Pluralistic Ignorance in the Bystander Effect
Informational Dynamics of Unresponsive Witnesses in Situations calling for Intervention

Rasmus K. Rendsvig

Department of Media, Cognition and Communication
University of Copenhagen

Link to published version on SpringerLink

Abstract. The goal of the present paper is to construct a formal explication of the pluralistic ignorance explanation of the bystander effect. The social dynamics leading to inaction is presented, decomposed, and modeled using dynamic epistemic logic augmented with 'transition rules' able to characterize agent behavior. Three agent types are defined: First Responders who intervene given belief of accident; City Dwellers, capturing 'apathetic urban residents' and Hesitators, who observe others when in doubt, basing subsequent decision on social proof. It is shown how groups of the latter may end in a state of pluralistic ignorance leading to inaction. Sequential models for each agent type are specified, and their results compared to empirical studies. It is concluded that only the Hesitator model produces reasonable results.

Keywords: bystander effect, pluralistic ignorance, social dynamics, social proof, social influence, dynamic epistemic logic

1 Introduction: The Bystander Effect and Pluralistic Ignorance

On March 13, 1964, in Queens, New York, Catherine Susan "Kitty" Genovese was raped and stabbed, the assailant fleeing multiple times during the ongoing assault that resulted in Genovese's death. Multiple residents witnessed parts of the nearly one hour long attack, without successfully intervening.

The foremost explanation put forth in the ensuing media coverage was *apathy* among urban citizens. Through the pressures of city life, "*homo urbanis*" has lost his sense of empathy for fellow man, all but grown indifferent to their quarrels (Latané and Darley, 1970, 1968). This explanation struck social psychologists John M. Darley and Bibb Latané as incorrect, whereupon they set out to provide an alternate explanation. The first publication of experimental results is the classic (Latané and Darley, 1968), which was soon followed by a vast amount of studies (see Latané and Nida (1981) for a review). In these studies, focus has been moved from why *urban citizens* do not help, to why *people in groups* are less prone to help. The collected data shows a robust tendency, namely that the chance of help being offered diminishes as the number of witnesses increases. This tendency will be referred to as *the bystander effect*.

An Explanation of the Intervention Process. One currently used text-book explanation (see e.g. (Myers, 2012)) of the bystander effect stems from (Latané and Darley, 1970), and involves three steps which each bystander must go through before he or she will intervene.[1] First, a bystander must *notice the event* in question. With busy street life involving traffic and pedestrians, the risk of overlooking a seizure is higher than on desolate streets. Where the problematic situation goes unnoticed, help will not be offered. Second, if noticed, the bystander must *interpret the event* in order to decide whether an emergency is occurring or not. In many cases, this will not prove to be a problem: situations involving car accidents or bleeding victims are seldom epistemically ambiguous. However, a man slumped on a bench may provide an epistemic conundrum, as he may be merely mumbling curses against the general youth, marooned following a too *Saké* intense business lunch, or moaning in pain from the onset of a seizure. Such ambiguities may be sought resolved by the acquisition of further information, readily present in the form of *social proof*: if other bystanders are not showing signs of distress, the event will be perceived as less critical and therefore ultimately bypassed. Third, in case the event is interpreted as requiring help, the bystander has to gauge *whether to take responsibility*: when alone, there is no question as to who should intervene, but when gathered in groups, *diffusion of responsibility* may arise. Such diffusion may be caused by uncertainty as to whether we are among the best qualified to handle the situation, whether others have already called for paramedics or are just about to act. When alone, the responsibility to intervene rests on one individual, but when in a group, the same pressure is apparently distributed among all, thereby diminishing the chances that anyone will act.

Pluralistic Ignorance. The goal of this paper is to model the social informational dynamics and decision procedures running the second of these steps, specifying conditions under which a group of agents in an ambiguous situation may choose to seek social proof in order to individually determine a correct course of action and the associated consequences thereof, hereby providing a detailed explanation for (this part of) the bystander effect. This narrower focus is taken as the second step of the bystander effect explanation constitutes an interesting informational dynamics in its own right, useful to the analysis of social situations in which neither distractions nor diffusion play important roles.

The second step of the dynamics revolve around a *belief state* often referred to as one of *pluralistic ignorance*: a situation in which everybody believes that everybody else believes a given proposition/endorses a given norm, while no-one in fact believes it/endorses it.[2] Pluralistic ignorance has been put forth as a decisive factor in a plethora of social situations, including the introduction of various unpopular norms such as college binge drinking (Prentice and Miller, 1993) and violent gang behavior (Bicchieri and Fukui, 1999), the persistence of poor strategies in light of poor firm performance (Westphal and Bednar, 2005) and lack of help seeking in class rooms (Miller and Mc-

[1] For a comprehensive walk-through of this explanation and supportive data, refer to (Latané and Darley, 1970).

[2] (Halbesleben and Buckley, 2004) provides an illuminating overview of the history and development of the term.

Farland, 1987); the allowance of ongoing mortgage deed merry-go-rounds in Denmark during the financial 'upswing' of 2007 (Hendricks and Rasmussen, 2012; Hansen et al., 2013).

Pluralistic ignorance may cause individuals not to act for more than one reason. One may be due to social inhibition – you may not wish to be the only one raising your hand to ask a question. Here, inaction results from vanity and social identity. There may be *doubt* – you may not wish to call the police if there is no cause for alarm. In case of doubt, inaction follows from incorrect information processing. Both causes may individually lead to inaction and may further co-occur. In the following, 'pluralistic ignorance' will be used to refer only to processes of the latter kind.

Structure of the Paper. In Section 2, an example of a social dynamics involving pluralistic ignorance which lead to unfortunate inaction is presented, and an informal sketch of the information processing involved is outlined, the structure of which is used as a guideline for the formal representation. In Section 3, elements from dynamic epistemic logic (DEL) are presented. DEL is the primary modeling tool, used to represent static epistemic states and belief revision in light of new information. Section 3 also presents the modeling of the occurrence of 'the accident'. In Section 4, it is shown how *agent types* may be defined for DEL, hereby augmenting the framework with a notion of *choice agency*. Three agent types relevant to the bystander effect are defined: First Responders, a 'good samaritan' type agent, who will choose to intervene in emergencies if she believes one such is occurring; City Dwellers, capturing the 'apathetic urban resident' and Hesitators, who observes others when in doubt. In Section 5, it is shown how Hesitators' *misinterpretation* of other Hesitators' choice to observe may lead to a state of pluralistic ignorance. Coupled with Hesitators basing subsequent action on social proof obtained through observation, it is further shown that this agent type will choose to evade the scene of the accident, though they privately believe there is cause for intervention. In Section 6, models run with each of the three agent types are compared to empirical studies. It is concluded that the Hesitator model is the only of the three producing reasonable results. In Section 7, we conclude.

2 From Ambiguity to Inaction

Consider the following example, inspired by (Halbesleben and Buckley, 2004). A firm is performing poorly and this is caused by the currently implemented business strategy. Every member of the board of directors has equal access to the relevant information pertaining to strategic choice and firm performance, which to each member strongly indicates that the prudent action is to change strategy. The available information is, however, not conclusive: the possibility that a strategy change will leave the firm performance-wise worse off cannot be ruled out. Hence, every board member is in an epistemically ambiguous situation, where they privately believe intervention will be fruitful, but all would prefer additional information before settling for a vote against the status quo. As the voting situation arises, all seek such further information from their peers, hoping that the votes of others will illuminate them. As all look to each other, nobody initially raises an objection, which is interpreted by others as the choice that no objection need be raised. Hereby, all conclude that their intelligent peers believe that the status quo should not be changed. This perceived consensus among peers is then seen as providing evidence to the conclusion that the currently implemented strategy is in fact desirable. Having thus reflected, board members choose not to intervene in the status quo, and the low firm performance continues another term or two. Though a fictitious example, pluralistic ignorance does occur on corporate boards, *often resulting in poor strategic decisions* (see (Westphal and Bednar, 2005)).

While the example differs in topic from a street accident, the two share common information dynamics. The street incident revolves around the commonly unwanted case of physical pain, the board meeting example focuses on the commonly unwanted event of a sub-optimal status quo strategy. Both involve uncertainty about whether an unwanted situation is the case or not, both require intervention in case the unwanted situation indeed is occurring, and both allow for the gathering of further information by observing peers.

Notice how the example implicitly utilizes *two* instances of pluralistic ignorance. First, there is one instance of what may be called *norm-based* pluralistic ignorance: though every member uses the decision rule "if in doubt, seek further information", they assume that others will follow a different rule, namely "if in doubt, raise an objection". This is an instance of pluralistic ignorance as everybody believes that everybody else follows a given norm (here, a decision rule), while in fact no-one follows it. The second instance of pluralistic ignorance is *proposition-based*: everybody ultimately believes that everybody else believes that status quo is fine, while everybody privately believes the strategy should be changed.

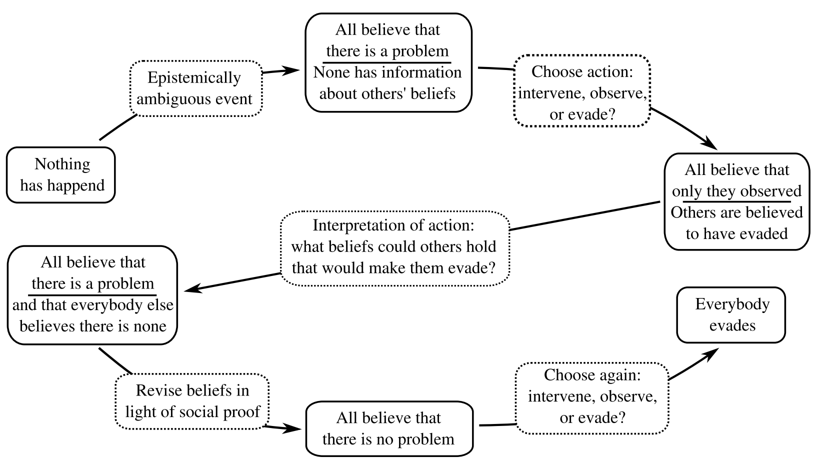

Fig. 1. Flowchart of the dynamics leading to the bystander effect. Boxes with solid lines represent epistemic states – boxes with dotted lines indicate events.

