The Mere Reminder of Others: A Cognitive Modelling Approach to the Implicit Bystander Effect

Chris M. Stolle (chstolle@edu.aau.at)
Department of Psychology, Lingnan University, 8 Castle Peak Road,
Hong Kong, China

Yi Huang (yihuang@ln.edu.hk)
Department of Psychology, Lingnan University, 8 Castle Peak Road,
Hong Kong, China

Abstract

The bystander effect suggests that people are less likely to assist in an emergency when others are present. Traditional theoretical accounts attribute this effect to top-down reflective processes, such as the diffusion of responsibility. However, recent research has proposed a two-system approach, suggesting that an individual's initial response to an emergency is personal distress and avoidance, which is further amplified by the presence of bystanders. In this study, we embed the two-system approach into an evidence accumulation model and argue that a higher distress and avoidance response causes slower evidence accumulation. We conducted a reaction time experiment where participants got exposed to faces or control stimuli and had to judge if a scene was dangerous. Our results confirm our hypothesis: Individuals exposed to faces had slower evidence accumulation for dangerous decisions. These findings contribute to a mechanistic understanding of how the anticipated bystander presence can influence early reflexive reactions to emergencies.

Keywords: bystander effect; drift diffusion model; cognitive modelling; perception

Introduction

The bystander effect is a phenomenon where individuals are less likely to intervene or help in an emergency when others are present. This was famously exemplified in the tragic murder of Kitty Genovese, where multiple witnesses observed the attack but failed to intervene. The aftermath of this case led to psychological research on the bystander effect.

Darley and Latané (1968) conducted the first controlled laboratory experiment to test this real-world observation. They instructed participants to discuss college-life problems with other students who were actually all confederates who played the role of bystanders. Participants were placed alone in a room and communicated with the others over a microphone. During the discussion, a confederate simulated an epileptic seizure. When participants were the only witness to the emergency, they always helped. However, when participants were grouped with more confederates, they showed slower response times and a diminished likelihood to help.

Later research has expanded the scope of the bystander effect to include general danger detection in the presence of others. First, the effect is not limited to emergencies involving human victims. An example is a study by Latané and Darley, where participants waited in a waiting room which suddenly filled up with smoke. Remarkably, when other individuals were present in the same room, participants often neglected to report or even recognise the emergence of the smoke (Latané & Darley, 1968). Second, bystanders do not necessarily need to be physically present to influence the situation, as demonstrated in Darley and Latané's (1968) classic study where the contact with bystanders was only over a microphone connection. The bystander effect has also been observed in virtual environments such as chat rooms (Markey, 2000), mail-lists (Barron & Yechiam, 2002), when observing cyberbullying on the Internet (Machackova, Dedkova, & Mezulanikova, 2015; for review see Rudnicki, Vandebosch, Voué, & Poels, 2023) and even when the social context was only implied or imagined (Garcia, Weaver, Moskowitz, & Darley, 2002). These findings highlight the far-reaching influence of the bystander effect across different contexts and situations.

Traditional theories attribute the bystander effect to situational top-down processes based on norms or social influences (for corresponding theories, see Latané & Darley, 1970; Latané & Nida, 1981). One widely accepted explanation is the diffusion of responsibility, which suggests that when there are multiple observers, individuals are less likely to intervene or provide assistance because they perceive the responsibility to act is shared among the group. The individual feels less accountable, resulting in inaction (Latané & Darley, 1970). Such explanations provide good descriptive understanding but cannot explain underlying psychological mechanisms (Fischer et al., 2011). More recent findings can also not be explained by traditional accounts. Evidence from bystander effects in rats suggests that the effect is not solely driven by human social norms (Havlik et al., 2020). Moreover, the bystander effect can be elicited by priming a social context (see Garcia et al., 2002), indicating that the effect can occur with implicated or imagined social influence.

In summary, while the traditional theories of top-down processes can explain aspects of the bystander effect, they cannot account for all bystander findings. The proposed top-down processes involving social norms and influences can hardly explain animal bystander behaviour. For studies where bystander presence is only imagined, classic top-down explanations reach their boundaries. Therefore, to account for these effects, bottom-up processes must be considered. In the following, we will discuss bottom-up perspectives and expound on a corresponding theory.

