Pluralistic Ignorance in the Bystander Effect
Theory and solutions

Jorik Melsen, Bram Wiggers, Dick Ameln

May 2019

1 Introduction

Governments around the world have conducted countless campaigns to combat the problem of drunk driving. Nevertheless, unfortunately drunk people still step into their cars after leaving a party or a bar, as is clear from the large number of DUI charges and thousands of fatalities resulting from alcohol related traffic accidents each year (see [1]). Sometimes people stop a drunk person when they notice he or she is planning to drive home. However, this is not always the case. Why do people not always intervene when they see these inebriated people fumbling with their car keys when leaving a public event to get to their car? This is an example of **the bystander effect**, which is the tendency for people not to intervene when witnessing problematic scenarios in a group, which they would have done would they have been alone.

In general the process of intervention in a problematic scenario occurs in three steps and at each step a different cause of the bystander effect can take place [2]. The first step consists of a bystander noticing an event. Clearly if a bystander doesn't notice an event, the bystander will not intervene. The second step occurs when the bystander tries to interpret the event he just witnessed. Some events can only be interpreted one way, for example a catastrophic event like a plane crash. Other events can be much more ambiguous. Imagine a crying toddler in a supermarket. A bystander is likely to think that the child is simply trowing a tantrum because the child's parents don't want to buy his favourite cereal. Another thing that might have happened is that one of the parents hit the child out of sight of the bystander. When a bystander is not certain whether an event is problematic or not he will try to get more information by observing other bystanders, gathering **social proof**. If the other bystanders do not appear to be worried, the bystander will think that the event is not problematic and will thus not intervene. This can lead to a situation in which all bystanders think an event is problematic, but are uncertain about it. Subsequently, the bystanders all observe each other, but to them it looks like all others don't show any signs of distress. Thus they all argue that the event is not problematic, even though they all believed the event was problematic. This phenomenon is know as **pluralistic ignorance**. In the third step, if a bystander does interpret the event correctly he still has to determine if he is going to take responsibility for the intervention. If the bystander is alone he is likely to intervene at this point, but if he is in a group of other bystanders **diffusion of responsibility** can occur. In the group, uncertainty can arise about who is most qualified to take action or whether perhaps someone has already acted (e.g. by calling the authorities). This uncertainty may prevent any of the bystanders from interacting, resulting in a situation in which no one intervenes.

Many different examples of the Bystander Effect and Pluralistic Ignorance exist. One of the first examples was the Abilene Paradox [3]. Jerry B. Harvey was enjoying a hot afternoon with his family in Coleman, Texas, when his father in-law suggested to have dinner in Abilene, 85 km away. Everyone agrees and Jerry, not wanting to spoil the fun, agrees as well. The drive is hot and long, and the food even worse. When the group returns, one of them sarcastically states how 'wonderful' the trip was. Everyone agrees that they did not want to go. The group is perplexed.

Everyone agreed to go on a trip which no one wanted to join. More serious problems exist as well. Consider a crime taking place in public and a crowd is able to see the crime and intervene. However, the crowd can see the crowd as well, and because people see others do not react tot the crime, they are themselves reluctant to react as well [4]. Another example is climate change. In this case, everyone is responsible and everyone knows that everyone is responsible. Next to this, everyone can see that others are not reacting, and therefore can assume that others are accepting the norm of polluting [5]. Diffusion of Responsibility can also be a valid reason for climate change. Pluralistic ignorance is also found in alcohol consumption on campus. Many students assume they are more uncomfortable with alcohol practices than the other students. Students even felt alienated from the campus, although this perceived deviance to peers was merely an illusion [6]. One problem which had a big impact on the history of the United States is the segregation of whites and blacks. O'Gorman [7] found that many whites overestimated the support of other whites for racial segregation, while themselves not being in favor of segregation. Data shows that in East North Central America the amount of people in favor was 14% whereas the same people believed that 48% of the other people were in favor of segregation. Nevertheless it took a long time for many of the whites to publicly express their problems with divide between white and dark skinned people. Lastly, this problem occurs in board rooms. When a certain strategy is not feasible anymore, directors often continue in the same path whereas they privately believe this is not the right direction. The members on the board in charge of the company could thus be in a state of pluralistic ignorance, resulting in the bystander effect[8].

In this paper we will mainly be focusing on the bystander effect that can arise in the second step of the above described process, paying special attention to pluralistic ignorance. We will be using the logical framework introduced in [9], which we will discuss in Section 2. In Section 3, three basic agent types will be introduced: **first responders** which will always intervene when they believe something is wrong, **hesitators**, which will first look for social proof by observing other bystanders and **city dwellers**, which are apathetic agents that never intervene, even when they know an event is problematic. When first responders are present clearly the bystander effect cannot occur, but when only hesitators or a mix of hesitators and city dwellers are present the bystander always occurs in the framework discussed. We will show this in Section 4, where we work out an example of pluralistic ignorance arising in a group of 4 hesitators. Since the bystander effect can have severe negative consequences, as in the drunk driving situation discussed above, we will discuss possible ways to prevent or circumvent the bystander effect in Section 5. We created an implementation build upon the framework we use, which will be discussed in Section 6. Finally we will have a final discussion in Section 7

2 Plausibility models

To model the bystander effect that occurs in the case of misinterpretation of a problematic event due to deceiving social proof, we will first discus a group of **plausiblity models**, which were introduced in [9]. This will include static models, describing a fixed scenario, and action models, which describe actions or events. Later we will discuss an update that, given a static model, can apply an action model. This results in a new static model that describes the first static model after the action(s) or event(s) described by the action model have taken place.

2.1 Epistemic plausibility models

First lets take a look at the static models. The base definition of a static model is given below.

Definition 1 (Epistemic Plausibility Model (EPM)) *: An epistemic plausibility model is denoted by* $\mathbf{S} = (S, \leq_i, \Phi)_{i \in \mathcal{A}}$*, where* S *is a finite set of states, and* $\leq_i$ *for all* $i \in \mathcal{A}$ *makes up a set of preoredered relations between the states in* S *and* Φ *is a set of the atomic doxastic propositions.*

Here the relations $\leq_i$ represent plausibility relations between states. If two states $s, t \in S$ are connected by $s \leq_i t$ or $t \geq_i s$ two states are said to be indistinguishable by agent $i \in \mathcal{A}$, i.e. both states are seen as plausible by the agent. However if we have for two states $s, t \in S$ the relation $s \leq_i t$ but not the relation $s \geq_i t$ (which we will express as $s <_i t$ in the rest of this work), then for agent $i \in \mathcal{A}$ state s seems more plausible than state t. On the other hand, if both $s \leq_i t$ and $t \geq_i s$ both states are deemed equally likely by agent $i \in \mathcal{A}$. Note that since $\leq_i$ is a preorder, the relations $\leq_i$ are reflexive and transitive.

Now a **doxastic proposition** P, which from now on we will simply refer to as a proposition, is a map that for an epistemic plausibility model $\mathbf{S}$, describes a subset $(P)_{\mathbf{S}} \subseteq S$. If we have a state s and $s \in (P)_{\mathbf{S}}$, proposition P is said to be true in s, more formally, it is true in the world $(\mathbf{S}, s) \models P$. Let $\top$ denote true and $\bot$ denote false, for two arbitrary propositions P and Q as defined above, we have the following:

$$
\begin{array}{ccc}
(\top)_{\mathbf{S}} = S & (\bot)_{\mathbf{S}} = \emptyset & (\neg P)_{\mathbf{S}} = S \backslash P_{\mathbf{S}} \\
(P \wedge Q)_{\mathbf{S}} = P_{\mathbf{S}} \cap Q_{\mathbf{S}} & (P \vee Q)_{\mathbf{S}} = P_{\mathbf{S}} \cup Q_{\mathbf{S}} & (P \to Q)_{\mathbf{S}} = (S \backslash P_{\mathbf{S}}) \cup Q_{\mathbf{S}}
\end{array}
$$

When we want to model a situation in which we have some designated actual state s_0 we can use a pointed EPM as defined by the following.

Definition 1.1 (Pointed Epistemic Proposition Model(PEPM)) *: A pointed epistemic proposition model is given by* $\mathbf{S} = (S, \leq_i, \Phi, s_0)_{i \in \mathcal{A}}$ *where* S, $\leq_i$ *and* Φ *are as defined in 1 and* $s_0 \in S$ *denotes the actual state of the model.*

For a model $\mathbf{S}$ as defined in 1, we will use the shorthand $\sim_i = \leq_i \cup \geq_i$, the so-called **indistinguishability relation**. The **information cell** of an agent $i \in \mathcal{A}$ in a state $s \in S$ is then given by $\mathcal{K}_i[s] = \{t : s \sim_i t\}$. Furthermore agent $i \in \mathcal{A}$ in a state $s \in S$ has a **plausibility cell** given by $\mathcal{B}_i[s] = \{t \in [K]_i[s] : t \leq_i u, \text{ for all } u \in \mathcal{K}_i[s]\}$. Here the information cell of an agent in a state s represents all states that are deemed plausible by the agent in state s and the plausibility cell contains the states that the agent finds most plausible in state s. Using these cells we can now also introduce the epistemic and doxastic notions of knowledge and believe, which are given by the following:

$$
(K_i P)_{\mathbf{S}} = \{s \in S : \mathcal{K}_i[s] \subseteq P_{\mathcal{S}}\} \qquad (B_i)_{\mathbf{S}} = \{s \in S : \mathcal{B}_i[s] \subseteq P_{\mathcal{S}}\}
$$