The dynamics involved in the example may be decomposed into eleven elements (see Fig. 1): six static states and five state-altering transitions. To start from scratch, in the first state, nothing has happened (1). This is followed by the occurrence of an event, epistemically ambiguous between being an accident (stabbing, onset of poor performance) or nothing of consequence (dispute, reasonable performance) (2). This event results in a second state, where everybody privately believes that an accident occurred, while remaining ignorant about the beliefs of others (3). Based on this state, one may choose to intervene (rush to help, object to strategic choice), may choose actively to evade (ignore the stabbing, withhold objection), or may choose to seek further information by observing the choice made by others. Crucially, the performed actions of evasion and observation are here considered to be epistemically ambiguous: in seeking further information, we do not want to flaunt our ignorance, so further observation is made discretely. Given the ambiguity of the accident, observation is chosen and executed (4). It is claimed here that a crucial further element for the dynamics is the resulting *mis-perception* of this choice when made by others: though we ourselves may choose to observe, when we see others do the same, we consider it plausible that they in fact chose to evade. Given this norm-based pluralistic ignorance, in the ensuing third state (5), all still believe there was an accident, while believing that all others evaded. To obtain information about the beliefs of others, their perceived actions must now be *interpreted* (6): given that you evaded, what may I conclude about your beliefs pertaining to the accident? Under the assumption that you are a reasonably decent person, only that you believed there was none. Such interpretations conducted by all then results in a fourth state (7) of proposition-based pluralistic ignorance: though we all believe there was an accident, we also believe that no-one else believes so. Revising our beliefs in the light of the obtained social proof (8), all conclude that no accident occurred (9). Given a further chance to act (10), evasion will be the natural choice, leading to the final state (11), where the accident in fact occurred, everybody believed so, but nevertheless chose to evade it, due to the social information dynamics.

3 Plausibility Models for States and Actions

The sketch presented above suggests several ingredients required for a suitable model, including propositional- and higher-order beliefs (beliefs about beliefs), belief change in light of new information and agent action. To that end, dynamic epistemic logic[3] with updating by action models with postconditions is a suitable framework: All higher-order beliefs are specified in relatively small models, factual change may be modeled using postconditions, and the dynamics may be built step-by-step, allowing for a detailed overview of each step. Step-by-step construction allows for easy replacement of single ‘modules’, whereby alternative runs may be investigated. For each such run, the dynamics terminate when either someone agent intervenes, or all agents choose to evade. The dynamics presented concern a three agent case, with group size variations presented in Section 6. Throughout, the same complete graph ‘all see all’ social network structure is assumed.

[3] Technically, no *logic* is introduced; the dynamics are investigated using only model theory.

3.1 Statics

Multi-agent Plausibility Frames. Where $\mathcal{A}$ is a finite set of *agents*, a *multi-agent plausibility frame* (*MPF*) is a structure $\mathsf{S} = (S, \leq_i)_{i \in \mathcal{A}}$ where S is a finite set of *states* and each $\leq_i$ is a well-preorder[4].

The idea behind plausibility frames is that such encode the knowledge and beliefs of a group of agents, $\mathcal{A}$, capturing which states each agent may tell apart, and how plausible these states are relative to one another. If two states s,t are connected by $\leq_i$, then i cannot tell these states apart, but if $s <_i t$ (i.e. $s \leq_i t$ and $t \not\leq_i s$), then i considers s *more plausible* than t.[5] Fig. 2 illustrates a simple plausibility frame F_1 with two states, s and t, and two agents, a and b. The arrow from t to s captures that $s <_b t$, i.e. that b cannot distinguish between s and t, but finds s strictly more plausible.[6] Reflexive arrows will only be drawn if they are the only arrows for a given agent. In Fig. 2, a cannot tell s from s nor t from t.

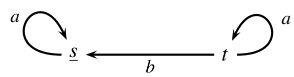

Fig. 2. A simple, two agent plausibility frame, denoted F_1.

Indistinguishability Relation; Information and Plausibility Cells. Given an MPF $\mathsf{S} = (S, \leq_i)_{i \in \mathcal{A}}$, the *indistinguishability relation* for agent i is the equivalence relation $\sim_i := \leq_i \cup \geq_i$. Further, the *information cell* of agent i at state s is $\mathcal{K}_i[s] = \{t : s \sim_i t\}$ and the *plausibility cell* of agent i at state s is $\mathcal{B}_i[s] = Min_{\leq_i} \mathcal{K}_i[s] = \{t \in \mathcal{K}_i[s] : t \leq_i s', \text{ for all } s' \in \mathcal{K}_i[s]\}$. The plausibility cell $\mathcal{B}_i[s]$ contains the worlds the agent find *most plausible* from the information cell $\mathcal{K}_i[s]$ and represent the "doxastic appearance" (Baltag and Smets, 2008, p. 25) of s to i.[7] Notice that $s \leq_i t$ means that s is at *least as plausible* as t for i.

In Fig. 2, it is the case that $s \sim_a s$ and $t \sim_a t$, but not $s \sim_a t$. Hence $\mathcal{K}_a[s] = \{s\}$ and $\mathcal{K}_a[t] = \{t\}$. For agent b, $s \sim_b s$, $t \sim_b t$ and $s \sim_b t$, entailing that $\mathcal{K}_b[s] = \mathcal{K}_b[t] = \{s,t\}$. However, even if the actual state is t, agent b finds s more (most) plausible, so it is the sole element in b's plausibility cell at t: $\mathcal{B}_b[t] = \{s\}$.

Doxastic Propositions. When considering plausibility frames, sets of states may be identified with propositions; e.g. the set $\{s\}$ could be identified with the proposition that intervention is desirable, and $\{t\}$ with the same proposition's negation. In this case, Fig. 1 would represent a situation in which 2 finds it more plausible that intervention is desirable than that it is not, while 1 would know whether or not this was the case.

More specifically, let a *doxastic proposition* (henceforth just *proposition*) be a map P that assigns to every MPF S with state-space S a subset $(P)_{\mathsf{S}} \subseteq S$.[8] Denote *true*, $\top$, *false*, $\bot$, and Boolean operations for arbitrary propositions P and Q by

[4] Reflexive and transitive binary relation where every non-empty subset has a minimal element, cf. (Baltag and Smets, 2008).

[5] The relation $\leq_i$ may therefore more appropriately be thought of as an *implausibility* relation, where $s \leq_i t$ is read 't is as implausible as s, or more so'.

[6] A heuristic to aid recall is that $s < t$ in form is similar to $s \leftarrow t$ when looking at the arrowhead. When $s \leq_i t$ and $t \leq_i s$, arrowheads are omitted altogether.

[7] The notation for information and plausibility cells are adopted from (Dégremont, 2010).

[8] Parentheses will be omitted where no confusion should arise.

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

Propositions with epistemic and doxastic modalities are given by

$$(K_iP)_{\mathsf{S}} := \{s \in S : \mathcal{K}_i[s] \subseteq P_{\mathsf{S}}\}$$
$$(B_iP)_{\mathsf{S}} := \{s \in S : \mathcal{B}_i[s] \subseteq P_{\mathsf{S}}\}$$

The Boolean case simply follows the immediate set-theoretic interpretation. Propositions with epistemic and doxastic operators represent statements of knowledge and belief: K_iP / B_iP reads 'agent i knows / believes that P', and $(K_iP)_{\mathsf{S}}$ / $(B_iP)_{\mathsf{S}}$ are the sets of states of these propositions given an MPF S. For knowledge, the definition entails that $s \in (K_iP)_{\mathsf{S}}$ iff for all t in i's information cell relative to s, t is a P-state. Belief has the same reading, but restricted to the plausibility cell for i at s.[9]

Example: King or Queen? Assume a situation with two players, a and b, where a has one card in hand, being either a King or a Queen. Let Q be the doxastic proposition "a has a Queen on hand" and K ditto for King. Over F_1, Fig. 1, set $(Q)_{\mathsf{F}_1} = \{s\}$ and $(K)_{\mathsf{F}_1} = \{t\}$ (hence $(\neg Q)_{\mathsf{F}_1} = (K)_{\mathsf{F}_1}$). In this case, $(K_aQ)_{\mathsf{F}_1} = \{s\}$, $(K_bQ) = \emptyset$ and $(B_bQ)_{\mathsf{F}_1} = \{s,t\}$. That is, in state s, agent a knows Q, in *no* state does b know Q, but b believes Q in both states. The structure thus models a situation in which a knows whether she holds a Queen, while b only has a belief this regarding. Whether this belief is correct depends on which of the two states is the *actual*. Combined with such a valuation of propositions, a plausibility frame is called a *model*. Denote F_1 with the described valuation $\mathbf{S_{Q/K}}$.

Epistemic Plausibility Models. Let a *valuation set* be a set Φ of doxastic propositions, considered the *atomic* propositions. An *epistemic plausibility model* (*EPM*) is an MP frame together with a valuation set Φ, denoted $\mathbf{S} = (S, \leq_i, \Phi)_{i\in\mathcal{A}}$. For $s \in P_{\mathbf{S}}$, write $\mathbf{S}, s \models P$, and say that P is *true* or *satisfied* at state s in model $\mathbf{S}$. A *pointed EPM* $\mathbf{S} = (S, \leq_i, \Phi, s_0)_{i\in\mathcal{A}}$ is an EPM with a designated state $s_0 \in S$, called the *actual state*. Where $s_0 \in P_{\mathbf{S}}$, write $\mathbf{S} \models P$.

EPMs and Kripke Models. Where epistemic plausibility frames are special instances of *Kripke frames* (see e.g. (Blackburn et al., 2001)), epistemic plausibility models are not special instances of *Kripke models*.[10] However, every EPM $\mathbf{S}$ gives rise to a Kripke model M_K. First, let Φ' be Φ where the functional nature of each doxastic proposition is ignored. Φ' may then be treated as a set of *atomic proposition symbols*. Second, define a *valuation map* $\|\cdot\| : \Phi' \longrightarrow \mathcal{P}(S)$, assigning to the elements of Φ' a set of states from

[9] The reason for using this atypical definition of propositions is that it allows us to speak about the *same* proposition across multiple models. This is practical as model transformations will play a large role in the latter. A Kripke model valuation may easily be extracted from set of doxastic atomic propositions; see (Baltag and Smets, 2008) for details.

[10] As opposed to the terminology of (Baltag and Smets, 2008; van Benthem, 2007; Demey, 2011) where epistemic plausibility models are Kripke models.

the state-space of the underlying frame. Simply let M_K and $\mathbf{S}$ be based on the same frame, and let the valuation map $\|\cdot\|$ for M_K be given by $\|P\| := P_{\mathbf{S}}$, for all $P \in \Phi$. The alternative definition of EPMs is used as it is natural when dealing with doxastic propositions rather than a syntactically specified language.

Relevant Propositions and the Initial State. To model the second step of the bystander effect dynamics for three agents, ten atomic propositions are required. First, use A to denote that an accident has occurred. This is the basic fact about which the agents must establish a belief. Second, each agent $i \in \mathcal{A} = \{a, b, c\}$ must choose one of three actions: either to intervene, I_i, to observe, O_i, or to evade the scene, E_i. The set of these atoms is denoted Φ. As the model constructed is temporally simple, the propositions are best read as "agent i has intervened / observed / evaded". It is assumed that no agent can perform two actions simultaneously, i.e. that $I_i \cap O_i = I_i \cap E_i = O_i \cap E_i = \emptyset$. Denote the set of all doxastic propositions obtainable from Φ and the above construction rules by Prop_Φ.