Many neuroscience studies indicate a need to consider bottom-up explanations. In Hortensius, Schutter, and de Gelder's (2016) study, participants watched a video and had

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

to decide whether or not it showed a dangerous situation. Participants with high personal distress exhibited lower motor corticospinal excitability (indicating less action preparation) with more bystanders. This effect was even more pronounced under cognitive load, suggesting that the bystander effect may reflexively impact action preparation. Further evidence comes from other human neuroimaging studies focussing on subconscious brain processes. For example, a study found that the presence of bystanders during an emergency was linked to decreased activity in areas associated with action preparation (Hortensius & de Gelder, 2014). Another study found increased activation in premotor brain areas when participants observed confrontations between others, regardless of their attention to the situation (Sinke, Sorger, Goebel, & de Gelder, 2010). This suggests that unconscious processes are activated when observing an emergency.

A recent perspective suggests that the bystander effect is influenced by reflexive bottom-up processes that occur before reflective, norm-based or social-influence-driven processes come into play (Hortensius & de Gelder, 2018). This perspective posits that when facing an emergency, a person's immediate distress triggers a fast fight-freeze-flight response (system I, see Figure 1A), which inhibits helping behaviours. This inhibitory effect is amplified in the presence of bystanders. Conversely, a slower and reflective system for sympathy (system II) promotes helping and counteracts the initial inhibition from system I. The net outcome of both systems determines the final behavioural outcome. The exact calculation of this net outcome was not clearly defined in the perspective paper of Hortensius and de Gelder (2018). If we strictly follow the proposed opponent process model approach (see Graziano & Habashi, 2010; Solomon & Corbit, 1974), as mentioned by Hortensius and de Gelder, the activity of both processes would be summed at any time point. However, given that the slow reflective system II lasts longer than the fast reflexive system I, this process would eventually always lead to helping behaviour. In this scenario, the assistance behaviours would be postponed, not hindered, contradicting the previous bystander literature. Therefore, we propose that the process must be able to terminate without resulting in helping. To fulfil this requirement, the net outcome function is not simply the sum of the two systems' current states of activation (i.e. their probability density function; PDF). Instead, both systems contribute to one cumulative density function (CDF) over time. The CDF has the advantage that it takes into account the activation of the systems at previous time points. The lower part of Figure 1B illustrates a CDF for the net outcome of both systems in Figure 1A. The CDF can be interpreted as an evidence accumulation process that determines the behavioural outcome.

Inspired by the reflexive bystander theory (Hortensius & de Gelder, 2018), we propose a drift-diffusion model (DDM; Ratcliff, 1978) approach to study implicit bystander effects. The DDM, an evidence accumulation model, is often used for modelling reaction time data for dichotomous decisions (for non-technical primers, see Alexandrowicz, 2020; Johnson, Hopwood, Cesario, & Pleskac, 2017). The DDM assumes that people accumulate information over time until a decision criterion is met. In the DDM, the two possible decisions are modelled as an upper and lower boundary, with the evidence accumulation process starting between them (commonly exactly in the middle). A decision is made when the evidence accumulation reaches one of the boundaries. This evidence accumulation process is exemplified in the upper half of Figure 1B. In the following, we show how the DDM applies to the bystander effect. Essentially, bystanding is a slowed or inhibited danger detection. Applied to the dichotomous decision framework of the DDM, bystanding can be investigated as deciding whether a situation is dangerous or safe. For modelling the decision-making process, the DDM has four main parameters that we briefly introduce. 1) Bias β: Before any evidence accumulation starts, people may be inclined towards one of the two alternatives. For danger detection, this means that people have specific presumptions that a situation usually is dangerous (high β) or safe (low β). In Figure 1B, the bias β marks the y-axis starting point of the evidence accumulation. Typically, β is represented by a value between 0 and 1, where .5 represents a starting value that lies exactly in the middle of both alternatives (i.e. neutrality). 2) Threshold α: This represents the amount of information that needs to be accumulated to reach a decision. A higher threshold means more information is required. Applied to danger detection, a person with a higher threshold either needs more time or a more vivid perception before drawing any conclusion about the danger of the situation. See the upper dotted line in Figure 1B. 3) Drift rate v: The average speed of evidence accumulation. It can be positive or negative, as the average evidence accumulation can go towards the upper or the lower alternative. A large absolute drift $|v|$ means a fast evidence accumulation. In danger detection, a high drift v would be interpreted as a fast evidence accumulation towards either the decision 'dangerous' or 'safe'. We hypothesise the drift rate v to be central to the bystander effect, as it directly corresponds to the current net outcome of system I+II, that we previously introduced as the CDF. 4) Non-decision time τ: Any reaction time (RT) components not part of the decision-making process, such as motor response time or visual perception time. In Figure 1B, τ is represented as a shift on the x-axis.