Lets first work out a simple example in detail applying what has been discussed so far. Suppose Alice and Bob went to a grocery store and bought a raspberry muffin and a plain vanilla muffin. After coming home Bob goes out to walk the dog and he comes back home to see Alice sitting in the kitchen, in front of a plate with some muffin crumbs on it. Now Bob does not know which of the two muffins Alice ate, in addition he also holds it for possible that she ate both muffins. We can then use a PEPM $\mathbf{S}$ with three state states $s, t, u \in S$ to denote the situation. Let proposition R and V denote "Alice has eaten the raspberry muffin" and "Alice has eaten the vanilla muffin" respectively. Now then s can represent the situation in which Alice ate only the raspberry muffin, t the situation in which Alice only ate the vanilla muffin and u the situation in which both muffins were eaten by Alice. Therefore we must have $(R)_{\mathbf{S}} = \{s, u\}$ and $(V)_{\mathbf{S}} = \{t, u\}$, but as discussed earlier, we thus must also have $(R \wedge Q)_{\mathcal{S}} = \{u\}$. Bob holds all three states for possible, but he remembers that Alice mentioned being very hungry while shopping earlier, so he finds it most plausible that she ate both muffins, which can be described by the relations $u <_b s$ and $u <_b t$. Furthermore he knows Alice likes both vanilla and raspberry muffins so he finds it equally plausible that she ate only the vanilla muffin or only the raspberry muffin, this can be given by the relations $s \leq_b t$ and $s \geq_b t$. Clearly Alice only holds for possible the state she is in, since she knows what she actually ate, $\mathbf{S}$ thus doesn't include any relations between states s, t and u for her. As it happens Bob is thinking in the right direction, Alice indeed ate both muffins and thus the actual state is given by u.

Now we can use a Kripke model to represent $\mathbf{S}$ by simply using the same states and relations, where we will use a directed arrow to denote $<_i$, pointing towards the more likely state, and a

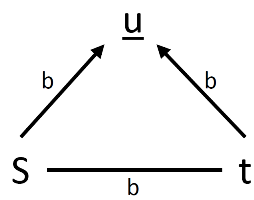

Figure 1: Simple example of a PEMP $\mathbf{S}$

line to denote a two way relation $\leq$ and $\geq$ between two states. We have defined propositions differently from the propositions used in Kripke model but let the proposition symbol represent the same thing and then the set of propositions Φ, $(P)_{\mathbf{S}}$ describing in which states a proposition is true, together form the valuation map of a Kripke model. However note that we will make the connection with Kripke models only to be able to use the graphical conventions associated with it, we will from now on only use the definitions and rules associated with EMPs and PMPs in the rest of this work. The simple scenario above can then be represented by figure 1, where reflexive relations are implied but not drawn and the underscore under state u denotes it is the true state.

Since bob has an indistinguishability relation between all states,the information cells of Bob in $\mathbf{S}$ can be given by $\mathcal{K}_b[s] = \mathcal{K}_b[t] = \mathcal{K}_b[u] = \{s, t, u\}$, whereas Alice only has reflexive relations so here information cells are given by $\mathcal{K}_a[s] = \{s\}$, $\mathcal{K}_a[t] = \{t\}$ and $\mathcal{K}_a[u] = \{u\}$. As discussed above from each state in S bob deems u as the most plausible situation and the plausibility cells for bob in S thus are given by $\mathcal{B}_b[s] = \mathcal{B}_b[t] = \mathcal{B}_b[u] = \{u\}$. On the other hand Alices plausibility cells are exactly the same as here information cells, thus giving $\mathcal{B}_a[s] = \{s\}$, $\mathcal{B}_a[t] = \{t\}$ and $\mathcal{B}_a[u] = \{u\}$. Using these cells we can now also look at the believes and knowledge Bob and Alice have. For Alice we have $(K_aR)_{\mathcal{S}} = (B_aV)_{\mathcal{S}} = \{s, u\}$ and $(K_aR)_{\mathcal{S}} = (B_aV)_{\mathcal{S}} = \{t, u\}$, which exactly match $(R)_{\mathbf{S}}$ and $(V)_{\mathbf{S}}$, which just means, as we already knew, that she always knows (and believes) what she has eaten. On the other hand for Bob we have $(K_bR)_{\mathcal{S}} = (K_bV)_{\mathcal{S}} = \emptyset$, showing that Bob indeed does not know which muffin(s) Alice ate. Nevertheless Bob's believes are given by $(B_bR)_{\mathbf{S}} = (B_bV)_{\mathbf{S}} = \{s, t, u\}$, again conforming that Bob believes Alice ate both the raspberry and the vanilla in all states of S (which is equivalent to writing $(B_b(R \wedge V)_{\mathbf{S}} = \{s, t, u\}$, which can also formally be show to be true using the above techniques).

2.2 Action plausibility models

So far we have only discussed static states, but if as we mentioned earlier if we want to describe the bystander effect we have to be able to make changes to the model that occur due to certain events or actions, like the problematic event or the reaction of the agents to the event. For this we will introduce action models, which in our framework will be defined as below.

Definition 2 (Action Plausibility Model (APM)) *: an action plausibility model is given by* $\mathbf{E} = (\Sigma, \leq_i, pre, post)_{i=\mathcal{A}}$*, where* Σ *is a finite set of states describing actions,* $\leq_i$ *are plausibility relations between these states, pre defines a precondition map on* Σ *and post describes a post condition map on* Σ*.*

To be more specific a **preconditions** $pre(\sigma)$ returns for an action state $\sigma \in \Sigma$, a propositional formula that has to be true in a static state $s \in S$ for the action to be applicable. On the other hand **postconditons** $post(\sigma)$ returns for an action state $\sigma \in \Sigma$ a propositional formula which will become true in a new static state which is formed by updating a state that satisfies $pre(\sigma)$ with the action of σ. How this update is performed will be discussed in the following Section (Section 2.3). Similar to before we can also define a pointed model, in which a state σ_0 is included which represents the actual event or action that has taken place.

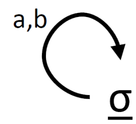

Figure 2: Simple example of a PAMP **E**

Definition 2.1 (Pointed Action Plausibility Model (PAPM)) *: an action plausibility model is given by* $\mathbf{E} = (\Sigma, \leq_i, pre, post, \sigma_0)_{i=\mathcal{A}}$*, where* Σ*,* $\leq_i$*, pre and post as in definition 2 and* σ_0 *describes the actual action(s) performed.*

Usually we will only look at an action model that represents one particular action or a combined set of related actions. This can be represented using a **doxastic program** Γ, which is similar to a proposition in an EPM. This thus means that a doxastic program corresponds to a subset of all actions/states in the models event space $\Gamma \subseteq \Sigma$. In the case of the bystander effect a doxastic program could for example encapsulate to the occurrence of a problematic event.

Lets now return back to the simple situation discussed at the end of Section 2.1. Alice sees Bob's questioning demeanour and decides to tell him she ate both muffins. This action can be represented by a program $\Gamma = \{\sigma\}$ where σ represents the action of Alice announcing that she ate both muffins and σ only has reflexive relations. This situation is given in figure 2.
Since Alice can only announce that she ate both muffins if she actually ate both the raspberry and the blueberry muffin, the precondition for σ is given by $pre(\sigma) = R \wedge V$. Since the action does not involve any actual change in the real world, the postcondition is given by $post(\sigma) = \top$.

2.3 Action priority updates

Now given an APM (which could for example be some doxastic program) we want to find an **action priority update product** between the old EPM and the APM, where this product represents the new EPM that describes the new static state after the action(s) or event(s) of the APM have occured. This update must take into account what new states are included in the new model as given by the the preconditions of the APM, how the new relations are formed between the new states and what propositional changes the postconditions of the APM bring about. The update is defined below.

Definition 3 (Action-Priority Update (APU)) *: is a binary operation denoted by* $\otimes$*, where the left argument is an EPM* **S** *as defined in 1 and the right argument is an APM* **E** *as defined in 2. The resulting product is again an EPM, given by* $\mathbf{S} \otimes \mathbf{E} = (S \otimes \Sigma, \leq_I^{\uparrow}, \Phi^{\uparrow})$*, in which the new set of states is given by* $S \otimes \Sigma = \{(s, \sigma) \in S \times \Gamma : (\mathbf{S}, s) \models (s_0, \sigma_0)\}$*. A new pre-order* $\leq_i^{\uparrow}$ *given by* $(s, \sigma) \leq_i^{\uparrow} (t, \tau)$ *will be included in the new EPM* $\mathbf{S} \otimes \mathbf{E}$ *iff* $\sigma <_i \tau$ *and* $s \sim_i t$*, or* $\sigma \leq_i \tau$ *and* $\sigma \geq_i \tau$ *and* $s \leq_i t$*.* $\Phi^{\uparrow}$ *is* Φ*, but with the requirement that for all* $P \in \Phi$ *we have* $P_{\mathbf{S} \otimes \mathbf{E}} = \{(s, \sigma) : s \in P_{\mathbf{S}} \text{ and } post(\sigma) \not\models \neg P\} \cap \{(s, \sigma) : post(\sigma) \models P\}$*.*

Similar to above, if we want to consider the situation in which we have one static state s_0 defining the actual state and one action state σ_0 which represents the actual action, we also want to get a new PEMP which has one state designated as the new actual state. This can be done by defining the following.