The initial state (where nothing has happened) may now be represented by the EPM $\mathbf{S_0}$, Fig. 3.

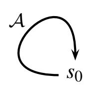

Fig. 3. The initial state where nothing has happened; all atoms are set to false: $(A)_{\mathbf{S_0}} = (I_i)_{\mathbf{S_0}} = (O_i)_{\mathbf{S_0}} = (E_i)_{\mathbf{S_0}} = \emptyset$

In this simple state, every agents knows *exactly* what has transpired so far: nothing. All propositions are false, everybody knows this, which is again known by all, etc. Among others, it is the case that $\mathbf{S_0}, s_0 \models \neg(A \vee I_1 \vee O_2 \vee E_3) \wedge \bigwedge_{i \in \mathcal{A}}(K_i(\neg A \wedge \bigwedge_{j \in \mathcal{A}} K_j \neg A))$.

The end conditions of runs mentioned on page 5 may now formally be specified: identify the *end of a run* with any EPM satisfying at its actual state either $\bigvee_{i \in \mathcal{A}} I_i$ or $\bigwedge_{i \in \mathcal{A}} E_i$, capturing respectively that at least one agent intervenes, or all evade.

3.2 Changing Models: Action Models and Action-Priority Update.

To capture factual and informational changes that occur due to events, a static epistemic plausibility model may be transformed using an *action model*, capturing the factual and epistemic representation of the event, and the *action-priority update product*. The guiding idea is that an action model encodes the belief and knowledge agents have about an *ongoing* event, the information from which is combined with the static model by taking the two models' *product*: the result is a new static model in which the agents' new information takes priority over that of the previous static model. The present formulation rests on (Baltag and Smets, 2008), with the addition of *postconditions*, as used in (van Ditmarsch and Kooi, 2008; Bolander and Birkegaard, 2011). The latter allows action models to not only change the knowledge and belief of agents, but also effectuate *ontic*[11] *changes*, needed when the environment or agents perform actions.

[11] Ontic facts are all non-doxastic facts, i.e. propositions that do not contain belief or knowledge operators.

Action Plausibility Models. A *(pointed) action plausibility model* (*APM*)

$$\mathbf{E} = (\Sigma, \leq_i, pre, post, \sigma_0)_{i \in \mathcal{A}}$$

is an MP frame $(\Sigma, \preceq_i)_{i \in \mathcal{A}}$ augmented with a *precondition map*, $pre : \Sigma \longrightarrow \mathsf{Prop}_\Phi$ and a *postcondition map* $post : \Sigma \longrightarrow \mathsf{Prop}_\Phi$ such that $post(\sigma) = \psi$ where $\psi \in \{\top, \bot\}$ or $\psi = \bigwedge_{i=1}^{n} \varphi_i$ with $\varphi_i \in \{P, \neg P : P \in \Phi\}$. Finally, $\sigma_0 \in \Sigma$ is *the actual event*.

Just as every world in an EPM represents a possible state of affairs, specified by the world's true propositions, so every action in an APM represents a possible *change*. *What* change is specified by the pre- and postconditions; preconditions determine what is required for the given action to take place, i.e. what conditions a world must satisfy for an action to be executable in that world, and postconditions what *factual* change the action brings about.

Example: King or Queen?, cont. Continuing the example above, let now a play her card face down. Assuming that a knows which cards she is playing, this situation may again be presented by the MPF F_1 in Fig. 1, with s representing the action 'a plays a Queen' and t 'a plays a King'. The precondition for s, that a is playing a Queen, is Q, that a has a Queen on hand, and *vice versa* for t and K. Hence $pre(s) = Q$ and $pre(t) = K$. Following the play, a will either no longer have a Queen on hand, or no longer have a King. Hence, $post(s) = \neg Q$ and $post(t) = \neg K$. With these pre- and postconditions, F_1 is an action model, call it $\mathbf{E_{Q/K}}$, representing the event where a is certain that she is playing a Queen, while b is uninformed about which of the two plays is the actual, finding it more plausible that a plays Q.

Doxastic Programs. Where $\mathbf{E_{Q/K}}$ represents a situation where a plays her card face down, a show-and-tell play by a is captured by the two strict subsets of the model, i.e. by the *doxastic programs* $\Gamma_Q = \{s\}$ and $\Gamma_K = \{t\}$. A doxastic program is the action model equivalent of a proposition, i.e. a subset of all actions in the models' event space: $\Gamma \subseteq \Sigma$. Over $\mathbf{E_{Q/K}}$, the program Γ_Q captures the event where a plays Queen and b sees this, and Γ_K the same for a playing King. In the ensuing, it will be assumed that doxastic programs contain the actual action.

An Accident Occurs. What is a suitable APM capturing both the factual change that the accident occurs, as well as a, b and c's information about this? Given that the accident in fact occurs, it is clear that the actual event σ_0 of the model must change the truth value of A from false in $\mathbf{S_0}$ to true in the ensuing EPM $\mathbf{S_1}$. Further, no agents perform actions during the event, so $post(\sigma_0) = A$.

Focusing on a, then how does she perceive the occurrence of accident? As "[m]ost emergencies are, or at least begin as, ambiguous events" (Latané and Darley, 1968, p. 216) a will at least be uncertain regarding whether it occurs or not, and therefore considers an alternative event τ_0 with $post(\tau_0) = \top$ possible.[12] Moreover, *ex hypothesi*, her perception of the event indicates that in fact A, so $\sigma_0 \prec_a \tau_0$. How does a perceive

[12] The postcondition $\top$ leaves all atomic propositions as they were in the previous model. This is specified by the action priority update product below.

that b and c perceive the accident? Not being telepathic, a cannot tell, and she considers it possible that both, neither, or either of b and c perceive the event as she does. It is assumed, though, that a perceives b and c during the event *as forming an opinion about whether or not A*. This assumption is made for two reasons: 1) it produces a smaller model, and 2) pertaining to social proof, only agents perceived as informed are interesting from a's point of view. Variations to this assumption would be interesting, but are not dealt with here. Finally, a must consider it possible that b and c are wrong about the way a perceives the event. Taking this into consideration, Fig. 4 illustrates a's doxastic perception of the accident.

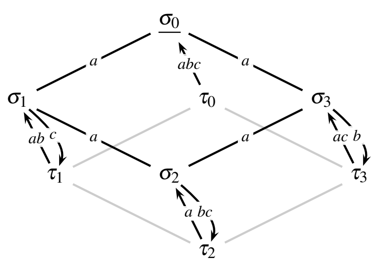

Fig. 4. Agent a's perception of the accident, σ_0. All σ_k actions have $post(\sigma_k) = A$; for all τ_k, $post(\tau_k) = \top$. Some links are gray only for presentation; they mirror the labels above. Diagonal links implied by transitivity are omitted; so are many links for b and c (see Fig. 5).

Agent a considers it more plausible that an accident is occuring, but is (at this point) agnostic about the beliefs of her peers; she finds it possible they all 'agree' (σ_0, τ_0), that she agrees with only b (σ_1,τ_1) or with only c (σ_3,τ_3), or that both b and c perceive the ongoing event as non-hazardous (σ_2,τ_2). Finally, she cannot rule out that b and c both find the event unproblematic and that they perceive a as doing the same (σ_7,τ_7).

Assuming that b and c perceive the accident in an identical manner, the model in Fig. 3 may be suitably duplicated and combined, resulting in the joint model $\mathbf{E_0}$, Fig. 5. Notice that b and c's perceptions are identical to a's. The only states not obtained from a duplication of Fig. 3 are σ_7 and τ_7. In fact, no one considers these possible, but neither can anyone rule out that others entertain them.

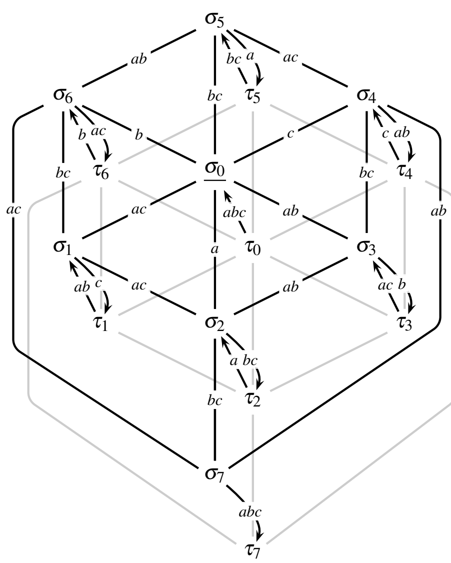

Fig. 5. The EPM denoted $\mathbf{E_0}$, representing the joint perception of the occurrence of an accident for three agents a, b and c. All σ_k actions have $post(\sigma_k) = A$; for all τ_k, $post(\tau_k) = \top$. Grey links are only for presentation; they mirror the labels above. Links implied by transitivity are omitted, e.g. between σ_6 and σ_4 for a. In essence, $\mathbf{E_0}$ captures 1) that the accident in fact occurs, 2) that this is ambiguous for every agent, who all find it more plausible that it does occur and 3) that no agents learns anything about others' perception of the event, except that *all have an opinion as to whether or not the accident occurred*.

To incorporate the (new) information from an action model or a doxastic program in an EPM, the two must be combined. A natural procedure for doing so is the *action-priority update product* (Baltag and Smets, 2008).

Action-Priority Update Product. The *action-priority update* is a binary operation $\otimes$ with first argument an EPM $\mathbf{S}$ with relations $\leq_i$ and second argument a doxastic program $\Gamma \subseteq \Sigma$ over some APM $\mathbf{E}$ with action space Σ and relations $\preceq_i$. The *APU product* is an EPM

$$\mathbf{S} \otimes \Gamma = (S \otimes \Gamma, \leq_i^{\uparrow}, \Phi^{\uparrow}, (s_0, \sigma_0))$$

where the updated state space is $S \otimes \Gamma = \{(s, \sigma) \in S \times \Gamma : \mathbf{S}, s \models pre(\sigma)\}$; each updated pre-order $\leq_i^{\uparrow}$ is given by $(s, \sigma) \leq_i^{\uparrow} (t, \tau)$ iff either $\sigma \prec_i \tau$ and $s \sim_i t$, or else $\sigma \simeq_i \tau$ and $s \leq_i t$;[13] the valuation set $\Phi^{\uparrow}$ is identical to Φ, with the requirement that for every atom $P \in \Phi$,

$$P_{\mathbf{S} \otimes \Gamma} = \{(s, \sigma) : s \in P_{\mathbf{S}} \text{ and } post(\sigma) \not\models \neg P\} \cup \{(s, \sigma) : post(\sigma) \models P\}$$

for states $(s, \sigma) \in S \otimes \Sigma$. Finally, (s_0, σ_0) is the new actual world.