Figure 1B illustrates the application of the reflexive bystander theory to a DDM process. We propose that, except for the drift rate v, the parameters remain constant across different bystander conditions for the following reasons. In their foundational paper, Hortensius and de Gelder (2018) suggested that in a dual-process model, the bystander effect strengthens the distress system I. Although their model suggests that the bystander effect does not affect system II, the one-sided change in system I leads to a reduced net outcome for both systems (i.e. a lower likelihood to help) at any given time. Hence, when a decision-maker without bystanders reaches a particular likelihood to help at time point

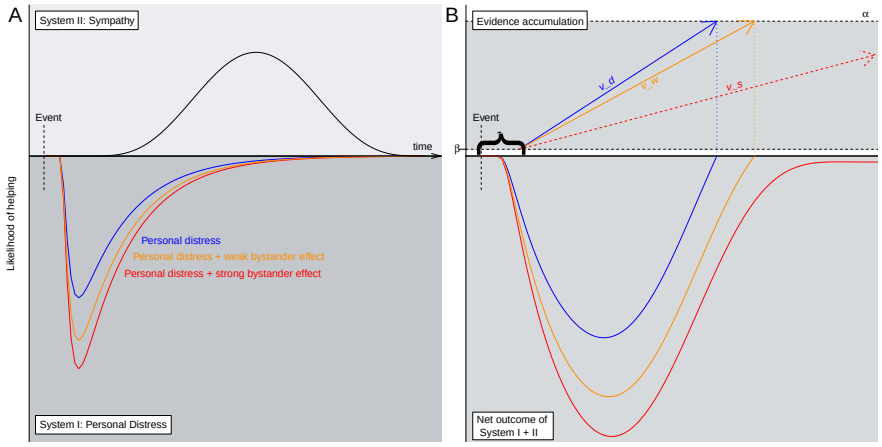

Figure 1: A two-system approach for explaining the bystander effect. Once observing an emergency, a fast reflexive fight-freeze-flight system I becomes active, which inhibits helping. The system I effect increases the more distress one experiences. The bystander effect takes place by increasing the distress effect on system I. After some time, the slower and reflective System II for sympathy starts to promote helping. Whether one helps or not depends on the net result of both systems (Hortensius & de Gelder, 2018). (A) shows the density functions of both processes. System I starts early and is more intense when there is a bystander effect. System II starts later but lasts longer. The lower part of (B) shows the net outcome cumulative functions when adding the likelihood functions of A together. Once the functions reach the upper boundary, system II has successfully outweighed the effects of system I and the person helps. The upper part of (B) maps the cumulative net outcome functions part onto the evidence accumulation process of the Drift Diffusion Model. First, there is the non-decision time τ. At the beginning of the decision-making process itself, all evidence accumulators start at the same point (with the same bias β) and accumulate information until they reach a threshold α. Since system I has affected the net outcome functions differently, all the evidence accumulators have different average drift rates until they reach the threshold (equal for all accumulators). The drift without bystanders (v_d) has the highest value. The drift rate is decreased with a stronger bystander effect (i.e. $v_d > v_w > v_s$).

μ, the same decision-maker exposed to bystanders will reach the same likelihood at time point $\mu + x$ with $x > 0$. This effect is due to changes within the decision-making process, which implies a change in the evidence accumulation rate (drift v). When two accumulation processes start at the same time point and have the same bias, yet one accumulation process is slower than the other, this implies a lower drift. The social influence literature backs this perspective. Germar, Schlemmer, Krug, Voss, and Mojzisch conducted an experiment where participants had to determine the majority colour of a bunch of dots on a screen. When shown the answers of others, participants increased the drift rate towards the response favoured by the others, even when the colour discrimination task was unambiguous (Germar et al., 2014). Further evidence supporting this comes from the first-person shooter task, a different domain of danger detection, where people decide if an image of a person includes a gun. Depending on the individual's skin colour in the image, the evidence accumulation towards the decision boundaries is affected: a lower drift towards the decision 'gun' is observed when the person is white compared to black (Pleskac, Cesario, & Johnson, 2018). This suggests that even visual stimuli without direct and conscious relevance to the target task may affect the accumulation drift.

