game $G = (\beta, y, r)$, a non-Bayesian behavior tuple x is a signaling equilibrium if it satisfies Lemma 5.1 and the following hold:

$$\mathbb{P}(\mathrm{A}) = P_{\mathrm{n}} \implies x_{\mathrm{n}}^{\mathrm{R}} = p^{-1}(P_{\mathrm{n}}) - \mathbb{P}(\neg\mathrm{S})y \quad (32)$$

$$\mathbb{P}(\mathrm{A}) = P_{\mathrm{vs}} \implies x_{\mathrm{v}}^{\mathrm{R}} = y \quad (33)$$

$$\mathbb{P}(\mathrm{A}) = P_{\mathrm{vs}} \implies x_{\mathrm{v}}^{\mathrm{T}} = 0 \quad (34)$$

$$\mathbb{P}(\mathrm{A}) = P_{\mathrm{vu}} \implies x_{\mathrm{v}}^{\mathrm{T}} = \frac{p^{-1}(P_{\mathrm{vu}})}{\mathbb{P}(\neg\mathrm{S})} \quad (35)$$

Furthermore, the behavior tuple x satisfying the above conditions is *essentially unique* for G, that is, any equilibrium x' satisfies the above conditions or has $P_{\mathcal{T}_I}(x') = P_{\mathcal{T}_I}(x)$.

For brevity, the proof is omitted, but largely follows from algebra and (8)–(9). A complete proof can be found in [25].

Parameter space is partitioned by the following thresholds:

$$E_{1U} := p(0), \quad (36)$$

$$E_{2U} := p(y - (P_{\mathrm{vu}}(t(y) - f(y))\beta + f(y)\beta)y), \quad (37)$$

$$E_{3U} := p(y - (P_{\mathrm{n}}(t(y) - f(y))\beta + f(y)\beta)y), \quad (38)$$

$$E_{4U} := p(1 - (P_{\mathrm{n}}(t(y) - f(y))\beta + f(y)\beta)y), \quad (39)$$

$$E_{5U} := p(1 - (P_{\mathrm{vs}}(t(y) - f(y))\beta + f(y)\beta)y), \quad (40)$$

$$E_{6U} := p(1), \quad (41)$$

and partitions:

$$E_1 := \{(\beta, y, r) : P_{\mathrm{vu}} < E_{1U}\}, \quad (42)$$

$$E_2 := \{(\beta, y, r) : E_{1U} \leq P_{\mathrm{vu}} \leq E_{2U}, \quad (43)$$

$$E_3 := \{(\beta, y, r) : E_{2U} < P_{\mathrm{vu}} \wedge P_{\mathrm{n}} < E_{3U}, \quad (44)$$

$$E_4 := \{(\beta, y, r) : E_{3U} \leq P_{\mathrm{n}} \leq E_{4U}, \quad (45)$$

$$E_5 := \{(\beta, y, r) : E_{4U} < P_{\mathrm{n}} \wedge P_{\mathrm{vs}} < E_{5U}, \quad (46)$$

$$E_6 := \{(\beta, y, r) : E_{5U} \leq P_{\mathrm{vs}} \leq E_{6U})\}, \quad (47)$$

$$E_7 := \{(\beta, y, r) : E_{6U} < P_{\mathrm{vs}}\}. \quad (48)$$

These partitions allow us to describe equilibrium accident probability with finer granularity.

Lemma 5.5: For any non-Bayesian game $G = (\beta, y, r)$, $G \in \cup_{i=1}^{7} E_i$, and

$$G \in E_1 \implies P_{\mathcal{T}_I}(G) = p(0) \tag{49}$$
$$G \in E_2 \implies P_{\mathcal{T}_I}(G) = P_{\mathrm{vu}} \tag{50}$$
$$G \in E_3 \implies P_{\mathrm{n}} < P_{\mathcal{T}_I}(G) < P_{\mathrm{vu}} \tag{51}$$
$$G \in E_4 \implies P_{\mathcal{T}_I}(G) = P_{\mathrm{n}} \tag{52}$$
$$G \in E_5 \implies P_{\mathrm{vs}} < P_{\mathcal{T}_I}(G) < P_{\mathrm{n}} \tag{53}$$
$$G \in E_6 \implies P_{\mathcal{T}_I}(G) = P_{\mathrm{vs}} \tag{54}$$
$$G \in E_7 \implies P_{\mathcal{T}_I}(G) = p(1) \tag{55}$$

This is proved by a technique similar to that of Lemma 4.1 in [1]. The only difference comes from using the endogenous accident probability defined by (31) rather than (24). For brevity, the full proof is omitted, but can be found in [25].