[13] $\preceq_i$ is from $\mathbf{E}$ and $\leq_i$ from $\mathbf{S}$. $\sigma \prec_i \tau$ denotes ($\sigma \preceq_i \tau$ and not $\sigma \succeq_i \tau$), $\sigma \simeq_i \tau$ denotes ($\sigma \preceq_i \tau$ and $\sigma \succeq_i \tau$).

The APU product gives priority to new information encoded in Γ over the old beliefs from $\mathbf{S}$ by the anti-lexicographic specification of $\leq_i^{\uparrow}$ that gives priority to the APM plausibility relation $\preceq_i$. The definition further clarifies the role of pre- and postconditions; if a world does not satisfy the preconditions of an action, then the given state-action pair does not survive the update, and if postconditions are specified, these override earlier ontic facts, else leave all as was.[14]

Example, concl.: King or Queen? The factual and doxastic consequences of a's play of the Queen is calculated by finding the APU product of $\mathbf{S_{Q/K}}$ and $\mathbf{E_{Q/K}}$. The result is (again) an EPM $\mathbf{S_{Q/K}} \otimes \mathbf{E_{Q/K}}$ with underlying frame F_1, with state space $\{(s,s),(t,t)\}$. The state (s,s) 'survives' as s from $\mathbf{S_{Q/K}}$ satisfies the preconditions of s from $\mathbf{E_{Q/K}}$, while (s,t) does not, as $pre(t) = K$ while $(K)_{\mathbf{S_{Q/K}}} = \{t\}$. Further, in the actual state, Q changes truth value from true in $s \in \mathbf{S_{Q/K}}$ to false in $(s,s) \in \mathbf{S_{Q/K}} \otimes \mathbf{E_{Q/K}}$ as a result of the postconditions of $s \in \mathbf{E_{Q/K}}$. Finally, the information of the agents have changed: e.g., b now *knows* that a does not hold a Queen ($(s,s) \models K_b \neg Q$).

Updating with the Accident. Updating the simple initial state model $\mathbf{S_0}$ with accident event model $\mathbf{E_0}$ results in a state of great uncertainty. In summary, every agent believes that A: an accident has occurred (though no-one *know*s), all know that their peers either believe or disbelieve A (none are indifferent between A and $\neg A$ states), and all consider it possible that *the others* consider it possible that all believe there is no accident.

Formally, updating the simple structure of $\mathbf{S_0}$ with $\mathbf{E_0}$ produces the EPM $\mathbf{S_1} := \mathbf{S_0} \otimes \mathbf{E_0}$ which shares frame with $\mathbf{E_0}$ (Fig. 5) and has (s_0, σ_0) as actual state. In $\mathbf{S_1}$, for all $i \in \{0,...,7\}$, $(s_0, \sigma_i) \in A_{\mathbf{S_1}}$, and $(s, \tau_i) \in (\neg A)_{\mathbf{S_1}}$. Among others, the following doxastic propositions are true at (s_0, σ_0): A, $\bigwedge_{i \in \mathcal{A}} B_i A$, $K_a(B_b A \vee B_b \neg A)$, $\neg K_a B_b \neg A$.

Based on this second static state, the agents must make their first decision, as specified by their *decision rules*, which jointly determine the *type* of agent they are.

4 Decisions: Agent Behavior Characterized by Transition Rules

Though agent action may be represented in the introduced DEL framework using suitable atoms and postconditions, the notion of agency in DEL is purely *doxastic*. To move from only *believing* agents to *acting* agents, a richer framework is called for. One possibility would be to introduce a game- or pay-off structure in parallel to the DEL framework or embed the entire dynamics modeled in a temporally extended game tree, whereby actions could be made 'rationally', based on utility maximization at end nodes. A drawback to this method is the large models required: every branch must be fully specified before decisions may follow. Further, considering all possible branches is a cognitively complex task, making the approach empirically unrealistic.

Instead, an alternative approach involves utilizing 'rule of thumb' decisions, brute-forced by the current beliefs of agents. This method, detailed below, forfeits "rational" decisions, but overcomes the two drawbacks of the game theoretic approach by letting

[14] The definition is based on (Baltag and Smets, 2008) for the anti-lexicographic order, adding postconditions from (van Ditmarsch and Kooi, 2008; Bolander and Birkegaard, 2011).

choice be dictated locally by current beliefs. Given an EPM, a set of doxastic programs provides a multitude of possible updates. In modeling a dynamic process, the modeler must choose which model is suitable for the next update, based on no strict directions from the to-be-updated EPM. However, environment or agent behavior will often be seen as dictated to some degree by facts or beliefs from the current EPM, thus used as a guideline. To incorporate the next action model choice in a formal manner, *transition rules* are introduced, locally specifying the next update as a function of the current EPM.

Transition rules are used to characterize agent behavior. Each behavior is specified by a set of transition rules, each with a *trigger condition* and a *goal formula*. If an EPM satisfies some trigger conditions, the ensuing EPM must satisfy the matching goal formulas. An APM that ensures that the goals are obtained then satisfies, or *solves*, the rules, and is seen as a possible choice for the agent in question. Hereby an EPM, a set of behavior-governing rules and a set of APMs jointly specify the transition to the next EPM.

Transition Rules. A *transition rule* $\mathcal{T}$ is an expression $\varphi \rightsquigarrow [X]\psi$ where $\varphi, \psi \in \mathsf{Prop}_{\Phi}$. Call φ the *trigger* and ψ the *goal*. If EPM $(\mathbf{S}, s_0)$ satisfies the trigger of a transition rule $\mathcal{T}$, $\mathcal{T}$ is said to be *active* in $\mathbf{S}$ (else *inactive*).

Specified below, transition rules may be used to choose the next update based on local conditions of the current EPM. E.g., updates by the 'environment' may be specified using atoms in the trigger. To exemplify, let R and W be atoms with resp. readings 'it rains' and 'the street is wet', then the transition rule $\mathcal{T}_1 = R \rightsquigarrow [X]W$ reads 'if it rains, then the next update must be such that after it, the street is wet'. Transition rules may also be used as agent decision rules for factual change, using $B_i\varphi/K_i\varphi$-formulas as triggers and suitable formulas as effects. E.g., the set of transition rules $\{B_iR \rightsquigarrow [X]U_i, B_i\neg R \rightsquigarrow [X]\neg U_i\}$ may be used to specify agent behavior relative to rain: if i believes it rains, then next i will have an umbrella, and if i believes it does not rain, then next i will not have an umbrella. Used thus, transition rules are akin to the *programs* and *knowledge-based programs* of (Fagin et al., 1995), here tailored to the DEL framework. They further instantiate one-step *epistemic planning problems*, in the terminology of (Bolander and Birkegaard, 2011).

Dynamic Modalities. Note that *transition rules are not doxastic propositions*: the "modality" $[X]$ has no interpretation, and construed as a formula, $\mathcal{T}_1$ has no truth conditions. Instead, transition rules are *prescriptions* for choosing the next action model. The choice of model is made by implementing a transition rule over an EPM $\mathbf{S}$ and a set $\mathbf{G}$ of doxastic programs over one or more APMs using *dynamic modalities.*

For any program Γ over APM $\mathbf{E}$, $[\Gamma]$ is a dynamic modality, and the doxastic proposition $[\Gamma]\varphi$ is given by

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

That is, a state s from $\mathbf{S}$ is a $[\Gamma]\varphi$-state iff every resolution of Γ over s is a φ-world in $\mathbf{S} \otimes \Gamma$. $[\Gamma]$-modalities are natural when φ is desired common knowledge among $\mathcal{A}$.

Further, where a Γ is doxastic program, let $[\Gamma]_i\varphi$ be given by

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

That is, a state s from $\mathbf{S}$ is a $[\Gamma]_i\varphi$-state iff every resolution of Γ over *s that is included in i's information cell relative to the actual world* (s_0,σ_0) *in* $\mathbf{S}\otimes\Gamma$ is a φ-world in $\mathbf{S}\otimes\Gamma$. Hence $\mathbf{S},s \models [\Gamma]_i\varphi$ iff $\mathbf{S}\otimes\Gamma,(s_0,\sigma_0) \models K_i\varphi$.

The $[\Gamma]_i$-modalities are natural when transition rules prescribe prescribe agent choices, as they ensure that the performing agent *knows her choice* following the action, while allowing others to be unaware of the choice made.

Solutions and Next APM Choice. A set of transition rules dictates the choice for the next APM by finding the transition rule(s)'s *solution*. A *solution* to $\mathcal{T} = \varphi \rightsquigarrow [X]\psi$ over pointed EPM $(\mathbf{S},s)$ is a doxastic program Γ such that $\mathbf{S},s \models \varphi \rightarrow [\Gamma]\psi$. Γ is a solution to the set $\boldsymbol{T}= \{\mathcal{T}_1,...,\mathcal{T}_n\}$ with $\mathcal{T}_k = \varphi_k \rightsquigarrow [X]\psi_k$ over $(\mathbf{S},s)$ if $\mathbf{S},s \models \bigwedge_1^n(\varphi_k \rightarrow [\Gamma]\psi_k)$, i.e. if Γ is a solution to all $\mathcal{T}_i$ over $(\mathbf{S},s)$ *simultaneously*.[15] Finally, a set of doxastic programs $\mathbf{G}$ is a solution to $\boldsymbol{T}$ over $\mathbf{S}$ iff for every t of $\mathbf{S}$, there is a $\Gamma \in \mathbf{G}$ such that Γ is a solution to $\boldsymbol{T}$ over $(\mathbf{S},t)$.[16]

If $\mathbf{G}$ is a solution to $\boldsymbol{T}$ over $\mathbf{S}$, then given a state from $\mathbf{S}$, the transition rules in $\boldsymbol{T}$ will specify one (or more) programs from $\mathbf{G}$ as the next choice. A deterministic choice will be made if $\mathbf{G}$ is selected suitably, in the sense that it contains a *unique* Γ for each s. In the ensuing, solution sets will be chosen thus.