Definition 3.1 (Pointed Action-Priority Update (PAPU)) *: is a binary operation denoted by* $\otimes$*, where the left argument is an PEPM* **S** *with actual state* s_0 *as defined in 1.1 and the right argument a PAPM* **E** *with an actual action state* σ_0 *as defined in 2.1. The resulting product is a new PEMP, given by* $\mathbf{S} \otimes \mathbf{E} = (S \otimes \Sigma, \leq_I^{\uparrow}, \Phi^{\uparrow}, (s_0, \sigma_0))$ *where* $S \otimes \Sigma$*,* $\leq_I^{\uparrow}$ *and* $\Phi^{\uparrow}$ *as in definition 3 and* (s_0, σ_0) *represents the new actual state.*

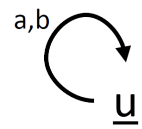

Figure 3: New PEMP that resulted from updating the PEMP in figure 1 with the PAMP in figure 2

Before, we established an initial PEPM for a simple example involving two muffins that Alice ate, followed by the action of Alice announcing that she ate the muffins given by a doxastic program over a PAPM. Using the update in definition 3.1 we can now describe a new state that occurs after applying the action to the static model. The product $S \times \Gamma$ results in the set of states $(s,\sigma),(t,\sigma),(u,\sigma)$. However we still have to check which of these states is actually included in the model by evaluating precondition $pre(\sigma) = R \wedge V$. Since $(\mathbf{S}, s) \not\models R \wedge V$ and $(\mathbf{S}, t) \not\models R \wedge V$, neither s nor t satisifes the precondition of σ, meaning that the states (s,σ) and (t,σ) will not be part of the new model. On the other hand $(\mathbf{S}, u) \models R \wedge V$ thus (u,σ) will be included in the model, which is the only and actual state of the model. Now since σ only contains reflexive relations the new model will only include relations $(u,\sigma) \leq_a (u,\sigma)$ and $(u,\sigma) \leq_b (u,\sigma)$. The new model can be represented by figure 3.

Since $u \in (R)_{\mathbf{S}}$ and $u \in (V)_{\mathbf{S}}$ but also $pre(\sigma) \not\models \neg R$ and $pre(\sigma) \not\models \neg V$, we have $(P)_{\mathbf{S}\otimes\mathbf{E}} = \{u\}$ and $(Q)_{\mathbf{S}\otimes\mathbf{E}} = \{u\}$. It is now straightforward to show that we also have $(B_a(R \wedge V))_{\mathbf{S}\otimes\mathbf{E}} = (B_b(R \wedge V))_{\mathbf{S}\otimes\mathbf{E}} = (K_a(R \wedge V))_{\mathbf{S}\otimes\mathbf{E}} = (K_b(R \wedge V))_{\mathbf{S}\otimes\mathbf{E}} = \{u\}$, i.e. now both Alice and Bob know that Alice ate both the raspberry and the vanilla muffin.

3 Transition rules

If a problematic event occurs no choice is involved and an AMP will always look the same for a fixed number of agents. However this is not the case when the agents themselves decide to perform a certain action, in this case an element of choice is involved. This could be approached by a range of game theoretic structures but we decided to stick with the relatively simple approach from [9]. The trick is to introduce **transition rules**, structures that given an EMP can provide a local update determined by an action satisfying certain believe or guideline an agent has. A transition rule consists of a **trigger condition** φ and a **goal formula** ψ, where both φ and ψ are propositional formulas. The trigger condition specifies a condition that has to be true in a state $s \in S$ of an EPM in order for the the transition rule to be applicable and the goal formula specifies what propositional change has to occur after applying an update.

A transition rule itself is neither an APM nor a proposition, it is a rules that prescribes an APM that brings about the desired states in the applicable states. The transition rule $\mathcal{T}$ can be given by $\varphi \rightsquigarrow [X]\varphi$, where X are programs Γ, over APMs, bringing about the desired change, $[\Gamma]$ are **dynamic modalities** and $([\Gamma]\psi)_{\mathbf{S}}$ are proposition given by

$$([\Gamma]\psi)_{\mathbf{S}} = \{s \in S : \forall \sigma \in \Gamma, \text{ if } (s,\sigma) \in S \otimes \Gamma \text{ then } (s,\sigma) \in (\psi)_{\mathbf{S}\otimes\Gamma}\}.$$

What this essentially means is that $[\Gamma]\psi$ is true in a state s of EPM $\mathbf{S}$ iff for every action σ that is part of the program Γ if the update on s leads to a new state (s,σ) in $\mathbf{S} \otimes \Gamma$ then we must have that in the new state ψ is true. You can also define these modalities on a per agent basis, which is useful when a transition rules has to specify a choice made by an agent. Γ is still a program but the modalities are now denoted by $[\Gamma]_i$ for an agent $i \in \mathcal{A}$ and the propositions by $([\Gamma]_i\psi)_{\mathbf{S}}$ are given by

First Responders	Hesitators	City Dwellers
$B_iA \rightsquigarrow [X]_iI_i$ $B_i\neg A \rightsquigarrow [X]_iE_i$	$K_iA \rightsquigarrow [X]_iI_i$ $B_iA \wedge \neg K_iA \rightsquigarrow [X]_iO_i$ $B_i \rightsquigarrow [X]_iE_i$	$B_iA \rightsquigarrow [X]_iE_i$ $B_i\neg A \rightsquigarrow [X]_iE_i$

Table 1: Decision rules for the three different agent types. For a first responder agent $i \in \mathcal{A}$ the first rule can be denoted by $\mathcal{F}_{1i}$, the second rule as $\mathcal{F}_{2i}$ and the set of rules becomes $\mathbf{F}_i = \{\mathcal{F}_{1i}, \mathcal{F}_{2i}\}$. Similarly if an agent $i \in \mathcal{A}$ is a hesitator or a city dweller we have $\mathcal{H}_{1i}$, $\mathcal{H}_{2i}$, $\mathcal{H}_{3i}$ and $\mathbf{H}_i = \{\mathcal{H}_{1i}, \mathcal{H}_{2i}, \mathcal{H}_{3i}\}$, or $\mathcal{C}_{1i}$, $\mathcal{C}_{2i}$ and $\mathbf{C}_i = \{\mathcal{C}_{1i}, \mathcal{C}_{2i}\}$.

$$([\Gamma]_i\psi)_{\mathbf{S}} = \{s \in S : \forall\sigma \in \Gamma,\ \text{if } (s,\sigma) \in S \otimes \Gamma \cap \mathcal{K}_i[(s_0,\sigma_0)] \text{ then } (s,\sigma) \in (\psi)_{\mathbf{S}\otimes\Gamma}\},$$

where this means that $[\Gamma]\psi$ is true in a state s of a PEPM $\mathbf{S}$ iff for every action σ that is part of the program Γ if the update on s leads to a new state (s,σ) that is part of the information cell of agent i relative to the new actual world (s_0,σ_0) in $\mathbf{S} \otimes \Gamma$, then we must have that in the new state ψ is true. This also automatically ensure that the agent knows which action he chose, which is of course what we want.

The solution to the transition rule $\mathcal{T}$ over a state s in EPM $\mathbf{S}$ is a program Γ for which $(\mathbf{S}, s) \models \varphi \rightarrow [\Gamma]\psi$. Now if we have a set of transition rules $\mathbf{T} = \{\mathcal{T}_1, ..., \mathcal{T}_n\}$ containing a transition rule $\mathcal{T}_i$ for a choice by each agent $i \in \mathcal{A}$ over a world $(\mathcal{S}, s)$, a program γ is a solution to $\mathcal{T}$ if $(\mathbf{S}, s) \models \bigwedge_1^n(\varphi_i \rightarrow [\Gamma]\psi_i)$ (where a formula ϕ_i can contain agent specific actions or epistemic and doxastic operators for agent i), that is Γ forms a solution to all transition rules $\mathcal{T}_i$ at the same time. Now we can also specify a solution $\mathbf{G}$, which makes up a set of doxastic programs, to a set of transition rules $\mathbf{T}$ over an EPM $\mathbf{S}$ iff for every state $s \in S$, there is a program $\Gamma \in \mathbf{G}$ that forms a solution to $\mathbf{T}$ over world $(\mathbf{S}, s)$. Now multiple solutions might exist to a set of transition rules, but a solution $\mathbf{G}$ can be chosen in such a way that a deterministic choice arises, i.e. $\mathbf{G}$ contains a unique Γ for each state s.

3.1 Agent types

Now we want to discuss how these transition rules can be used to model the aforementioned agent types: first responders, city dwellers and hesitators. First of all lets introduce a specific type of transition rule, a so called **decision rule**, which is a transition rule with a doxastic trigger condition. In these transition rules we will need to include proposition describing three types of actions performed by an agent. After a problematic event an agent $i \in \mathcal{A}$ can intervene, denoted by proposition I_i, an agent $i \in \mathcal{A}$ can evade the problematic scenario, denoted by proposition E_i, or an agent $i \in \mathcal{A}$ can choose to observe in order to gather social proof on whether the event is really problematic, this is denoted by proposition O_i. Table 1 then describes the decision rules in response to a problematic event for the different agent types.

First responders thus intervene when they believe an event is problematic ($\mathcal{F}_{1i}$) and will evade when they think an event is not problematic ($\mathcal{F}_{2i}$). A first responder could be any kind of assertive agent, in the context of an accident on the streets it could for example be a police office or a medic, but it could also be a normal person with the right mindset. City dwellers will always evade, no matter if they believe an event is problematic ($\mathcal{C}_{1i}$) or not ($\mathcal{C}_{2i}$). A city dweller is an apathetic agent that does not care about the others around it. Finally a hesitator only intervenes if it knows an event is problematic ($\mathcal{H}_{1i}$), it will look for social proof it it believes an event is problematic but does not know it is ($\mathcal{H}_{2i}$) and it will evade when it believes an event is not problematic ($\mathcal{H}_{3i}$). In reality this often represent a typical person.