Regarding the remaining three parameters, we do not anticipate any direct effects. 1) For the bias β, bystander effects are unlikely to affect an individual's preference towards one of the two alternatives before the decision-making process. 2) Although changes in the threshold α are theoretically possible, they are not considered by the model proposed by Hortensius and de Gelder (2018). Theoretically, people could increase their threshold to avoid premature, inappropriate behaviour, but this contradicts our theory: When they adapt their criterion in response to bystanders, this would imply that there is not a dual process (of system I and system II) involved, but that only one system would adapt its criterion. 3) Changes in the non-decision parameter τ are theoretically possible and compatible with the model. The stronger activation of system I could lead to slower motor responses. However, this effect would not be due to shifts in system I of the 2-system approach but rather a pure freezing effect in others' presence. Therefore, we argue that this may not constitute the main bystander effect.

Building on the work of Hortensius and de Gelder (2018), we propose to adopt a reflexive perspective and hypothesise that the mere presence of others triggers the bystander effect. Previous studies on bystander behaviours were either conducted in a complex social context, making it challenging to ascribe behaviour to specific processes (e.g. Darley & Latané, 1968; Latané & Darley, 1968), or they implicitly modelled a social context in which others' actions were inconsistent with the emergency (e.g. Hortensius & de Gelder,

2014; Hortensius et al., 2016). Here, we introduce a new paradigm in which participants must quickly classify scenes as dangerous or safe. The bystander effect is primed using displayed facial images. The facial expressions are manipulated to align or conflict with the emergency, altering the perceived context. We hypothesise that the bystander effect would manifest independently of the emergency congruence of the facial expression.

We hypothesise that the bystander effect can be attributed to increased activation of system I (see Hortensius & de Gelder, 2018), and this effect could be explained by a slower evidence accumulation for the decision to help. Our primary hypotheses are as follows: (1) slower RT when the bystander effect is primed, regardless of whether the facial expression is congruent with the emergency. (2) In the DDM, we expect a decreased drift v for the decision to help.

Methods

Participants

We recruited ninety-three participants from the Prolific online platform (prolific.co). We excluded six participants who refused to answer and two for low accuracy ($< 60\%$). The final sample consisted of eighty-five participants (37 female, 2 unknown) with an average age of 24.53 years ($SD = 4.81$). The Ethics and Discipline Committee of Lingnan University approved the study. All participants provided informed consent and were paid 4.5£ to compensate them for their time.

Stimuli

For our study, we designed an image classification task that involved scenes depicting either a safe and non-threatening environment or a dangerous situation involving a fire. The scenes were sourced from the Comprehensive Disaster Dataset (Niloy et al., 2021). This is a rich dataset that contains labelled dangerous (such as fire) and non-dangerous naturalistic situations. Sourcing both situation categories (dangers/non-dangerous) from the same dataset alllowed us to reduce systematic differences between the images occurring from non-category influences. Further, we deliberately chose naturalistic scenes to closer resemble real-world danger detection, as it is for relevance in typical bystander situations. We randomly sampled (without replacement) 48 images from each of four different categories: wildness-safe, wildness-fire, urban-safe and urban-fire. The images were then cut to 150x150 pixels and transformed to a grey scale. We visually inspected each image to ensure that the images were suitable for the study. A single researcher filtered out any images on which humans were present or where the content was not clearly categorisable as being safe or dangerous. After filtering, replacement images were sampled from the same categories. To expand the image set further, we mirrored all the images horizontally. In total, the final scene data set contained 384 images.

The face images were sourced from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). The database contains face images of individuals showing a specific expression (e.g. smiling) and wearing a grey T-shirt against a white background. Further, each face was rated by a large sample on different criteria (e.g. attrativeness). We sampled face images of six different individuals, each showing a smile (emergency-incongruent) and a neutral (emergency-congruent) facial expression, resulting in twelve images. Our sampling criteria included attractiveness (within Q1 and Q3), age (within Q1 and Q3), trustworthiness (>Q2), race (perceived as white by 100% of raters) and gender (50% female).