Example: Looping System. Consider the very simple 'system', consisting of an EPM $\mathbf{S}$ with $s_0 \in P_{\mathbf{S}}$, and APM $\mathbf{E}$ with $pre(\sigma_0) = P$, $post(\sigma_0) = \neg P$, and the set $\boldsymbol{T}= \{\mathcal{T}_0,\mathcal{T}_1\}$ of transition rules:

$$\begin{array}{l}\mathcal{T}_0 = P \rightsquigarrow [X]\neg P\\ \mathcal{T}_1 = \neg P \rightsquigarrow [X]P\end{array} \quad \mathbf{S}: \boxed{\begin{array}{c}s_0\\P\end{array}} \quad \mathbf{E}: \boxed{\begin{array}{cc}\sigma_0 & \sigma_1\\ \langle P;\neg P\rangle & \langle\neg P;P\rangle\end{array}}$$

With $\Gamma_0 = \{\sigma_0\}$ and $\Gamma_1 = \{\sigma_1\}$, $\mathbf{G} = \{\Gamma_0,\Gamma_1\}$ is a solution to $\boldsymbol{T}$ over $\mathbf{S}$. For $\mathcal{T}_1$, $\mathbf{S},s_0 \models \neg P \rightarrow [\Gamma_0]P$ as $s_0 \notin (\neg P)_{\mathbf{S}}$. For $\mathcal{T}_0$, it is easy to check that $\mathbf{S}\otimes\Gamma,(s_0,\sigma_0) \models \neg P$, entailing that $\mathbf{S},s_0 \models [\Gamma_0]\neg P$. As Γ_0 is unique, this is chosen as next update. It should be easy to see that $\mathbf{G}$ is also a solution to $\boldsymbol{T}$ over $\mathbf{S}\otimes\Gamma_0$, where Γ_1 is chosen. Further re-application of $\boldsymbol{T}$ loops the system.[17]

Three Agent Types. Transition rules may be used to provide general characterizations of agent behavior determined by belief. Rules with a doxastic trigger will be referred to as *decision rules*, by sets of which an abundance of possible *agent types* may be defined. Of interest are the following three, corresponding to three types of human behavior relevant to the bystander effect.

---

[15] Note the analogy with numerical equations; for both $2+x=5$ and $\{2+x=5, 4+x=7\}$, $x=3$ is the (unique) solution.

[16] The definition is altered to suit transition rules using $[X]_i$ "modalities" by suitable replacing $[X]$ with $[X]_i$ and $[\Gamma]$ with $[\Gamma]_i$ throughout.

[17] For a definition of *system*, see (Rendsvig, 2013a).

First Responder:	**City Dweller:**	**Hesitator:**
$B_iA \rightsquigarrow [X]_iI_i$	$B_iA \rightsquigarrow [X]_iE_i$	$K_iA \rightsquigarrow [X]_iI_i$
$B_i\neg A \rightsquigarrow [X]_iE_i$	$B_i\neg A \rightsquigarrow [X]_iE_i$	$B_iA \wedge \neg K_iA \rightsquigarrow [X]_iO_i$
		$B_i\neg A \rightsquigarrow [X]_iE_i$

Table 1. Decision rules specifying three agent types. Denote by $\mathcal{F}_{1i}$ and $\mathcal{F}_{2i}$ resp. the upper and lower first responder rule indexed for i, and set $\boldsymbol{F}_i := \{\mathcal{F}_{1i}, \mathcal{F}_{2i}\}$ and treat $\mathcal{C}_{1i}, \mathcal{C}_{2i}, \mathcal{H}_{1i}, \mathcal{H}_{2i}, \mathcal{H}_{3i}$, $\boldsymbol{C}_i$ and $\boldsymbol{H}_i$ in a similar manner.

The First Responder will intervene if she believes there is an accident, otherwise not. First Responders thus reflect the normally expected, but not witnessed, behavior in relation to emergencies. A City Dweller will evade the scene no matter what his beliefs, hereby reflecting the media's grim picture of the "apathetic" urban citizen, ignoring the murder of Kitty Genovese. Finally, a Hesitator will choose to observe if she believes but does not *know* that there is an accident, and will else evade. Hereby the Hesitator rules capture (part of, see below) the behavior used as explanation for the bystander effect (e.g. by Latané and Darley (1968) when they write "it is likely that an individual bystander will be considerably influenced by the decisions he perceives other bystanders to be taking" (p. 216)). Presently, focus will be on Hesitators, with comments on First Responders and City Dwellers. The latter two are subjects of Section 6.

Possible Choices. To implement either of the rule sets, a suitable set of doxastic programs for X to range over must be specified. It seems natural to assume that when an agent is intervening, then this is epistemically unambiguous for all agents. When b and c see a choose either to observe or evade, it seems more plausible that they cannot tell these actions apart, as neither action has an observable, distinguishing mark.[18] It is assumed that agents find it more plausible that others evade than that they observe.[19] In sum, these consideration give rise to the APM $\mathbf{E_1}_i$ of Fig. 6.

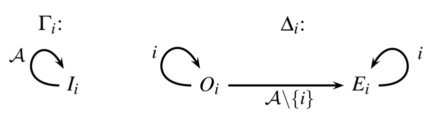

Fig. 6. The APM $\mathbf{E_1}_i$, representing the three moves available to i as well as the doxastic perception of these for the remaining agents. State names specify postconditions; all preconditions are $\top$. If i chooses to intervene, Γ_i will be next APM choice, whereas if i chooses to either observe or evade, Δ_i will be used, with actual event resp. O_i or E_i.

[18] Unless evading entails leaving the scene or observation is performed in a non-discrete manner. Often this is not the case, though: "Among American males it is considered desirable to appear poised and collected in times of stress. ... If each member of a group is, at the same time, trying to appear calm and also looking around at the other members to gauge their reactions, all members may be led (or misled) by each other to define the situation as less critical than they would if alone. Until someone [intervenes], each person only sees other nonresponding bystanders, and ... is likely to be influenced not to act himself." (Latané and Darley, 1968, p. 216); "... Apparent passitivity and lack of concern on the part of other bystanders may indicate that they feel the emergency is not serious, but it may simply mean that they have not yet had time to work out their own own interpretation or even that they are assuming a bland exterior to hide their inner uncertainty and concern." (Latané and Rodin, 1969, p. 199).

[19] The second quote in the previous note seems to indicate the plausibility of this assumption.

$\mathbf{E_1}_i$ does not facilitate simultaneous choice, in the sense that it does not contain a solution to e.g. $\{\mathcal{H}_{1a}, \mathcal{H}_{1b}\}$ over $\mathbf{S_1}$. Combining, however, a copy of $\mathbf{E_1}_i$ for each of a, b and c while respecting the doxastic links in an intuitive way may easily be done. Specifically, a combined APM $\mathbf{E_1}$ may be obtained by taking the reflexive, transitive closure of the *Cartesian graph product* $\mathbf{E_1}_a \square \mathbf{E_1}_b \square \mathbf{E_1}_c$ (see e.g. (Hammack et al., 2011) for definition) and specifying pre- and postconditions as follows: for $(s,t,u) \in \mathbf{E_1}_a \square \mathbf{E_1}_b \square \mathbf{E_1}_c$, let $pre(s,t,u)_{\mathbf{E_1}} = \top$ and $post(s,t,u) = post(s)_{\mathbf{E_1}_a} \wedge post(t)_{\mathbf{E_1}_b} \wedge post(u)_{\mathbf{E_1}_c}$. The resulting APM $\mathbf{E_1}$ has eight mutually disconnected components of four types, see Fig. 7.

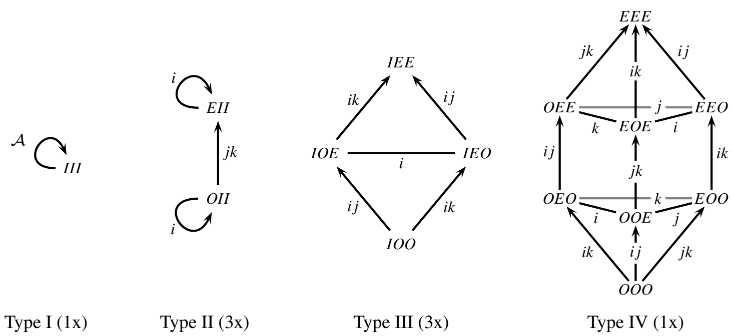

Fig. 7. The APM $\mathbf{E_1}$, representing simultaneous move programs, omitting (most) reflexive and transitive arrows; states are labeled with postconditions, with IOE representing $\langle \top; I_i O_j E_k \rangle$, etc.; all preconditions are $\top$. Notice the rise in dimensions; Type I is a point, Type IV is a cube.

In Fig. 7, the Type I model is obtained by the Cartesian product of Γ_i, Γ_j and Γ_k; Type II from Γ_j, Γ_k and Δ_i; Type III from Γ_i and Δ_j, Δ_k and Type IV from $\Delta_i, \Delta_j, \Delta_k$. Notice that reflexive and transitive closure is required to ensure that all $\leq_i$'s are pre-orders. Euclidean closure is not required, but is doxastically reasonable: in e.g. the Type III component, i should not be able to distinguish between IOE and IEO, nor consider either more plausible.

Doxastic programs over $\mathbf{E_1}$ identical to each of the four sub-model types give rise to the desired solution space $\mathbb{E}_\Gamma$. Let $\mathbb{E}$ be the set consisting of all pointed $\mathbf{E_1}$ models, and let $\mathbb{E}_\Gamma$ be the set of all doxastic programs over models in $\mathbb{E}$ such that each $\Gamma \in \mathbb{E}_\Gamma$ contains exactly one of the type I-IV sub-models. Then $\mathbb{E}_\Gamma$ contains a unique solution to every combination of the three agent types. (recall, def. of solution set: for every $s \in S$.)

Over $\mathbf{S_1}$ and $\mathbb{E}_\Gamma$, the Type I program is the unique solution to $\boldsymbol{F}_1 \cup \boldsymbol{F}_2 \cup \boldsymbol{F}_3$; the Type II program with $i = 2$ and $\sigma_0 = EII$ is the unique solution to $\boldsymbol{C}_1 \cup \boldsymbol{F}_2 \cup \boldsymbol{F}_3$ – with $\sigma_0 = OII$, it is the unique solution to $\boldsymbol{H}_1 \cup \boldsymbol{F}_2 \cup \boldsymbol{F}_3$; the Type III program with $i = 1, j = 2, k = 3$ and $\sigma_0 = IEO$ is unique solution to $\boldsymbol{F}_1 \cup \boldsymbol{C}_2 \cup \boldsymbol{H}_3$; the Type IV program with $\sigma_0 = OOO$ is unique solution to $\boldsymbol{H}_1 \cup \boldsymbol{H}_2 \cup \boldsymbol{H}_3$. Interestingly, the Type IV program with actual state OOO reflects an implicit norm-based pluralistic ignorance: though every agent is observing, each perceive the situation as one where they are the only ones doing it.

Choosing to Observe. Given the above, the system based on $\mathbf{S_1}$ with $\boldsymbol{H}_1 \cup \boldsymbol{H}_2 \cup \boldsymbol{H}_3$ as rules over $\mathbf{E_1}$ dictates the type IV program with $\sigma_0 = OOO$ as next APM choice.