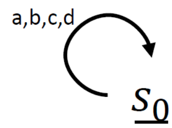

Figure 4: The initial PEMP $\mathbf{S}_0$ of the bystander example, in which nothing has happened yet. All atomic propositions are false: $(A)_{\mathbf{S}_0} = (I_i)_{\mathbf{S}_0} = (O_i)_{\mathbf{S}_0} = (E_i)_{\mathbf{S}_0} = \emptyset$.

4 The bystander effect worked out

Now we have the logical framework from [9] in place, we can start to explore the bystander effect. More specifically we will look at the bystander effect that arises when agents believe something problematic has happened, but are uncertain about it. Like mentioned in the introduction, if there is a first responder among the agents in the presence of a problematic event the first responder will always intervene, so there is no way the bystander effect can occur. On the opposite site of the spectrum we have the city dweller which will always evade the problematic event, which is not really what the bystander effect is about. Therefore in the following we will focus on the situation where only hesitators are present, which as we will see in due course will lead to pluralistic ignorance. Nevertheless we will also make some comments about first responders and city dwellers since they can still be of interest, for example note that the bystander effect can also occur if a mix of hestitators and city dwellers is present.

For now we will begin with working out an example of the bystander effect surrounding drunk driving, the first example we discussed in the introduction. As we saw shortly after however, a wide range of scenarios in which the bystander effect can occur. The first few steps will be exactly the same in all of these scenarios, but in the next to last step differences may occur, we will deal with them when our discussion gets to that point.

4.1 The initial situation

Consider the situation in which 5 people are present in a small bar. One of them is very intoxicated, the four other are either sober or have only drunken a moderate amount of alcohol. It is these four that will be our hesitator agents, we will simply refer to them as agent a, b, c and d. The set of agents is thus given by $\mathcal{A} = \{a, b, c, d\}$. We can then now try to constrict our initial PEMP $\mathbf{S}_0$ (see definition 1.1). A proposition A could denote that a problematic event occurs and as we saw before we will also need propositions E_i, O_i and I_i for all agents $i \in \mathcal{A}$. Initially all of these are false in our model and these propositional atoms make up Φ of our initial PEMP. Since an agent can only perform one of the actions at the same time we have that at any point we must have $(I_i)_\mathbf{s} \cap (O_i)_\mathbf{s} = (O_i)_\mathbf{s} \cap (E_i)_\mathbf{s} = (E_i)_\mathbf{s} \cap (I_i)_\mathbf{s} = \emptyset$. We only need one state $s_0 \in S_0$ initially, where we have the reflexive relation $s_0 \leq_i s_0$ for all $i \in \mathcal{A}$. The PEMP can thus be described by figure 4.

Since the information cell of every agent contains the only state in the model, every agent knows what has happened, that is they know nothing has happened yet. After a problematic event, the nature of which we will go into in the next subsection, the end condition of the run of actions is met when either at least one agent intervenes, $\bigvee_{i \in \mathcal{A}} I_i$, or when all agents evade, $\bigwedge_{i \in \mathcal{A}} E_i$.

4.2 A problem arises

The inebriated person gets up and takes out his car keys, he is planning on driving home. Like mentioned earlier drunk driving often leads to serious accidents, so this event could certainly be considered problematic. Then in the PAMP modelling this event, the actual event σ_0 should contain the postcondition $post(\sigma_0) = A$. Which states that the truth value of A, which now describes the problematic event of the drunk going for his car, should change from false in the actual state s_0 of initial PEMP $\mathbf{S}_0$, to true in the actual state of the PEMP $\mathbf{S}_1$ after the event has occurred.

Since all agents are hesitators we will assume their tought process is identical and we will thus start by looking at the event from the point of view of one of the agents, let's take agent a. Agent a will not be completely certain that event is problematic, for example the person might be less drunk than agent a perceives (just to be clear, we think you should only drive when you are completely sober!) or the person might just be looking for his house key and might not plan on actually driving home. Agent a therefore considers an other event τ_0 possible in which no problematic event has occurred, described by postcondition $post(\tau_0) = \top$. But it does find it more plausible that the event is problematic, so a relation between the event states is given by $\sigma_0 < \tau_0$. Since the three other people at the bar (not including the drunk), agents b, c and d, all witness the event we make the assumption that a perceives the other agents as also forming an opinion about the event, that is whether it is problematic or not. Agent a being a normal human cannot read the other agents mind, it therefore considers it possible that the other agents, like itself, consider the event to be problematic. However he also cannot rule out the situation in which any combination of the other agents find it more likely that the event was not problematic. This situation is then captured by the PAMP in figure 5.

We see that the PAMP in the image correctly captures that agent a finds it most plausible that a problematic event has occurred. Also depicted is that agent a is uncertain about the other agents, on one hand it considers possible that all other agents also find it more plausible that the event is problematic, as given by the event pair (σ_0, τ_o). Nevertheless agent a also holds for possible that one of the other agents finds it more plausible that the event is not problematic, denoted by event pairs (σ_1, τ_1), (σ_2, τ_2) and (σ_3, τ_3). Furthermore a can also not be sure that even 2 or all 3 other agents find it more likely no problematic event has transpired, as displayed by event pairs (σ_4, τ_4), (σ_5, τ_5), (σ_6, τ_6) and (σ_7, τ_7).

Now there is still one other thing agent a has to take into account, agent a can also not be sure what other agents think about a's own state of mind. They could hold it for possible that agent a does not find the event problematic. To model this additional state pairs (σ_k, τ_k) can be introduced which are connected by a relation $\sigma_k >_j \tau_k$ for $i \neq j$. If we then also assume that agents b, c and d perceive the event in the same way as agent a the complete $PAMP$ is given by figure 6.

The initial PEMP $\mathbf{S}_0$ can now be updated with PAMP $\mathbf{E}_0$ using a PAPU (See definition 3.1). The resulting PEMP $\mathbf{S}_1$ has exactly the same structure as $\mathbf{E}_0$, where every new state is simply a pair (s_0, σ_k), residing at the same location as σ_k, or a pair (s_0, τ_k), at the same location as τ_k. The new actual state is (s_0, σ_0) and for all $k \in 0, ..., 15$ we have $(s_0, \sigma_k) \in (A)_{\mathbf{S}_1}$ and $(s_0, \tau_k) \in (\neg A)_{\mathbf{S}_1}$. From $\mathbf{S}_1$ we can conclude that in the actual state (s_0, σ_0) we have the following important propositions:A, $\bigwedge_{i \in \mathcal{A}} B_i A$, $\bigwedge_{i \in \mathcal{A}} \neg K_i A$, $\bigwedge_{i \in \mathcal{A}} K_i \bigwedge_{j \in \mathcal{A}} (B_i A \vee B_i \neg A)$, $\bigwedge_{i \in \mathcal{A}} K_i \bigwedge_{j \in \mathcal{A} \setminus \{i\}} \neg B_i B_j A$, $\bigwedge_{i \in \mathcal{A}} K_i \bigwedge_{j \in \mathcal{A} \setminus \{i\}} \neg B_i B_j \neg A$. The most important things to take away from this is that all agents believe the event is problematic, but do not know it is , all agents know the other agents either believe the event is problematic or believe it is not problematic, and all agents hold it for possible that the other agents believe nothing has happened.

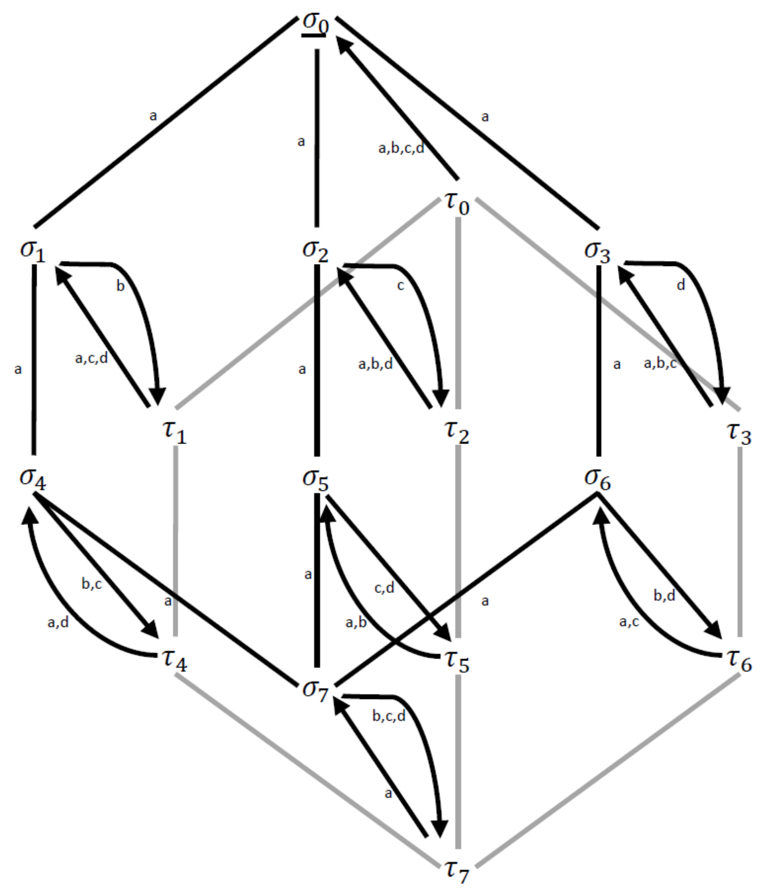

Figure 5: This PAMP with actual event σ_0 describes agent a's perception of the event. All σ_k events describe a problematic event has taken place, captured by $post(\sigma_k) = A$, while all τ_k events describe no problematic event has occurred, as given by $post(\tau_k) = \top$. The precondition for all states is $\top$, since the event happens no matter what. Reflexive relations and transitive relations have been left out. The labels of the grey relations are left out for clarity of the picture, but they simply mirror the corresponding parallel black relations.