To create the face-coloured triangles, we scrambled the final neutral face images. We defined an isosceles triangle with a horizontal base whose length corresponded roughly to the shoulder width shown in face images. The height corresponded to the visible height of the face stimuli (including the neck and shoulder). Any pixels outside the triangle were set transparent, while pixels within the triangle were ordered randomly, resulting in scrambled triangles.

Experimental task and procedure

In this task, participants were required to determine whether the scene was dangerous or safe. Before the onset of each scene, we primed participants with either a display of faces (in the experimental condition), face-coloured triangles (in the control condition), or no priming stimuli to create a sense of bystander presence. The experiment was programmed using the lab.js framework (Henninger, Shevchenko, Mertens, Kieslich, & Hilbig, 2022) and hosted on our website. Participants were required to complete the experiment in a closed room without any disturbances. Instructions were provided on the screen, and participants had to successfully complete a quiz to demonstrate their understanding before beginning the experiment. The experiment consisted of 384 trials that were divided into two experimental blocks. Each block began with six training trials, followed by 186 experimental trials. Participants were instructed to utilise the physical keyboard of their computer/laptop to make responses using the F and J keys.

Figure 2 shows an example trial. First, priming stimuli (happy faces, neutral faces, or triangles) were presented for a random time between 2 and 4 seconds. Then, the scene was shown. Participants had 1.5 seconds to determine whether the scene was dangerous (i.e. contained fire) or safe (no fire). If participants responded too slowly, a blue screen was shown for 3 seconds, indicating they should respond more quickly. If they responded within 1.5 seconds, but the response was wrong, a red X appeared below the scene for 1 second. Lastly, an inter-trial interval with a fixation cross was shown for 500ms.

The experiment followed a 2 x 3 x 3 mixed design, with within-subject factors of scenes (dangerous vs. safe) and the number of priming stimuli/bystanders (0 vs. 3 vs. 6), and the between-subject factor of the priming stimuli type (happy faces, participant N = 28 vs. neutral faces N = 27 vs. triangles N = 30). The within-subject design was balanced so that each six-factor combination occurred equally often.

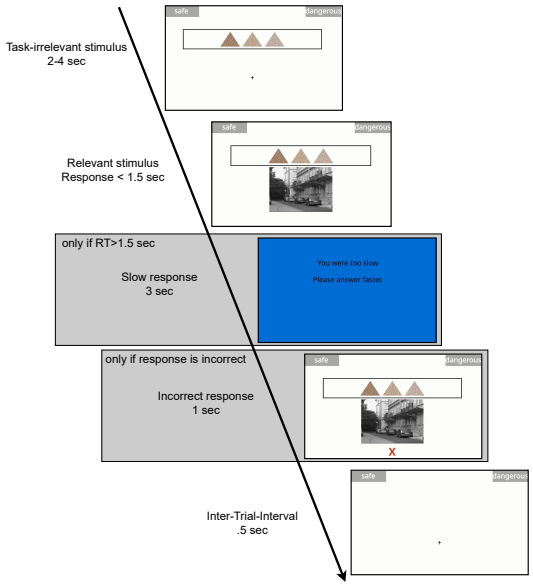

Figure 2: Screens within one trial.

Data analysis

We estimated Hierarchical Bayesian multilevel models to predict the diffusion model parameters and the logarithmic reaction time. The Bayesian approach allowed us to simultaneously estimate the DDM parameters and calculate common regression models. A disadvantage of the Bayesian approach is the necessity of defining priors that could influence the modelling results. In our modelling, we dealt with this issue by defining little-informative (and practically *uninformative*) priors. For each dependent variable, the multilevel models followed the same structure, with the parameters being predicted by a subject-specific intercept and the within-subject variables (scene type *st* - dichotomous; number of priming stimuli *np* - 3 levels). Additionally, the subject-specific intercept was predicted by a general intercept μ and the between-subjects variable (priming stimuli type *pt* - 3 levels). For a given person *i* and the predicted variable/parameter *y*, the following model was calculated:

$$y \sim \beta_{0i} + \beta_{1.} st + \beta_{2.} np + \varepsilon_{0.}$$
$$\beta_{0i} \sim \mathcal{N}(\mu, \sigma^2) + \beta_{.1} pt + \varepsilon_{.0}$$

In order to improve the signal-to-noise ratio and avoid fast guesses, we excluded trials with $RT < 300ms$ and missing answers from the DDM and RT analysis.