That is, every agent chooses to observe. The ensuing EPM $\mathbf{S_2} := \mathbf{S_1} \otimes \Gamma_{IV}$ contains 128 states, but is easily described. Take Γ_{IV} and replace every state with a complete copy of $\mathbf{S_1}$'s frame, connect two states from two different copies $((s,\sigma_1) \leq_i (s',\sigma_2))$ iff $s = s'$ and $\sigma_1 \leq_i \sigma_2$, and finally take the reflexive-transitive closure. The new actual world is $((s_0,\sigma_0), OOO)$, satisfying

$$A,\ O_a \wedge O_b \wedge O_c \text{ and } \bigwedge_{i\in\mathcal{A}} (K_i O_i \wedge B_i \bigwedge_{i\in\mathcal{A}\setminus\{i\}} E_i).$$

The latter captures the post-factual effects of the mentioned norm-based pluralistic ignorance of Γ_{IV}: all have the belief that *they individually were the only ones to observe, while all others evaded*. Importantly, had EPM $\mathbf{E_1}_i$ on page 15 been such that agents perceived actions according to their own decision rules, i.e. found observation more plausible than evasion in others, all would have correct beliefs about the actions of others.

5 Action Interpretation and Social Proof

Albeit all agents have formed the belief that their co-witnesses evaded, none has formed any beliefs rationalizing these choices: All are still doxastically indifferent between whether the others believe A or $\neg A$. Neither does any agent have means of deducing others' beliefs, given the introduced formal framework. Such a deduction would require e.g. the ability to rationalize by *forward induction*, which requires information from both past play and future possibilities (**?**) couched in a game framework representing preferences, rationality, etc. (see e.g. (Rendsvig and Hendricks, 2013) for an implementation). Though structures akin to game trees may be defined using EPMs, APMs and protocols (van Benthem et al., 2007; Dégremont, 2010), a simpler, more superficial construct may be used to facilitate the reasoning. The suggested approach utilizes an 'inverse' version of decision rules, brute forcing conclusions about belief from observations about action.

In making decisions, our beliefs about the relevant state of affairs dictate our action, up to error and human factors. Hence the route from beliefs to actions is often functional. As this function will often not be injective, moving from actions to beliefs is not as straightforward, since multiple different belief states may result in the same action. Having to provide a rationalization of a given action will therefore often include abductive reasoning. An abductive hypothesis to rationalize an action allows inferring as explanation of the observed action a previous belief state of the acting agent. Below, such hypotheses are called *interpretation rules*.

Interpretation Rules. An *interpretation rule* is a doxastic proposition $\varphi \to [\mathbf{S}]B_i\psi$, with φ called the *basis* and ψ the *content*, with the underlying idea that on the basis of an action (e.g. E_i), agents may deduce something about the content of i's beliefs (e.g. that $B_i\neg A$).

Doxastic propositions involving the modality of the consequent are given by

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

where S is the domain of $\mathbf{S}$, and where $s \in s'$ means that s is a *predecessor*[20] of s'. Hence $[\mathbf{S}]\chi$ is true in $(\mathbf{S}', s')$ just in case s''s predecessor in $\mathbf{S}$ was a χ-world. The modality is included to respect the temporal aspect introduced by updates, and $\mathbf{S}$ is to be substituted with the EPM based on which i made the choice in question.

A set of interpretation rules may in general be implemented using an APM where the preconditions of each state is a conjunction of interpretation rules with different bases with a conjunct for each action to be interpreted. Hereby each state represents a different hypothesis regarding the acting agent's type, i.e. how the agent made decisions. The plausibility order then specifies the 'abductive hierarchy' of such hypotheses.[21]

To simplify, agents are given only *one* hypothesis about types, the hypothesis also being *correct* in the sense that the interpretation rules are (close to) the converse of the transition rules that are in fact applied. Hence the interpretation rule model APM $\mathbf{E_2}_{ij}$ that determines how agents $\mathcal{A}\backslash\{j\}$ interprets the actions of j is a one state model. Let $\rho_j \in \mathbf{E_2}_j$ and set

$$\begin{aligned} pre(\rho_j) := E_j &\to [\mathbf{S_1}]B_j\neg A & \wedge \\ I_j &\to [\mathbf{S_1}]K_jA & \wedge \\ O_j &\to [\mathbf{S_1}]B_jA \wedge \neg K_jA \end{aligned}$$

Applying such rules for all agents may be done by sequential application of $\mathbf{E_2}_j$ for each $j \in \{a,b,c\}$ on $\mathbf{S_2}$ (and the resulting models).[22] Call the APU product $\mathbf{S_3}$.

Belief-based Pluralistic Ignorance. By application of interpretation rules, agents establish beliefs about each others previous beliefs, but even though the applied interpretation rules were correct, the obtained beliefs are wrong: $\mathbf{S_1} \models B_bA$, but $\mathbf{S_3} \models B_a[\mathbf{S_1}]B_b\neg A$. This is a direct consequence of the mis-perception of actions occurring in $\mathbf{E_1}$.[23]

In the actual world of $\mathbf{S_3}$, the agents are in a state of belief-based pluralistic ignorance with respect to A: $\mathbf{S_3} \models \bigwedge_{i\in\mathcal{A}}(B_iA \wedge \bigwedge_{j\in\mathcal{A}\backslash\{i\}} B_iB_j\neg A)$, cf. the definition in (Hansen, 2011, ch. 6). To see this, notice what happens when a interprets the actions of b and c over $\mathbf{S_2}$. In $\mathbf{S_2}$, the most plausible copy of $\mathbf{S_1}$ (Fig. 4) is the one in which all states satisfy $O_a \wedge E_b \wedge E_c$ (Fig. 7). Of these 16 states, only 4 satisfy both $pre(\rho_b)$ and $pre(\rho_c)$, namely the successors of $\sigma_2, \tau_2, \sigma_7$ and τ_7, and only the first is in a's plausibility cell relative to $(s_0)_{\mathbf{S_2}}$. As all other states in this $\mathbf{S_1}$-copy are deleted upon update with $\mathbf{E_2}_b;\mathbf{E_2}_c$, it follows that a's plausibility cell $\mathcal{B}_a[(s_0)_{\mathbf{S_3}}]$ contains only the state $((((s_0,\sigma_2),OEE),\rho_b),\rho_c)$, which in turns satisfies $A \wedge B_b\neg A \wedge B_c\neg A$. Hence

[20] When constructing APU products, a state in the product model is an ordered pair (s,σ) of a state s and and action σ. In this pair, s may again be such a pair. Say that a *predecessor* of s' is any s that occurs in any of the ordered pairs of s', including s' itself.

[21] It is possible to give agents a choice of interpretation by invoking transition rules with interpretation rules as possible solutions. In the present, agents are given *no choice of interpretation*, and this construction is consequently skipped for simplicity.

[22] Each $\mathbf{E_2}_j$ functions as a *truthful public announcement* of $pre(\rho)$, for which the order of announcements does not matter (Baltag and Smets, 2009): states are deleted, the remaining orderings staying as previous. Deleting simultaneously or in some sequence makes no difference.

[23] Though time has passed, beliefs have not changed, and this is known to all: $\mathbf{S_3} \models \bigwedge_{i\in\mathcal{A}} K_i(\bigwedge_{j\in\mathcal{A}}[\mathbf{S_1}]B_j\varphi \to B_j\varphi)$ for $\varphi \in \{A,\neg A\}$.

$\mathbf{S_3} \models B_a(A \wedge B_b\neg A \wedge B_c\neg A)$. Analogous reasoning for b and c shows that $(s_0)_{\mathbf{S_3}}$ is a state of pluralistic ignorance w.r.t. A.

Again, importantly, had EPM $\mathbf{E_1}_i$ been defined so that agents considered observation more plausible than evasion, this state of pluralistic ignorance would *not* have arisen.

Social Proof. In the portrayal of the bystander effect, witnesses alter their beliefs following their mutual act of orientation, and in the light of the newly obtained information that no one else believes that there is cause for alarm, concludes that no intervention is required. To represent the revised beliefs of agents[24], introduce a new operator $SB_{i|G}$, representing the beliefs of agent i when socially influenced by her beliefs about the beliefs of agents from group G. $SB_{i|G}$ is defined using simple majority 'voting' with a self-bias tie-breaking rule: let

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

with tie-breaking parameters α, β given by

$$\alpha = \begin{cases} ½ & \text{if } s \in (B_i\varphi)_{\mathbf{S}} \\ 0 & \text{else} \end{cases} \qquad \beta = \begin{cases} ½ & \text{if } s \in (B_i\neg\varphi)_{\mathbf{S}} \\ 0 & \text{else} \end{cases}$$

This definition leaves agent i's 'social beliefs' w.r.t. φ undetermined (i.e. $\neg(SB_{i|G}\varphi \vee SB_{i|G}\neg\varphi)$) iff both i is agnostic whether φ and there is no strict majority on the matter.

Applying the notion of social belief to A in $\mathbf{S_3}$, it is easily seen that $\mathbf{S_3} \models \bigwedge_{i\in\mathcal{A}} SB_{i|\mathcal{A}}\neg A$. That is, upon incorporating social proof, all agents 'socially believe', contrary to their private beliefs, that no accident occurred.

Action under Influence. Notice that none of the three agent types introduced so far will change their action if presented again prompted to intervene, observe or evade. First Responders will again intervene, City Dwellers will again choose to evade, and Hesitators will again, irrespective of social proof, choose to observe.

To make Hesitators pay heed to the observation they chose to make, their decision rules are changed (in the ensuing section, a fusion of the two types is defined). Let an 'influenced' agent act in accordance with the following rules:

Influenced:
$SB_{i|G}A \rightsquigarrow [X]_iI_i$
$SB_{i|G}\neg A \rightsquigarrow [X]_iE_i$

Note that an Influenced agent acts like a First Responder who bases her actions on social beliefs.

[24] Strictly speaking, in the present model agents do not *revise* their beliefs. An additional operator is instead introduced to facilitate comparison with private beliefs. A belief revision policy may easily be defined using decision rules to the effect that agents update their beliefs under the suitable circumstances, see (Rendsvig, 2013b).

A Hesitator-now-turned-Influenced presented with the choice to intervene, observe or evade (as given by $\mathbb{E}_\Gamma$) will choose to evade. More precisely, if a, b and c are Influenced, the unique next APM choice will be the Type IV program with $\sigma_0 = EEE$. The actual world in the ensuing EPM $\mathbf{S_4}$, the final step of the model, will then satisfy

$$A \wedge \bigwedge_{i \in \mathcal{A}} B_i A \wedge \bigwedge_{i \in \mathcal{A}} SB_{i|G} \neg A \wedge \bigwedge_{i \in \mathcal{A}} E_i.$$

The last conjunct is an (unfortunate) end condition, as specified on page 8. Hereby, informational dynamics leading to an observable bystander effect has been modeled.