4.3 Looking for social proof

Now that the drunk has gotten up with his car keys in hand, the other four people in the bar must decide whether to intervene, observe or evade the situation. Since they are all hesitators they will use the set of decision rules $\mathcal{H}_i$ (see table 1) to make their action choice. However, to apply any of these rules we will have to find at an appropriate set of doxastic programs that can solve for X. For this we have to take into account that if an agent intervenes, it is fair to assume al agents will know the agent is intervening, i.e. the act of intervening is epistemically unambiguous for all agents. On the other hand if an agent choices either to evade or to observe it seems likely the other agents cannot tell the difference between these two actions, since the two don't have any distinguishing traits that are detectable by the other agents. A crucial assumption made in [9] is that agents find it more plausible that the other agents evade than that they observe, which as we will see later is a direct cause of pluralistic ignorance in this model. These considerations will lead to AMP $\mathbf{E}_{1i}$ in figure 7.

The programs in $\mathbf{E}_{1i}$ do not allow the agents to perform more than one of the actions, that is the programs cannot satisfy more than one rule in $\mathcal{H}_i$ at the same time, which satisfies the con-

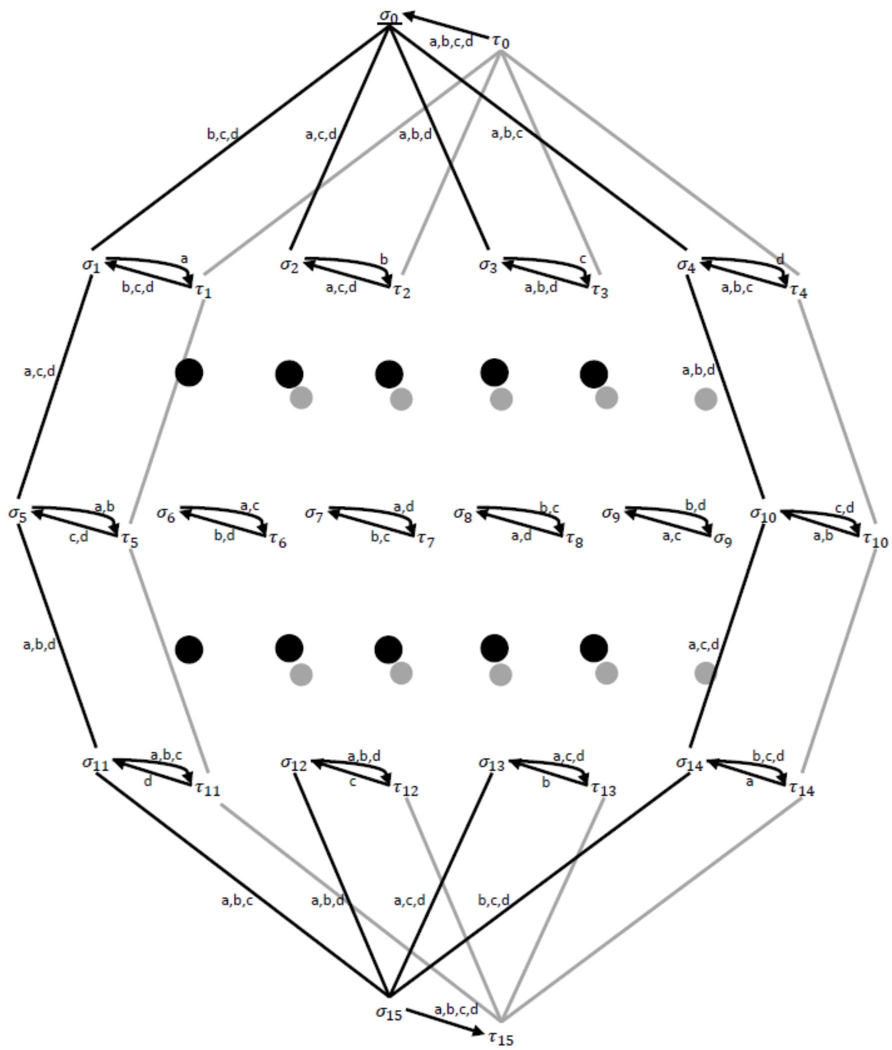

Figure 6: PAMP $\mathbf{E}_0$ with actual event σ_0 describes the joint perception of agents a, b, c and d of a problematic event. To prevent the figure from becoming too messy some of the relations have been left out, at the location of the circles. Event states in these region should be connected by relations $\sigma_k \leq_i \sigma_l$ and $\sigma_k \geq_i \sigma_l$ iff agent i considers the same state most plausible in σ_k as in σ_l. The precondition of all events σ_k and τ_k is $pre(\sigma_k) = pre(\tau_k) = \top$, furthermore for all events σ_k we have postcondition $post(\sigma_k) = A$ and for all events τ_k we have postcondition $post(\tau_k) = \top$. Again reflexive and transitive relations have been left out and grey relations mirror the corresponding black relations.

dition we mentioned in section 4.1. To combine the appropriate programs from $\mathbf{E}_{1i}$ for all $i \in \mathcal{A}$ into one single PAMP E_1 one can take the reflexive, transitive closure of the **Cartesian graph product** of the programs (see e.g. [10]), ensuring that the relations of E_1 are in proper preorder. The Cartesian graph product $\mathbf{E}_k \Box \mathbf{E}_l$ of two AMPs includes the events $\Sigma_k \times \Sigma_l$ and there is a relation $(\sigma_k, \sigma_l) \leq_i (\tau_k, \tau_l)$ between two event states $(\sigma_k, \sigma_l), (\tau_k, \tau_l) \in \Sigma_k \times \Sigma_l$ if either $\sigma_k = \tau_k$ and $\sigma_l \leq_i \tau_l \in \mathbf{E}_l$ or $\sigma_l = \tau_l$ and $\sigma_k \leq_i \tau_k \in \mathbf{E}_k$. Additionally the preconditions and postconditions of the product APM are conjunctions of respectively the preconditions and postconditions of the separate APMs. Since we have four hesitators and in the actual state of (s_0, σ_0) of $\mathbf{S}_1$, all four hesitators believe a problematic event has occurred the decisions rules $\mathcal{H}_i$ dictate that we take

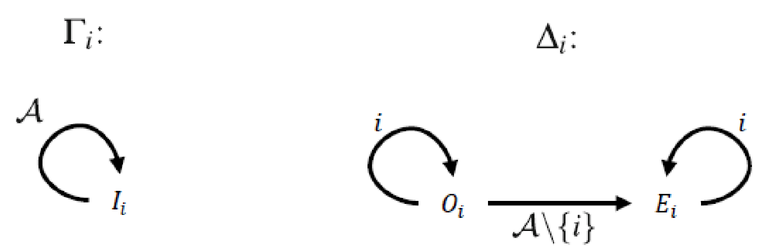

Figure 7: APM $\mathbf{E}_{1i}$ represents three different programs. First of all lets note that the precondition of all actions is $\top$ and that the states have been named to reflect their postcondition (e.g. $post(O_i) = O_i$). Γ_i represents the choice of agent $i \in \mathcal{A}$ to intervene, Δ_i with actual action O_i represents the choice of agent $i \in \mathcal{A}$ to observe and Δ_i with actual action E_i represents the choice of agent $i \in \mathcal{A}$ to evade

the reflexive, transitive closure of the Cartesian graph product $\Delta_a \Box \Delta_b \Box \Delta_c \Box \Delta_d$ with actual state $O_a O_b O_c O_d$. This gives rise to E_1 given in figure 8.

We are again in the position to update our static state, this time this will correspond to taking the PAPU of $\mathbf{S}_1$ and $\mathbf{E}_1$. This new EPM $\mathbf{S}_2 = \mathbf{S}_1 \otimes \mathbf{E}_1$ contains 512 states so we will not be able to represent the ensuing EPMs graphically, fortunately enough $\mathbf{S}_2$ is easy to describe. Take $\mathbf{E}_1$ and swap all event states with a complete duplicate of the frame of $\mathbf{S}_1$, then connect states from different duplicates by $(s, \sigma_k) \leq_i (s', \sigma_l)$ iff $s = s'$ and $\sigma_k \leq_i \sigma_l$ and then finally also enforce the reflexive and transitive closure. Among these 512 different states is the actual world $((s_0, \sigma_0), OOOO)$, in which the following propositions of importance hold: A, $O_a \wedge O_b \wedge O_c \wedge O_d$ and $\bigwedge_{i \in \mathcal{A}} (K_i O_i \wedge B_i \bigwedge_{j \in \mathcal{A} \setminus \{i\}} E_j)$. Crucially we thus see that all agents believe they were the only agent to observe, while ate the same time all other agents evaded. In reality of course all the hesitating agents observed.

4.4 Interpretation of social proof

Currently all four people believe that the other three people in the bar have evaded the situation of the potential drunk driver, however they have not yet interpreted this believe and they still think that it is equally likely that the other people do believe the drunk walking towards his car is problematic as that the other people do not believe the drunk walking towards his car. Given the current framework there is however no way for agents to interpret their believe about the actions of other agents. There are again a large number of ways this interpretation could be included, but we again choose to use the relatively straightforward way adopted in [9]. They argue that our believes about the situation at hand often determine which actions an agent makes, meaning that we could describe the transition from believe to action in a functional manner. Nevertheless one action could be caused by many different believes, therefore they state that abductive reasoning is needed.