We intentionally chose not to impose any theory-based model restrictions because our new DDM-based explanation for the bystander effect is first to be tested through data. This approach allowed us to account for our uncertainty in the effect by allowing the different conditions to potentially affect any of the parameters in the model (as recommended by Pleskac et al., 2018). For the DDM, we used participants' RT in milliseconds and their decisions (upper = dangerous, lower = safe) as input variables. This implied a positive evidence accumulation for dangerous scenes (*st* = dangerous) and a negative evidence accumulation for safe scenes. As such, we estimated two drift rates (delta safe and delta danger).

For the DDM parameters, we estimated the *st* predictor only for the non-decision time tau. We did not include it as a regressor for the other parameters because the decision process focused on danger detection. The threshold parameter α accounted for biases towards one alternative by incorporating the model parameter β. For the drift rate *v*, different values were expected in dangerous vs non-dangerous situations, as the drift goes in opposite directions. Therefore, we accounted for this by estimating two independent *v* parameters.

We conducted model estimations using the R-implementation of JAGS (Su & Yajima, 2021), with the JAGS implementation of the Wiener distribution (Wabersich & Vandekerckhove, 2014) for the DDM. DDM parameters were estimated along with the multi-level predictors. For all analyses, MCMC samples were drawn with six chains, each consisting of 100 000 iterations, of which 25 000 were burn-in. The thinning rate was set to 10. At the end of the model estimations, all parameters had an R-Hat < 1.01.

We report the mean parameters along with their 95% Highest Density Interval. We further provide the probability of direction (*pd*; Makowski, Ben-Shachar, & Lüdecke, 2019), which indicates the likelihood that the effect goes in one specific direction. A *pd* value can be between .5 and 1, with values closer to 1 indicating a higher probability. Effects are expected to likely exist with $pd > .97$ (Makowski et al., 2019).

Results

We analysed the logarithmic reaction time and found a significant effect towards slower reaction times for smiling faces *sf* but not for neutral faces *nf*. However, both face types tend to have a slower reaction time, consistent with our first hypothesis. We further found slightly faster RTs when the stimulus to classify was dangerous *d* or when 3 or 6 priming stimuli (*p3* or *p6*) were present. The detailed Bayesian multilevel regression results are described in the following regression equation, with the mean weights in the first, the 95% HDI in the second and the probability of direction in the third line.

$$\begin{array}{lccc} log(RT) \sim \beta_{0i} - .02 * d & -.01 * p3 & -.02 * p6 \\ & [-.02,-.01] & [-.02,-.01] & [-.02,-.01] \\ & pd > .999 & pd > .999 & pd > .999 \end{array}$$

$$\begin{array}{lccc} \beta_{0i} \sim -.49 & +.09 * sf & +.04 * nf \\ & [-.53,-.44] & [.02,.16] & [-.03,.11] \\ & pd > .999 & pd = .996 & pd = .881 \end{array}$$

Our proposed model (see Figure 1) hypothesised only effects on the drift rate for the decision to help. Results are listed in Table 1. As expected, we found no differences between conditions for the threshold α or the bias β. Notably, β had a value of .5, indicating no bias towards the safe or risky option. For the drift rates, we did not find an effect for the drift towards the safe option. However, we did find a negative drift effect when the priming stimuli were smiling

Table 1: Main analysis covering the whole 2 x 3 x 3 mixed design. Bayesian random intercept fixed slope model on the diffusion model parameters. Each column represents a predictor. The mean and a 95% highest density interval (HDI, in square brackets) are listed for each estimation. Stars represent the probability of direction *pd*.