6 Comparison to Empirical Studies

The presented sequence of models, transition rules and updates conjoined captures important informational aspects of the observable bystander effect, given that the model is accepted. As presented, the sequence may be regarded as one possible execution of a broader, implicit system. Other runs of this system may be constructed by varying parameters relevant to the bystander effect.

In this section, the effect of changing two parameters will be presented. The first change is of agent types, where it is seen that for non-mixed groups, both City Dwellers and (Influenced) Hesitators will produce the observable bystander effect. The second variation is group size, and it is shown that of non-mixed populations, only (Influenced) Hesitator behavior varies as a function of group size.

Let us briefly outline the implicit system before changing parameters. The system has initial state $\mathbf{S_0}$ on page 8, where everybody knows nothing has happened and end conditions either $\bigvee_{i \in \mathcal{A}} I_i$ or $\bigwedge_{i \in \mathcal{A}} E_i$. $\mathbf{S_0}$ is updated with the occurrence of the accident, $\mathbf{E_0}$ on page 9 resulting in $\mathbf{S_1}$ on page 12, where all believe an accident has occurred, while having no information about others' beliefs. Apart from adding further agents to the population, these steps will remain fixed.[25] Next, agents make a first decision over $\mathbb{E}_\Gamma$ and $\mathbf{S_1}$ is updated with the next APM choice.[26] Depending on agent types, the run might end at $\mathbf{S_2}$. If not, the interpretation rule model on page 18 is applied for all agents, and a second decision is made based on the outcome, possibly involving the aggregation of the perceived beliefs of others'. Again, if the system does not satisfy one of the end conditions, it will continue, in which case the interpretation rule model is re-applied (suitably altered to accommodate the temporal shift), followed by decisions, etc.

In the run described in the previous sections, two *different* agent types were used. For the first choice made, agents were assumed to be Hesitators, making them choose to observe. For their second choice, they were assumed to be Influenced, making them act on their social beliefs. [27] To facilitate comparison of models, this 'mixed' type may be properly defined as *Influenced Hesitators*:

[25] Concerning $\mathbf{E_0}$, it should be obvious how the APM must be altered to include further agents, while maintaining complete higher-order ignorance.

[26] Again, it should be obvious how $\mathbb{E}_\Gamma$ may be altered to accommodate for a larger population.

[27] The shift was made to ease the exposition. Influenced agents require the notion of social beliefs, not necessary for Hesitators' first choice.

City Dweller:	**Hesitator:**	**Influenced Hesitator:**
$B_iA \rightsquigarrow [X]_iE_i$	$K_iA \rightsquigarrow [X]_iI_i$	$K_iA \rightsquigarrow [X]_iI_i$
$B_i\neg A \rightsquigarrow [X]_iE_i$	$B_i\neg A \rightsquigarrow [X]_iE_i$	$B_i\neg A \rightsquigarrow [X]_iE_i$
	$B_iA \wedge \neg K_iA \rightsquigarrow [X]_iO_i$	$\neg K_iA \wedge \neg(SB_{i\|G}A \vee SB_{i\|G}\neg A) \rightsquigarrow [X]_iO_i$
First Responder:		$O_i \wedge SB_{i\|G}A \rightsquigarrow [X]_iI_i$
$B_iA \rightsquigarrow [X]_iI_i$		$O_i \wedge SB_{i\|G}\neg A \rightsquigarrow [X]_iE_i$
$B_i\neg A \rightsquigarrow [X]_iE_i$		

Table 2. Re-specification of agent types.

Notice that Influenced Hesitators behave as a mixture of Hesitators (first three rules) and First Responders (last two), but who take social proof into account. Notice the difference between third rule for Hesitators and the same for Influenced Hesitators. The latter requires that Influenced Hesitators have undetermined social beliefs before they choose to observe. The altered First Responder rules (rows four and five) capture that if the agents has observed and have determined social beliefs, observation gives way to intervention or evasion.[28]

Varying Types and Group Size. The table below summarizes the end conditions and the EPM number where they arise relative to agent types and group size. For end conditions, I_k and E_k represent $\bigvee_{i\in\mathcal{A}} I_i$ and $\bigwedge_{i\in\mathcal{A}} E_i$ respectively being satisfied in EPM $\mathbf{S_k}$, with k rising as described in the previous sections. '–' means that no end conditions are met.

	1	2	3	$k \geq 4$
FR	I_2	I_2	I_2	I_2
CD	E_2	E_2	E_2	E_2
H	–	–	–	–
IH	I_3	I_3	E_3	E_3

Table 2. Summary of end conditions and times as a function of agent type and group size.

In words, for any group size, First Responders will intervene in $\mathbf{S_2}$, i.e. immediately following the accident. At the same time, City Dwellers will evade the scene, no matter the group size. 'Simple' Hesitators, as defined in Table 2, will never reach an end conclusion, as they will never come to either believe there is no accident, or know that there is one. That all these three agent types' actions are invariant over group size is due to their inherently *non-social* decision rules. The 'social', or Influenced, Hesitators will however change their behavior according to group size: they will intervene immediately if the group size is small enough for their private belief not to be 'overridden' by social proof. If the group size is 3 or above, Influenced Hesitators will conclude, by the mis-perception of others' choice to observe as an act of evasion and the resulting state

[28] The requirement that i must have observed before acting on social beliefs ensures that agents do not intervene immediately after seeing the accident (a private belief that A would imply that $SB_{i|G}A$, as agents then hold no beliefs regarding others' beliefs).

of pluralistic ignorance, that enough agents believe that no accident occurred for themselves to be 'socially convinced' that this is the case. Consequently, they will choose to evade in $\mathbf{S_3}$ for any group size of 3 or above.

At group size 2 these agents still decide to intervene because they use their own belief to break the tie between what they perceive as an even split on whether an accident is happening.[29]

Comparison to Empirical Studies. Running the system with each of the four agent types may be considered as providing four different models of the bystander effect, each of which may be compared to empirical results to evaluate consistency with data. Table 2 allows for only a simple comparison, checking whether end conditions as a function of group size correlates properly with the observed.

A wide variety of studies have been performed on the inhibiting effect of the presence of others in situation requiring intervention (see e.g. (Latané and Nida, 1981) for a meta-study). Many of these have different, more specific foci, e.g. the role of diffusion of responsibility, friendship, gender, and more. As the focus of this paper is the second step of the bystander effect, only studies on the effect of social proof on the perception of the accident are relevant. Alas, no one such has been found that provides suitable data for all group sizes. Comparison can therefore only be made using multiple studies invoking different experimental settings.

Inherently, the presented models are *deterministic*, while experimental data provides information about the *percentage* of individuals who intervene. Given this, data will *not* be correctly matched. To evaluate tendency of correctness, acceptance or rejection of models are therefore based not on strict consistency, but on the weak requirement that a model must correctly match the binary experimental end conditions in strictly more than 50% of cases.

Smoky Room and the Rejection of FR, CD and H. The classic 'smoky room' experiment of (Latané and Darley, 1968), specifically designed to test the hypothesis of the second step of the explanation of the bystander effect (see p. 2), has served as a strong guide for the construction of the models, and provides data which allows the rejection of three of them. In the study, groups of size 1 or 3 where sat in a waiting room, completing questionnaires. The groups of size 3 either consisted of 1 individual naive to the experiment and 2 of the experimenters' confederates, or 3 naive subjects. While completing the questionnaire, smoke was introduced to the room through a visible vent, ambiguously indicating either an emergency (e.g. fire) or not (e.g. steam). As the possible accident will have dire consequences for the subjects themselves, the experimenters assumed that no diffusion of responsibility arose.

The experiment was stopped when either one agent intervened, or after six minutes of smoke introduction and questionnaire completion, at which point smoke was heavy. Compared to the model, these end conditions are identified with $\bigvee_{i\in\mathcal{A}} I_i$ and $\bigwedge_{i\in\mathcal{A}} E_i$, respectively.

Of the subjects that were alone, 75% reported the smoke, a number high enough to warrant the rejection of the City Dweller model, which would have it that all evade.

[29] Cf. the tie-breaking rule used in the definition of social beliefs.

Likewise, the Hesitator model is rejected, as it would have it that individuals would continuously observe (instead of completing their questionnaires). Both the First Responder and Influenced Hesitator models score better than 50%.

With 2 confederates in the room, only 1 in 10 naive subjects intervened. With 75% of individuals intervening when alone, it should be expected that 98% of groups of size 3 with three naive subjects would intervene if individuals acted independently,[30] but only 38% of the 8 groups did so.

The First Responder model is 10% correct in the confederates condition[31] and 38% in the naive group condition, hereby falling below the 50% mark. The Influence Hesitators model does better: it is 90% correct of the confederates condition[32] and 62% correct in the naive group condition. Hence, it fairs better than 50% overall.

Interestingly, the response time for intervention in the three naive subjects case was considerably longer than the single subject case, indicating 1) that individuals in groups did pay heed to social proof before acting, and 2) that in many cases (38%), social proof from only two peers was not enough to preclude intervention.

Is One Additional Witness Enough for Intervention Inhibition? The smoky room study only compares groups of size 1 and 3, whereby it does not supply sufficient data to evaluate the IH model for group sizes of 2 or above 3. To evaluate the model for group size of 2, another classic experiment may be consulted, namely the 'lady in distress' case of (Latané and Rodin, 1969). Three conditions where tested in this experiment: with a lone, naive subject, with one naive subject and one confederate, and with two naive subjects, again filling out questionnaires. In a simulated accident, the female interviewer faked a fall in an easily accessible and audible adjacent room. The fall was indicated by a loud crash, a scream and subsequent moaning of complaints and hurt. Contrary to the smoky room, this accident is ambiguous between either a serious accident (e.g. broken leg) or a not-so-serious one (weakly sprained ankle).

In the first condition, 70% of the alone subjects intervened, with a strong drop to 7% when an inactive confederate was introduced. With two naive subjects, 91% of groups would be expected to intervene if subjects acted independently, whereas only 40% of such groups in fact did so.

These percentages strongly contradict the IH model for group size 2, as the expectation is that neither of the two agents would be sufficiently influenced by each other to not act on their private beliefs. Both would therefore intervene following observation. This makes the model incorrect in 60% of cases, making it worse than a random bet.

Partly, the model may misfire as the experiment does not conform to the pluralistic ignorance explanation of the bystander effect. Specifically, the experiment does not preclude the possibility of a mix of social proof and diffusion of responsibility effects, given that the accident in question did not put the subjects in faked danger. An experiment precluding diffusion effects may be conjectured to show a higher degree of intervention, yielding a better fit.