This reasoning will be captured in so called **interpretation rules**. Interpretation rules describe propositions of the form $\varphi \to [\mathbf{S}] B_i \psi$, where φ will be named the **basis** and ψ will be called the **content**. The basis describes an action, and from this action the interpretation rule tries to deduce some of the believes of $i \in \mathcal{A}$ about the content. We have seen modalities before in transition rules, but now their interpretation is somewhat different, the proposition of a modality is now captured by

$$(\mathbf{S}_\chi)_{\mathbf{S}'} = \{s' \in S' : \exists s \in S \text{ such that } s \in s' \text{ and } (\mathbf{S}, s) \models \chi\},$$

here $s \in s'$ denotes that state s is a **predecessor** of state s', e.g. in our example s_0 is a predecesor of (s_0, σ_0) and $((s_0, \sigma_0), OOOO)$. What this thus means is that $[\mathbf{S}]\chi$ is true in a world $(\mathbf{S}', s')$ iff

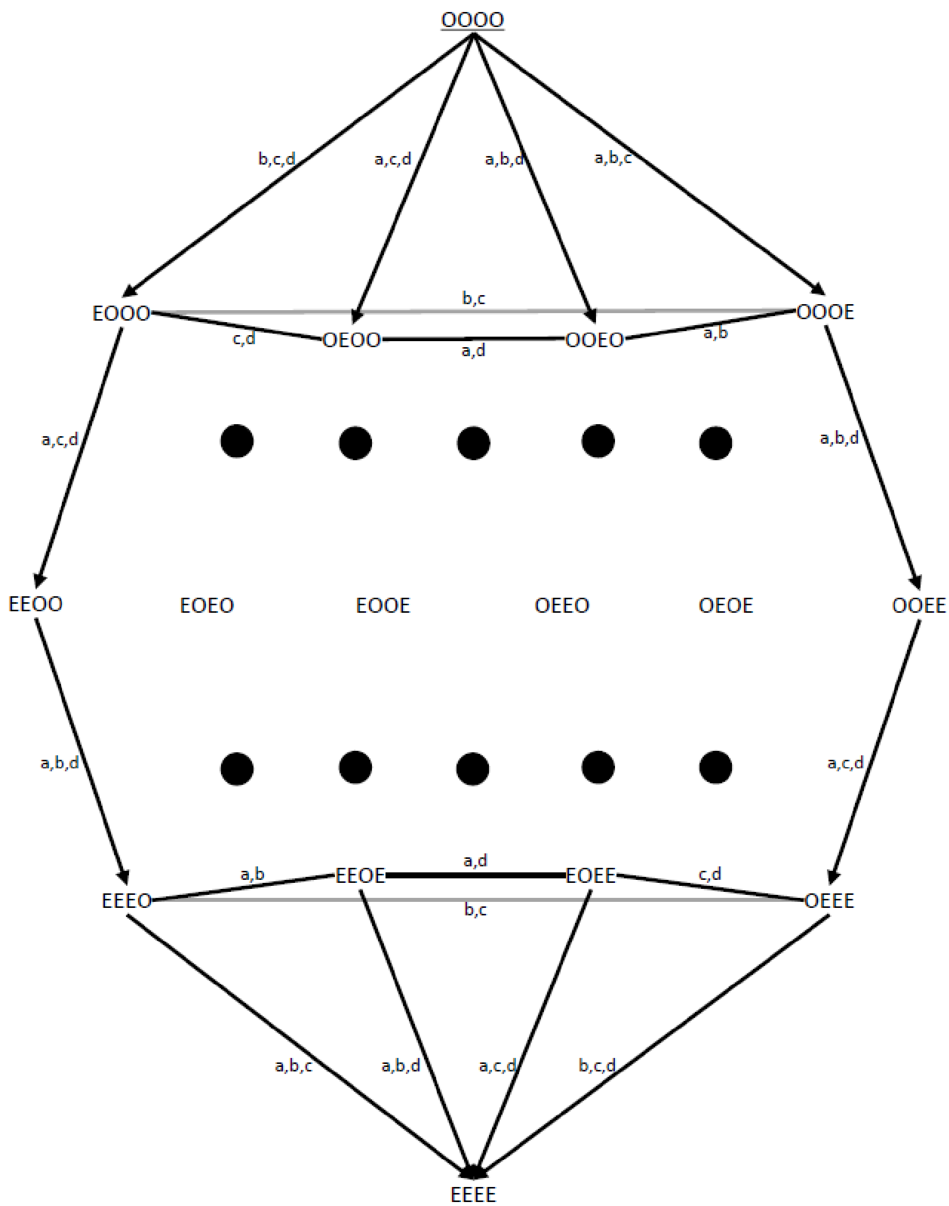

Figure 8: PAPM $\mathbf{E}_1$ with actual state $OOOO$, note that in the figure labels of the actions have been left out for brevity, also note the similarities with figure 6. Some of the relations have been left out for clarity of the figure, relation $\sigma_k \leq_i \sigma_l$ should be added if the action of agent $i \in \mathcal{A}$ is the same in state σ_k as in state σ_l and the number of evading agents in σ_k is greater or equal than the number of evading agents in state σ_l (this also hold for states in the same layer). Preconditions for all states are $\top$ while postconditions are reflected by the labels of the state, e.g. action state $post(EOOE) = E_a \wedge O_b \wedge O_c \wedge E_d$.

the predecessor of s' in $\mathbf{S}$ was a world $(\mathbf{S}, s)$ in which χ was true. The modality now thus allows agents in a model $\mathbf{S}'$ to reason about past models $\mathbf{S}$.

Interpretation rules will be implemented by putting a conjunction over all action that should be interpreted in the precondition of an AMP. In this APM different states can correspond to different hypotheses about the agents behavior. We will however only need one state, i.e. one hypothesis, to give an accurate conjunction of interpretation rules that almost exactly capture the logic of the decision rules in $\mathbf{H}_j$. An APM $\mathbf{E}_2 j$ then describe how all agents $i \in \mathcal{A} \backslash \{j\}$ interpret any action performed by agent j using only a one state σ_j with preconditions given by

$$pre(\sigma_j) = (I_j \rightarrow [\mathbf{S}_1] K_j A) \wedge (O_j \rightarrow [\mathbf{S}_1](B_j A \wedge \neg K_j A)) \wedge (E_j \rightarrow [\mathbf{S}_1] B_j \neg A)$$

Instead of creating a combined PAMP we update model $\mathbf{S}_2$ by sequentially applying $\mathbf{E}_{2j}$ (which has only one state which is thus automatically the actual state) for each $j \in \{a, b, c, d\}$ on $\mathbf{S}_2$ in its intermediate successors, which is allowed since the order of application does not matter. The resulting new PEMP after applying all updates will be named $\mathbf{S}_3$.

In $\mathbf{S}_2$ the most plausible states for agent $i \in \mathcal{A}$ from the actual state (the belief cell $\mathcal{B}_i[((s_0, \sigma_0), OOOO)]$) corresponded to states in which all other agents $j \in \mathcal{A} \backslash \{i\}$ evaded. Now of those states all of the states $s' \in \mathbf{S}_2$ containing an agent j that evades E_j, if in the predecessor of that state s', $s \in \mathbf{S}_1$ agent j believed the event was problematic it will not be included in the new model $\mathbf{S}_3$. Thus the believe cell at the new actual state for each agent $i \in \mathcal{A}$ in the new model will only contain states in which all other agents $j \in \mathcal{A} \backslash \{i\}$ both evaded and believe that no problematic event has occurred. What this means is that in the new PEMP $\mathbf{S}_3$ every agent believes that all other agents do not believe a problematic event has occurred, that is in the new actual world we have $\bigwedge_{i \in \mathcal{A}} B_i \bigwedge_{j \in \mathcal{A} \backslash \{i\}} B_j \neg A$. Note that if we had constructed $\mathbf{E}_1$ in such a way that agents find it more likely that other agents observe than evade, agents would now correctly believe that all other agent believe a problematic event has occurred.

4.5 Acting on the new believes

Now all people believe the other three people believe it is not a problem that the drunk is walking to his car. In the bystander effect that would mean that now the agent therefore concludes on this social proof that this event must indeed not be problematic and that there is therefore no cause to intervene. This would bring the agents in a state of pluralistic ignorance. But how do we represent this in our model? In [9] they introduce a new operator $SB_{i|G}$ which described the believe of agent $i \in \mathcal{A}$ under influence of social proof extracted from a group of agents G (G being all relevant agents involved in the problematic event, in our example simply $\mathcal{A}$). They argue that it is appropriate to use a simple majority vote on the believes on believes, including two parameters that regulate the self bias (the believe of the agent itself), this would be given by

$$s \in (SB_{i|G}\varphi) \text{ iff } \alpha + |\{j \in G : s \in (B_i B_j \varphi)_{\mathbf{S}}\}| > \beta + |\{j \in G : s \in (B_i B_j \neg\varphi)\}|,$$

where $\alpha = 0.5$ iff $s \in (B_i \varphi)_{\mathbf{S}}$ and $\alpha = 0$ otherwise, $\beta = 0.5$ if $s \in (B_i \neg\varphi)_{\mathbf{S}}$ and $\alpha = 0$ otherwise.