	alpha	beta	delta safe	delta danger	tau
global estimates					
μ	1.53*** [1.44, 1.62]	.50*** [.48, .52]	-2.84*** [-3.06, -2.63]	3.07*** [2.84, 3.34]	.38*** [.36, .39]
between subjects effects					
smiling face	.05 [-.07, .17]	.00 [-.02, .03]	.16 [-.14, .46]	-.37* [-.72, -.02]	.03** [.01, .06]
neutral face	.05 [-.08, .17]	.01 [-.01, .03]	.21 [-.08, .52]	-.34* [-.69, .01]	.00 [-.02, .03]
within subjects effects					
danger	–	–	–	–	.00 [-.01, .00]
prim. 3	-.01 [-.05, .03]	.00 [-.02, .01]	-.03 [-.14, .08]	-.05 [-.16, .06]	-.01** [-.01, .00]
prim. 6	-.03 [-.06, .01]	.00 [-.02, .01]	.04 [-.07, .16]	-.01 [-.12, .09]	-.01** [-.01, .00]

Note: probability of direction .97 * .99 ** .999 *** 1

faces ($-.37[-.72, -.02]$) or neutral faces ($-.34[-.69, .01]$). Following our second hypothesis, the average evidence accumulation speed slowed down when faces were present. This effect is stable and similar whether the faces are smiling or neutral. However, the number of priming stimuli did not affect the drift rate.

The non-decision parameter tau was negatively affected by smiling face primes ($.03[.01, .06]$) but not by neutral face primes ($.00[-.02, .03]$). This means that the non-decision time takes longer if the face is smiling but not if it is neutral. The non-decision time was slightly shortened by the presence of 3 ($-.01[-.01, .00]$) and 6 ($-.01[-.01, .00]$) priming stimuli.

Discussion

We proposed a cognitive model of the bystander effect from a reflexive bottom-up perspective. We argued that the decision to intervene may be understood as a drift-diffusion process. The mere anticipated presence of others slows this process down, leading to a lower average drift. To test this theory, we mapped our approach onto the drift-diffusion model. We then conducted a Bayesian hierarchical multilevel analysis in which we allowed all factors used in the current analysis to affect the model parameters. We confirmed our model assumptions and showed that the bystander effect may partially be explained by a reflexive bottom-up process that leads to a slower information uptake in the presence of others.

Our models' central hypothesis is that bystanders' presence causes a slowing in the evidence accumulation. Based on pure reaction time data, we only found statistically relevant effects for facial stimuli incongruent with the emergency, such as smiling faces. However, when analysing the data with the DDM, we found similar negative effects on the drift towards the decision 'danger detected' in emergency-congruent and emergency-incongruent facial stimuli. This supports our proposed mapping of the slowed decision-making due to an increased system I process (Hortensius & de Gelder, 2018) and is further in line with previous studies that found social influence (Germar et al., 2014) and social expectation (Pleskac et al., 2018) effects on the drift rate. The emergency-incongruent stimuli also positively affected the non-decision time. This could indicate a freezing response but requires further investigation. In combination with the drift rate effect, this may have further led to a significant effect on RT.

We further found small negative effects of priming stimuli on the non-decision time. We do not have a definitive explanation for this effect but propose that it might be related to an increased level of action preparation when seeing something compared to when seeing only the fixation cross. Systematic investigation of this non-decision time effect is required.

Our study has limitations that we will discuss in the following. First, we did not find effects of the number of priming stimuli on the drift rate. This speaks against a classic bystander interpretation, which says that the more people are present, the stronger the effect. However, it may align with previous literature suggesting that priming of bystanders is already enough for the effect to occur (Garcia et al., 2002). Therefore, the implicit bystander effect may be interpreted as a priming effect. However, our results regarding this are only preliminary and may require further exploration in future.

Second, for the priming stimulus type, we only used neutral (emergency-congruent) and smiling (emergency-incongruent) faces. However, introducing other facial expressions such as fear or anger would cover the emotion spectrum more comprehensively and allow a better understanding of the impact of bystanders' emergency congruency on the bystander effect. However, given our current results, we attribute the main bystander effect to differences in the evidence accumulation speed (i.e. drift rate v). These effects were stable independently of the facial expression.

Third, we only conducted a fire detection task and found effects on the drift rate, which we interpret as a bystander effect. However, it is still to be explored how far the found effect generalises. Evidence accumulation speed could slow down 1) only for fire detection, 2) for danger detection in general, or 3) for any task. In line with a pure bystander effect, we hypothesise that the effect generalises over 2) danger detection in general but not over 3) other tasks. Future research deploying different task types is required to draw conclusions about the generalisability.