---

[30] See (Latané and Darley, 1968) for calculation of hypothetical baseline based on the alone condition.

[31] In a mixed population model, using City Dweller agents for the two confederates.

[32] Again in a mixed population model, using City Dweller agents for the two confederates.

The meta-study (Latané and Nida, 1981) strongly indicates that determining social influence occurs in groups of size 2. Summarizing 33 studies with face-to-face interaction, the effective individual probability of helping was 50%, with an effective individual response rate in groups only 22%. Most of these studies involved groups of 2.[33] Hence, it seems that the model misfires when it comes to groups of size 2.

The obvious parameter to tweak for a better fit is the self-biased majority voting definition of social beliefs, which does not put enough weight on the other in the 2 person case. Changing this to one favoring the perceived beliefs of the other would yield a better model for the 2 subject case, while it would not alter the results for the group size 3 case. Table 3 summarizes the effect of the Influenced Hesitator model run using a tie-breaking rule favoring the *opposite* belief of ones own.[34]

	1	2	3	$k \geq 4$
IH′	I_3	E_3	E_3	E_3

Table 3. Summary of end conditions as a function of agent type and group size for others-biased social beliefs.

Though this model does not fair very well on the data from the smoky room and lady in distress studies, it does at least do better than a random bet.

7 Conclusions

Accepting the individual modeling steps as reasonable explications of the reasoning steps occurring in bystander effect-like scenarios, the constructed dynamics allows for a number of conclusions about such phenomena:

- The sub-optimal choice for all to evade is *not* a consequence of "apathetic" agents: City Dweller evasion is not influenced by group size.
- The sub-optimal choice for all to evade is a direct result of considering social proof in a state of proposition-based pluralistic ignorance: Influenced Hesitators with *correct* beliefs about their peers beliefs would choose to intervene.
- Proposition-based pluralistic ignorance arises due to norm-based pluralistic ignorance: that all agents assume others are evading when they themselves are observing is a necessary condition for the state of proposition-based pluralistic ignorance to arise.
- Subjects *do not* incorporate social proof by self-biased majority voting, but rather the opposite.

[33] How well these individual studies conform to the pluralistic ignorance explanation of the bystander effect has not been checked.

[34] I.e., interchanging the α, β tie-breaking parameters. Alternatively, social beliefs could be defined by weighing others' perceived beliefs higher than one's own, or by moving to a threshold rule requiring e.g. perceived agreement with all peers as done in (Seligman et al., 2013; Christoff and Hansen, 2013)).

Several venues further for both formal and empirical research present themselves. As no pure agent type group fits data very well, two possibilities are worth investigating. First, how well will models with mixed groups perform? That not all subjects chose to intervene in the single agent case seems to indicate that at least some behave as City Dwellers; that some chose to intervene in the three agent case indicates that some act as First Responders. With suitable proportions of each agent type, a model may be produced which will match data more closely with an average of end conditions of runs based on random picks from the mixed population. To fit both population mix and the social influence parameter, a data set from a large-scale smoky room-style study would be required.

Finding implementable resolution strategies for the pluralistic ignorance state could be of benefit, if these turn out to work in practice. Some such have been suggested in the literature; in the study of Schroeder and Prentice (1998), information on the subject diminished the alcohol consumption among college students. How information should provoke changes in agent type in the present framework is an open question. A shorter term strategy for obtaining help is suggested by Cialdini (2007): Single out an individual and ask *only her* for help. If an agent is singled out, the inaction of others will no longer be eligible as a source of information about the event. Hence the agent is forced to act on her private beliefs, in which case both Influenced Hesitator models predicts intervention. Of formal studies, Proietti and Olsson (2013) show how a state of pluralistic ignorance state may be dissolved by a series of announcements of private beliefs heard by matrix neighbors. Specifying an agent type replicating the behavior and varying only the network structure of the model might provide further insights into the fragility of the phenomenon.

For a complete model of the bystander effect, both the first (noticing the event) and third step (assuming responsibility) of the explanation provided in the introduction must be modeled. The former may rest less on information processing than features of physical space: as more people are present, less may notice the event e.g. due to obscured line of sight. Modeling the third step may require a more expressive logical framework in which beliefs regarding agent types may be held: if all agents falsely believe a First Responder is present, all may believe that intervention is required while no-one will take action.

Acknowledgements. The author would like to thank the editors for organizing the CPH-LU workshops on social epistemology, as well as the participants of said workshops for valuable comments and discussion. The work has benefited especially from discussions with Henrik Boensvang and Vincent F. Hendricks. Carlo Proietti is thanked for his corrections to the manuscript. Finally, a warm Thank You to the two anonymous reviewers: the comments, criticisms and correction provided by your thorough reading of the original submission have been invaluable.

Bibliography

Baltag, A. and Smets, S. (2008). A Qualitative Theory of Dynamic Interactive Belief Revision, *in* G. Bonanno, W. van der Hoek and M. Wooldridge (eds), *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, number Loft 7 in *Texts in Logic and Games, Vol. 3*, Amsterdam University Press, pp. 9–58.

Baltag, A. and Smets, S. (2009). Talking your way into agreement: Belief merge by persuasive communication, *in* M. Baldoni, C. Baroglio, J. Bentahar, G. Boella, M. Cossentino, M. Dastani, B. Dunin-Keplicz, G. Fortino, M.-P. Gleizes, J. a. Leite, V. Mascardi, J. Padget, J. Pavón, A. Polleres, A. E. F. Seghrouchni, P. Torroni and R. Verbrugge (eds), *Proceedings of the Second Multi- Agent Logics, Languages, and Organisations Federated Workshops*, CEUR Workshop Proceedings, Vol. 494, Turin, Italy, pp. 129–141.

Bicchieri, C. and Fukui, Y. (1999). The great illusion: ignorance, informational cascades, and the persistence of unpopular norms, *Business Ethics Quarterly* **9**(1): 127–155.

Blackburn, P., de Rijke, M. and Venema, Y. (2001). *Modal Logic*, Cambridge University Press.

Bolander, T. and Birkegaard, M. (2011). Epistemic planning for single- and multi-agent systems, *Journal of Applied Non-Classical Logics* **21**(1): 9–34.

Christoff, Z. and Hansen, J. U. (2013). A Two-Tiered Formalization of Social Influence, *in* D. Grossi and O. Roy (eds), *Logic, Rationality, and Interaction*, Lecture Notes in Computer Science, Springer, pp. 68–81.

Cialdini, R. (2007). *Influence: The Psychology of Persuasion*, HarperCollins Publishers.

Dégremont, C. (2010). *The Temporal Mind: Observations on the logic of belief change in interactive systems*, PhD thesis, University of Amsterdam.

Demey, L. (2011). Some remarks on the model theory of epistemic plausibility models, *Journal of Applied Non-Classical Logics* **21**(3-4): 375–395.

Fagin, R., Halpern, J. Y., Moses, Y. and Vardi, M. Y. (1995). *Reasoning About Knowledge*, The MIT Press.

Halbesleben, J. R. and Buckley, M. R. (2004). Pluralistic ignorance: historical development and organizational applications, *Management Decision* **42**(1): 126–138.

Hammack, R., Imrich, W. and Klavzar, S. (2011). *Handbook of Product Graphs*, Discrete Mathematics and its Applications, 2 edn, CRC Press.

Hansen, J. U. (2011). *A logic toolbox for modeling knowledge and information in multi-agent systems and social epistemology by*, PhD thesis, Roskilde University.

Hansen, P. G., Hendricks, V. F. and Rendsvig, R. K. (2013). Infostorms, *Metaphilosophy* **44**(3): 301–326.

Hendricks, V. F. and Rasmussen, J. L. (2012). *Nedtur! Finanskrisen forstået filosofisk*, Gyldendal Business, Copenhagen.

Latané, B. and Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies., *Journal of Personality and Social Psychology* **10**(3): 215–21.

Latané, B. and Darley, J. M. (1970). *The Unresponsive Bystander: Why Doesn't He Help?*, Appleten-Century-Crofts, New York.

Latané, B. and Nida, S. (1981). Ten Years of Research on Group Size and Helping, *Psychological Bulletin* **89**(2): 308–324.

Latané, B. and Rodin, J. (1969). A Lady in Distress: Inhibiting Effects of Friends and Strangers on Bystander Intervention, *Journal of Experimental Social Psychology* **5**: 189–202.

Miller, D. and McFarland, C. (1987). Pluralistic ignorance: when similarity is interpreted as dissimilarity, *Journal of Personality and Social Psychology* **53**: 298–305.

Myers, D. G. (2012). *Social Psychology*, 11 edn, McGraw-Hill.

Prentice, D. and Miller, D. (1993). Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm, *Journal of Personality and Social Psychology* **64**: 243–56.

Proietti, C. and Olsson, E. J. (2013). A DDL Approach to Pluralistic Ignorance and Collective Belief, *Journal of Philosophical Logic* **March**: 1–17.

Rendsvig, R. K. (2013a). Aggregated Beliefs and Informational Cascades, *in* D. Grossi and O. Roy (eds), *Logic, Rationality, and Interaction*, Lecture Notes in Computer Science, Springer, pp. 337–341.

Rendsvig, R. K. (2013b). Look and Learn: Extracting Information from Actions of Others, *http://staff.science.uva.nl/˜ulle/esslli-workshop-2013/papers/rendsvig.pdf* .

Rendsvig, R. K. and Hendricks, V. F. (2013). Social Proof in Extensive Games, *under review* .

Schroeder, C. M. and Prentice, D. A. (1998). Exposing Pluralistic Ignorance to Reduce Alcohol Use Among College Students, *Journal of Applied Social Psychology* **28**(23): 2150–2180.

Seligman, J., Girard, P. and Liu, F. (2013). Logical dynamics of belief change in the community, *Under submission* .

van Benthem, J. (2007). Dynamic logic for belief revision, *Journal of Applied Non-Classical Logics* **17**(2): 129–155.

van Benthem, J., Gerbrandy, J. and Pacuit, E. (2007). Merging Frameworks for Interaction : DEL and ETL, *in* D. Samet (ed.), *Proceedings of TARK 2007*, pp. 72–81.

van Ditmarsch, H. and Kooi, B. (2008). Semantic Results for Ontic and Epistemic Change, *in* G. Bonanno, W. van der Hoek and M. Wooldridge (eds), *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, number Loft 7 in *Texts in Logic and Games, Vol. 3*, Amsterdam University Press, pp. 87–117.

Westphal, J. D. and Bednar, M. K. (2005). Pluralisitc Ignorance in Corporate Boards and Firms' Strategic Persistence in Response to Low Firm Performance, *Administrative Science Quarterly* **50**: 262–298.