We can now characterize any game G; Lemma 5.5 gives a restriction on equilibrium accident probability, and Lemmas 5.1 and 5.4 yield driver behavior under this restriction. This behavior shows that Bayesian and non-Bayesian drivers induce the same accident probability and social cost.

Lemma 5.6: For any endogenous game $G = (\beta, y, r)$,

$$P_{\mathcal{T}_B}(G) = P_{\mathcal{T}_I}(G). \tag{56}$$

Proof: By Lemma 5.5, G belongs to one of the equilibrium families E_1–E_7, so we can show this by cases.

Equilibrium accident probability is restricted as a function of game parameters by Lemma 5.5 for non-Bayesian agents and by [1, Lemma 4.1] for Bayesian agents. This gives the desired result immediately unless $G \in E_3$ or $G \in E_5$. In the first case, note that Bayesian agents choose $x_{\mathrm{n}}^{\mathrm{R}} = 0$, $x_{\mathrm{vs}}^{\mathrm{R}} = 0$, and $x_{\mathrm{vu}}^{\mathrm{R}} = \mathbb{P}(\neg \mathrm{S})y$ as a consequence of [1, Lemma 4.2]. Similarly, non-Bayesian agents choose $x_{\mathrm{n}}^{\mathrm{R}} = 0$, $x_{\mathrm{v}}^{\mathrm{R}} = 0$, and $x_{\mathrm{v}}^{\mathrm{T}} = y$ by Lemma 5.5. Then, by (24) and (31),

$$P_{\mathcal{T}_B}(G) = p(\mathbb{P}(\neg \mathrm{S})y) = P_{\mathcal{T}_I}(G),$$

which is the desired result.

The proof is similar if $G \in E_5$, so we are finished. ■

Lemma 5.7: For any endogenous game $G = (\beta, y, r)$,

$$\mathcal{J}_{\mathcal{T}_B}(G) = \mathcal{J}_{\mathcal{T}_I}(G). \tag{57}$$

Proof: Equilibrium behavior is given by [1, Lemma 4.2]) for Bayesian agents, and by 5.5 for non-Bayesian agents. Expanding (4) and applying algebra with this behavior gives the desired result. ■

VI. Conclusion

This work considered a class of models describing how human agents respond to road hazard information sharing. We first showed that games of this kind with incomplete information can be equivalently interpreted as imperfect information games, removing an assumption on human rationality. We used this fact to prove that models with an endogenous accident probability can describe scenarios that those with an exogenous accident probability cannot; namely, that social cost can be increasing with information quality. Future work could consider additional alternative descriptions of human behavior, and possibly use a heterogeneous population.

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] 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] 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.

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

[5] 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.

[6] B. L. Ferguson, P. N. Brown, and J. R. Marden, “The Effectiveness of Subsidies and Tolls in Congestion Games,” *IEEE Transactions on Automatic Control*, vol. 67, pp. 2729–2742, jun 2022.

[7] 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.

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

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

[10] S. Shah, S. Amin, and P. Jaillet, “Optimal Information Provision for Strategic Hybrid Workers,” May 2022. arXiv:2205.02732 [cs].

[11] K. Basu, “The Traveler’s Dilemma: Paradoxes of Rationality in Game Theory,” *The American Economic Review*, vol. 84, no. 2, pp. 391–395, 1994.

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

[13] R. J. Aumann, “Rationality and Bounded Rationality,” *Games and Economic Behavior*, May 1997.

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

[15] G. Gigerenzer and H. Brighton, “Homo Heuristicus: Why Biased Minds Make Better Inferences,” *Topics in Cognitive Science*, vol. 1, no. 1, pp. 107–143, 2009.

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

[17] K. D. Edwards, “Prospect theory: A literature review,” *International Review of Financial Analysis*, vol. 5, pp. 19–38, Jan. 1996.

[18] G. Gigerenzer, W. Gaissmaier, E. Kurz-Milcke, L. M. Schwartz, and S. Woloshin, “Helping Doctors and Patients Make Sense of Health Statistics,” *Psychological Science in the Public Interest*, vol. 8, pp. 53–96, Nov. 2007.

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

[20] 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.

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

[22] J. C. Harsanyi, “Games with Incomplete Information Played by "Bayesian" Players, I-Iii,” *Management Science*, vol. 14, pp. 159–182, Nov. 1967.

[23] 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, pp. 613–624, jun 2022.

[24] B. L. Ferguson, P. N. Brown, and J. R. Marden, “Avoiding Unintended Consequences: How Incentives Aid Information Provisioning in Bayesian Congestion Games,” in *61st IEEE Conference on Decision and Control*, apr 2022.

[25] B. Gould and P. Brown, “Rationality and Behavior Feedback in a Model of Vehicle-to-Vehicle Communication,” *arXiv:2307.03382 [cs]*, July 2023.