However we argue that this is not always appropriate. For example if a serious, yet ambiguous accident has occurred it seems more realistic that a hesitator would already intervene if it believes at least one other agent believes the event is indeed an accident. This could be represented by an alternate operator $SB^*_{i|G}$ for which we have in the case of a problematic event

$$s \in (SB^*_{i|G} A)_{\mathbf{S}} \text{ iff } B_i A \wedge B_i \bigvee_{j \in G \backslash \{i\}} B_j A \text{ or } B_i \neg A \wedge B_i \bigwedge_{j \in G \backslash \{i\}} B_j A.$$

where the second disjunct ensures that in the case $B_i A$ only $SB^*_{i|G} \neg A$ if all other agents believe the event is not problematic. In settings were an event is less problematic and where agents experience more social inhibition by other agents, majority voting might be appropriate. This could for example be the situation in which a lecture is unclear for some reason, but no students dare to raise their hand since they believe (a majority) of other do not believe the lecture is unclear.

On the other hand it could take place in a business meeting in which information is introduced that shows the company is going through a rough patch. Nobody dares to introduce an alternate strategy to the current one because they believe that their peers do not believe the current strategy is problematic and therefore do not want to risk embarrassment by making a suggestion that others might deem useless. Another point to be made is that hesitators could be more distinct, in the sense that some might intervene faster than others. Rather than using a fixed value of α and β it might therefore be useful to be able to vary them, where appropriate values can be taken based on the nature of the problematic event and potentially taking into account differences in assertiveness between hesitators.

For our present example it seems reasonable to use $SB^*_{i|G}$, considering the severeness of potential outcomes if the drunk is allowed to drive his car. However since we have that in the actual state of $\mathbf{S}_3$ an agent believes all other agents do not believe the event is problematic, we have that $B_iA \wedge B_i \bigvee_{j\in G\setminus\{i\}} B_jA$ is false for all four hesitators but on the other hand we have that $B_i \bigwedge_{j\in G\setminus\{i\}} B_j\neg A$. Consequently we have that in the actual state $SB^*_{i|G}A$ is false but $SB^*_{i|G}\neg A$ is true, that is the agent has an influenced believe that the event is not problematic.

But what does the agent do with this influenced believe? We can do this by adding two new decision rules to $\mathbf{H}_i$ that regulate what happens after an hesitator observes and as a consequence forms an influenced believe about the accident. These new rules can be given by

$$\mathcal{H}_{4i} = SB^*_iA \rightsquigarrow [X]_iI_i,$$

$$\mathcal{H}_{5i} = SB^*_i\neg A \rightsquigarrow [X]_iE_i,$$

represent that an agent will intervene if he has an influenced believe that an event is problematic and will evade if he has an influenced believe that the event is not problematic. Simply adding the operator SB_i instead of SB^*_i when one wants to majority voting, or some other form of voting when one also varies α and β. To ensure that the agent will not invoke $\mathcal{H}_{2i}$ again when the agent has formed an influenced believe on the event, we can modify it to $\neg K_iA \wedge \neg(SB^*_iA \wedge SB^*_i\neg A) \rightsquigarrow [X]_iO_i$.

Now in our example we have $SB^*_i\neg A$, so we want to invoke $\mathcal{H}_{5i}$. This results in an PAMP $\mathbf{E}^*_1$, which is simply $\mathbf{E}_1$ but instead with actual state $EEEE$ instead of $OOOO$. To arrive at a new PEMP $\mathbf{S}_4$ one thus has to perform the PAPU of $\mathbf{S}_3$ and $\mathbf{E}^*_1$. The result is that in the actual state of $\mathbf{S}_4$ we have that $\bigwedge_{i\in\mathcal{A}} E_i$, i.e. all agents evade. $\mathbf{S}_4$ thus satisfies one of the end conditions mentioned in Section 4.1, pluralistic ignorance has occured and all agents evaded the problem. This thus means that none of the 4 (mostly) sober people in the bar intervene with the drunk getting in his car. As we mentioned in the introduction results of this could be dramatic, possibly costing the life of the drunk or even worse of innocent passerby.

The methods we have discussed can be used to explore any problematic event, with any number of agents, consisting of any agent type. We have however focused on the case of 4 hesitators because only pure hesitator agent groups will experience pluralistic ignorance (the bystander effect can happen with a mix of hesitators and evaders however, just not the specific case of pluralistic ignorance), and because if we chose more than 4 agents the process will contain an exponentially increasing number of states and relations. Where using 4 agents could still give a clear picture but at the same time also gives an insight in the complexity of the problem.

5 Combating the bystander effect

Different methods have been proposed to combat pluralistic ignorance in literature. We first discuss several of these proposed methods. After that we will look at how some of these methods could be integrated in the model we discussed from [9].

5.1 Group Composition

Westphal et. al. ([8]) proposes dense friendship ties to moderate the effect of pluralistic ignorance as it increases the propensity for directors to speak up. One of the reasons for pluralistic ignorance given by Westphal et. al. is the loss of face when speaking up against the norm. Because friends will not condemn you for a different opinion the possibility to speak up is bigger. Another reason for pluralistic ignorance is the inability to judge the average belief of a group. Because friends see each other more often in informal settings, it is possible to speak more freely and discuss your private concerns. When people are able to discuss their own concerns, they might discover that others share their opinion and discuss their concerns with the whole group. Next to this Westphal et. al. find that demographic homogeneity decreases pluraistic ignorance. One of the reasons for this is that similarity on salient demographic characteristics increase social integration and level of communication between group members. This will increase trust and possibility to voice concerns in board rooms. This is contrary to recent efforts to increase diversity in board rooms. In a large scale survey, Westphal et. al. proved their hypothesis and showed that pluralistic ignorance decreases when number of friendship ties increase in outside directors increases.

5.2 Structured Debate

Katzenstein [11] proposes two methods to structure the debate. The Dialectical Inquiry method makes the participants look for the underlying assumptions of a plan, and develop a counter plan on opposite underlying assumptions. With this counter plan, differences come to light with the current plan and it is easier to voice concerns to the current plan. The other method is the Devils Advocate method. This plan rests on the same idea of the Dialectical Inquiry, finding opposite ideas of the current plan. However, instead of developing a counter plan, the participants critique the current plan until the plan is changed. This method encourages people to voice possible concerns, even if they do not feel these concerns are the norm. The structured debate methods are very useful in board room applications of pluralistic ignorance, but can not be used in situations like the drunk driver.

5.3 Increase Communication Skills

Another way to break the cycle of pluralistic ignorance is to increase communication skills [5]. Geiger and Swim propose to increase communication skills such that people have less problems speaking against the norm. Their idea rests on the assumption that people expect to be respected less by those who disagree with them than being disliked by those who oppose. By teaching people how to phrase opposition without losing respect of the group, it is possible to break pluralistic ignorance. This can be used in the climate change debate [5], by teaching a bartender how to intervene [1] or saying no to a beer on campus [6].

5.4 Anonymity

In an anonimity setting, participants do not know from each other who they are personally. Several, positive and negative, impacts on group behavior have been found [12]. One important impact of anonymity on the pluralistic ignorance is that it reduces *groupthink*. In groupthink, participants may withold concerns to make them not look weak, unintelligent or not commited to the group efforts. By making participants anonymous, it is easier for them to talk freely and their opinions will not be connected to their personal competence. This way it is a lot easier to express unpopular ideas. On the other hand, anonymity also has some negative impacts. Whereas it is easier to voice unpopular ideas, the possibility of *free riding* is also larger. Next to this, anonymity can not be used together with the positive influence of friendship ties, as participants do not know who they are working with. Anonymity can be used in board meetings [8], the climate change debate [5] but could also have had a big influence on the 1968 white segregation discussion [7].

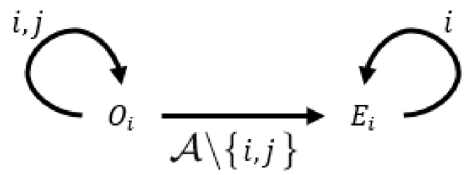

Figure 9: Modified version of Δ_i of E_{1i} in figure 7, that now describes the situation in which i and j are friends and agent i decides to observe.

5.5 Modelling solutions

First of all we will look at how we could implement friendship ties. In context of our example this could correspond to some or more of the four people, not including the drunk (but of course friendship ties with the drunk person could potentially also influence the situation), being friends. If a friend sees the drunk getting up reaching for his car keys, if the friend believes this is problematic it is very likely he will make at least some mention of it to his friend.

Lets say we introduce the possibility of friendship ties to or models, e.g. by extanding EPMs, APMs and APUs with a new relation of type $i \bowtie j$ which indicates a friendship tie between agents $i, j \in \mathcal{A}$. Now since friendship ties are only important in the case an agent observes, the easiest way to include their effect is by adapting E_{1j} in such a way that when an agent observes, his friends know he is observing. For example if $i \bowtie j$ we could edit Δ_i of E_{1j} to give the situation in figure 9.

The reasoning behind this is that if two friends are in the presence of a problematic event, if one of them chooses to observe the situation is is likely that he addresses the event to his friend, not necessarily stating he believes a problem has occurred, this act of addressing the event then forms a distinguishing mark between observing and evading. Note that when a friend evades one should still use Δ_i. Now if an agent knows another agent is observing, this automatically ensures that in the actual state of $\mathbf{S}_3$ he will believe this observing agent believes a problematic event has occurred. In the case the $BS^*_{i|G}$ operator is used this already breaks the bystander effect, in the case $BS_{i|G}$ is used it contributes to breaking the bystander effect, breaking the bystander effect if enough friendship ties are present in a group of agents. Alternatively one could introduce a form of communication, by which a friend asks a friend, or tells a friend about his believe on the event, but this will lead to the same result while adding unnecessary complexities if communication is only used for this.