Acknowledgments

This work was supported by the Shenzhen University-Lingnan University Joint Research Programme (#SZU-LU003/2324) to Yi Huang. Funders had no role in study design, data collection, data analysis, data interpretation, writing of the manuscript, or the decision to submit.

References

Alexandrowicz, R. W. (2020). The diffusion model visualizer: an interactive tool to understand the diffusion model parameters. *Psychological Research*, *84*(4), 1157–1165.

Barron, G., & Yechiam, E. (2002). Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, *18*(5), 507–520.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4, Pt.1), 377–383.

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., … Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, *137*(4), 517–537.

Garcia, S. M., Weaver, K., Moskowitz, G. B., & Darley, J. M. (2002). Crowded minds: The implicit bystander effect. *Journal of Personality and Social Psychology*, *83*(4), 843–853.

Germar, M., Schlemmer, A., Krug, K., Voss, A., & Mojzisch, A. (2014, February). Social influence and perceptual decision making: A diffusion model analysis. *Personality and Social Psychology Bulletin*, *40*(2), 217–231.

Graziano, W. G., & Habashi, M. M. (2010). Motivational processes underlying both prejudice and helping. *Personality and Social Psychology Review*, *14*(3), 313–331.

Havlik, J. L., Vieira Sugano, Y. Y., Jacobi, M. C., Kukreja, R. R., Jacobi, J. H. C., & Mason, P. (2020, July). The bystander effect in rats. *Science Advances*, *6*(28), eabb4205.

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573.

Hortensius, R., & de Gelder, B. (2014). The neural basis of the bystander effect - The influence of group size on neural activity when witnessing an emergency. *NeuroImage*, *93*, 53–58.

Hortensius, R., & de Gelder, B. (2018). From empathy to apathy: The bystander effect revisited. *Current Directions in Psychological Science*, *27*(4), 249–256.

Hortensius, R., Schutter, D. J. L. G., & de Gelder, B. (2016). Personal distress and the influence of bystanders on responding to an emergency. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(4), 672–688.

Johnson, D. J., Hopwood, C. J., Cesario, J., & Pleskac, T. J. (2017). Advancing research on cognitive processes in social and personality psychology: A hierarchical drift diffusion model primer. *Social Psychological and Personality Science*, *8*(4), 413–423.

Latané, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, *10*(3), 215–221.

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: why doesn't he help?* Englewood Cliffs, NJ: Prentice-Hall.

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, *89*(2), 308–324.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Machackova, H., Dedkova, L., & Mezulanikova, K. (2015). Brief report: The bystander effect in cyberbullying incidents. *Journal of Adolescence*, *43*(1), 96–99.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, *4*(40), 1541.

Markey, P. (2000). Bystander intervention in computer-mediated communication. *Computers in Human Behavior*, *16*(2), 183–188.

Niloy, F. F., Arif, Nayem, A. B. S., Sarker, A., Paul, O., Amin, M. A., … Rahman, A. M. (2021). A novel disaster image dataset and characteristics analysis using attention model. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 6116–6122). (arXiv:2107.01284 [cs])

Pleskac, T. J., Cesario, J., & Johnson, D. J. (2018). How race affects evidence accumulation during the decision to shoot. *Psychonomic Bulletin & Review*, *25*(4), 1301–1330.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Rudnicki, K., Vandebosch, H., Voué, P., & Poels, K. (2023). Systematic review of determinants and consequences of bystander interventions in online hate and cyberbullying among adults. *Behaviour & Information Technology*, *42*(5), 527–544.

Sinke, C., Sorger, B., Goebel, R., & de Gelder, B. (2010). Tease or threat? Judging social interactions from bodily expressions. *NeuroImage*, *49*(2), 1717–1727.

Solomon, R. L., & Corbit, J. D. (1974). An opponent-process theory of motivation: I. Temporal dynamics of affect. *Psychological Review*, *81*(2), 119–145.

Su, Y.-S., & Yajima, M. (2021). R2jags: Using r to run 'jags' [Computer software manual]. (R package version 0.7-1)

Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior Research Methods*, *46*(1), 15–28.