Its also interesting to discuss the increase of communication skills, or rather education in general. For example if one of the people in our example was the bartender, if the bartender had received a training course involving how to handle potential drunk drivers or over drinking in general, it would become a lot more likely that the bartender intervenes. This is effect is however not limited to our example, it could apply to any problematic event and any form of education. If for example medics, police officers or people with any training in first-aid are present near a potential accident it might make it more likely for them to intervene, as opposed to the situation in which they did not have any training. We could thus see education as a means of changing a hesitator to a first responder in some scenarios. This has no direct consequences for the model we used, however the effect of education is thus that it can lead to a higher change of a fist responder being present during a problematic event, at the same time then also reducing the change the bystander effect takes place.

Now anonymity might not work in all scenarios, it won't help in the drunk driving example we discussed, nevertheless it might be of interest in some scenarios as discussed in Section 5.4.

Anonymity could for example take the form of broadcasting an anonymous message about your believes on the problematic event. The effect of this anonymous message send by agent $i \in \mathcal{A}$ is then that after the other agents receive it, we will have the situation that $\bigwedge_{j\in\mathcal{A}\setminus i} B_j \bigvee_{k\in\mathcal{A}\setminus j} B_k A$ but also $\bigwedge_{j\in\mathcal{A}} \neg B_j \forall_{i\in\mathcal{A}\setminus j} B_i A$. That is the subset of agents which excludes the broadcaster of an anonymous message all believe there is is some other agent that believes the event is problematic, while at the same time no agent believes a particular other agent believes the event is problematic, i.e. the broadcast receivers believe some agent believe the event is problematic but they are not aware which one. In the case that $BS^*_{j|G}$ is used if an agent $j \in \mathcal{A}\setminus i$ believes an event is problematic, after it receives the broadcast $B_j \bigvee_{k\in\mathcal{A}\setminus j} B_k A$ and thus $BS^*_{j|G}$ will become true, causing the agent to intervene and to break the bystander effect.

Many other potential ways to combat pluralistic ignorance could exist, the framework in principle also allows for a range of extensions that try to implement these methods, e.g. by adding a a larger variety of agent types or by introducing new action such as public broadcasts.

6 Implementation

To support the theory described in this report, an application was developed that simulates the steps defined in section 4 while keeping track of internal plausibility models. The goal of this part of the project was to develop a framework for illustrating how pluralistic ignorance can give rise to the bystander effect and for experimenting with different prevention mechanisms. To achieve this, the plausibility models and transition rules were modeled in an object oriented fashion in Java. The front end of the application provides an intuitive way of inspecting the epistemic state of the model in the different steps. The application depicts a situation in which several agents witness the occurrence of an event that requires an intervention. The setting that was chosen for the simulation is that of a firm, where several board members witness an event that renders the current business strategy inadequate, and which causes the firm to lose money. The agents are board members, who can intervene by objecting against the strategy.

The application shows a graphic illustration of the agents in a meeting room. The user can advance the simulation to the next step using a button. The information panel underneath the illustration provides a brief explanation of what happens in each step of the simulation. A more detailed and formal explanation of how the underlying knowledge structures are updated is given in section 4 of this report. Because the plausibility models describing the intermediate steps of the bystander effect get very large (up to more than 100 states for the 3 agent case), we do not explicitly show the action models in the application. Instead, the application shows the first and second order beliefs of the agents as well as the propositions and their valuation. The beliefs are drawn directly from the underlying EPM generated by the current step of the simulation, using the rules defined in section 2. The user can view the beliefs of the different agents in the model by clicking on the respective agent in the illustration. Furthermore, the user can toggle whether or not to show introspective beliefs, and beliefs about propositions corresponding to negated actions. By default, these beliefs are omitted, such that only those beliefs are shown which are most relevant for the occurrence of the bystander effect. Finally, the propositions field shows a list of all propositions in the current model, together with their description, and their truth value in the true state of the current EPM.

The user can vary the number of agents between 3 and 4. The minimum number of agents needed for the bystander effect to occur is 3. While in theory it would be possible to use the same knowledge structures for a number of agents larger than 4, this would drastically increase the complexity of the models and lead to excessively long computation times. Currently, all agents in the simulation are of the hesitator type. This design choice was made because the bystander effect only occurs when there are hesitators among the agents who witnessed the event. However, the data structures used in the implementation also support the other agent types, so the application could be easily extended to include city dwellers and first responders.

As an example of a method to prevent the bystander effect, a friendship tie (as introduced in Section 5.1) can be added between two agents in the simulation. When this option is selected, the model simulation will follow the same steps, but in the three agent case the outcome will be different as at least one agent will intervene after revising its social beliefs. The other methods for preventing the bystander effect are not included in this version of the application, but these are left as a possible direction for future work.

7 Discussion

Above we have discussed the bystander effect arising due to misinterpretation of social proof, potentially leading to a state of pluralistic ignorance followed by evasion of a problematic scenario. To model facet of the bystander effect we have used the model introduced in [9]. It uses static models that are updated with action model describing a certain change in the static state. In evaluation of the bystander effect a couple of key steps are nicely modelled in this framework: The initial state in which nothing happens, an ambiguous problematic event arises, agents decide what action to perform, agents misinterpret the actions of others, agents revise their believes on basis of this faulty social proof and then eventually all agents evade. Since the bystander can cause serious harm in certain scenarios, as in the drunk driving example we have explored in detail, we therefore explored ways to prevent the bystander effect and discussed how some of these can be implemented, by extending the model if necessary. Most importantly here were friendship ties, which give agents insight into each other actions, and education which can increase the number of first responders. We have also based an implementation on the framework and these steps, which explores the situation in which only hesitators are present and also implements friendship ties as discussed by us as a potential solution to the bystander effect. The application supports our theory and shows that friendship ties can indeed break the bystander effect.

Overall the framework from [9] forms a good basis to model the relevant type of bystander effect. One large drawback however is that the outcome is completely deterministic, different from what many empirical studies observe [13]. To get a more expressive model we have already argued that it might be favourable to change the notion of influenced believe described by operator $SB_{i|G}$ to include a larger variety in the properties of hesitators and also allowing the model to take into account the severity of a problematic event, where possibly in more severe events hesitators require less social proof to intervene. Another improvement, also already mentioned in [9], is that rather than using decision rules and interpretation rules to use a more powerful approach, e.g. drawing on frameworks from game theory. This would allow for a richer description of the bystander effect potentially taking into account different settings in which the effect takes place, e.g. in the sense of agent characteristics, event type, number of agents, prior beliefs or knowledge (for example when an agent has expertise involving the problematic event at hand),etc. Moreover it might also be interesting to include communication between agents, in a more in dept manner than what we described in Section 5.5. This would give a more realistic situation, considering humans frequently use their communication abilities in a large variety of scenarios in real life.

Finally if extra detail is added to the framework in the form of believes about other agents their types, also the bystander effect arising in the third step of the interpretation process discussed in the introduction could be modelled [9]. For example if hesitators believe a first responder is present, it might believe it is best for the first responder to intervene (first). But if this believe is false this could lead to the situation in which all hesitators believe an event is problematic, but they would still not intervene because they believe it is best a non existing first responder intervenes. Also this step of the bystander effect might benefit from including more detailed agent types which capture could capture expertise in a certain relevant field of study.

References

[1] U. S. D. of Transportation: National Highway Traffic Safety Administration, "https://www.nhtsa.gov/," accessed:1-6-2019:.

[2] B. Latané and J. M. Darley, *The unresponsive bystander: Why doesn't he help?* Appleton-Century-Crofts, 1970.

[3] J. B. Harvey, "The abilene paradox: the management of agreement.," *Organizational Dynamics*, 1974.

[4] D. A. Prentice and D. T. Miller, "Pluralistic ignorance and the perpetuation of social norms by unwitting actors," in *Advances in experimental social psychology*, vol. 28, pp. 161–209, Elsevier, 1996.

[5] N. Geiger and J. K. Swim, "Climate of silence: Pluralistic ignorance as a barrier to climate change discussion," *Journal of Environmental Psychology*, vol. 47, pp. 79–90, 2016.

[6] D. A. Prentice and D. T. Miller, "Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm.," *Journal of personality and social psychology*, vol. 64, no. 2, p. 243, 1993.

[7] H. J. O'GORMAN, "Pluralistic ignorance and white estimates of white support for racial segregation," *Public Opinion Quarterly*, vol. 39, no. 3, pp. 313–330, 1975.

[8] J. D. Westphal and M. K. Bednar, "Pluralistic ignorance in corporate boards and firms' strategic persistence in response to low firm performance," *Administrative Science Quarterly*, vol. 50, no. 2, pp. 262–298, 2005.

[9] R. K. Rendsvig, "Pluralistic ignorance in the bystander effect: Informational dynamics of unresponsive witnesses in situations calling for intervention," *Synthese*, vol. 191, no. 11, pp. 2471–2498, 2014.

[10] R. Hammack, W. Imrich, and S. Klavžar, *Handbook of product graphs*. CRC press, 2011.

[11] G. Katzenstein, "The debate on structured debate: Toward a unified theory," *Organizational Behavior and Human Decision Processes*, vol. 66, no. 3, pp. 316–332, 1996.

[12] B. Gavish and J. H. Gerdes Jr, "Anonymous mechanisms in group decision support systems communication," *Decision Support Systems*, vol. 23, no. 4, pp. 297–328, 1998.

[13] B. Latané and S. Nida, "Ten years of research on group size and helping.," *Psychological bulletin*, vol. 89, no. 2, p. 308, 1981.