

Coffee or Date? A Bayesian Account of Indirect Disclosure in High-Stakes Social Interactions

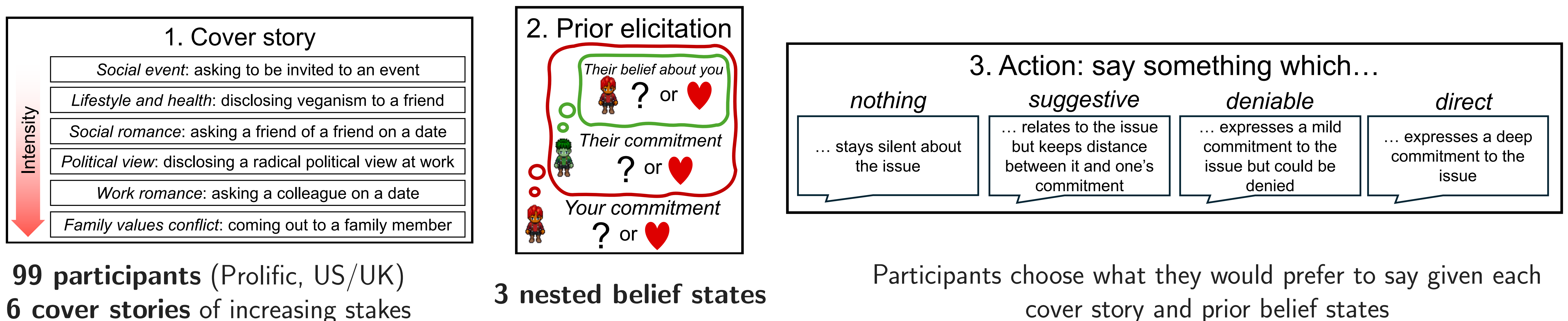
Victor Btesh^{1*} David Lagnado¹

¹Department of Experimental Psychology, University College London *victor.btesh.19@ucl.ac.uk

Introduction

- We rarely say what we mean, especially when unsure what others think or feel. Coffee instead of a date, a question instead of a claim.
- Game-theoretic accounts (Pinker et al., 2008; Qian et al., 2024) show how deniability can be optimal, but not how the games motivating it emerge from what the speaker thinks the listener wants, knows, and believes about them.
- What computations underpin how we adapt our actions to the specific mental states of others? What costs and rewards are we optimising for in high-stakes social interactions?

Experiment



Computational Model

A pragmatic listener L_1 infers the speaker's commitment from action a and chooses a binary response $r \in \{-, +\}$, where positive is coded 1, as follows:

$$U_{L_1}(r; a, c) = r [\Psi_{ac} + \log p(c_{\text{speaker}}=c | a)]$$

The sophisticated speaker S_2 reasons about this listener and weights three motives when choosing how directly to disclose:

Positive response

Wants the listener to respond well to action a

Strategic ambiguity

Wants to keep the listener uncertain about their true commitment c

Motivated alignment

Wants to learn about the listener's commitment, insofar as it matches their own

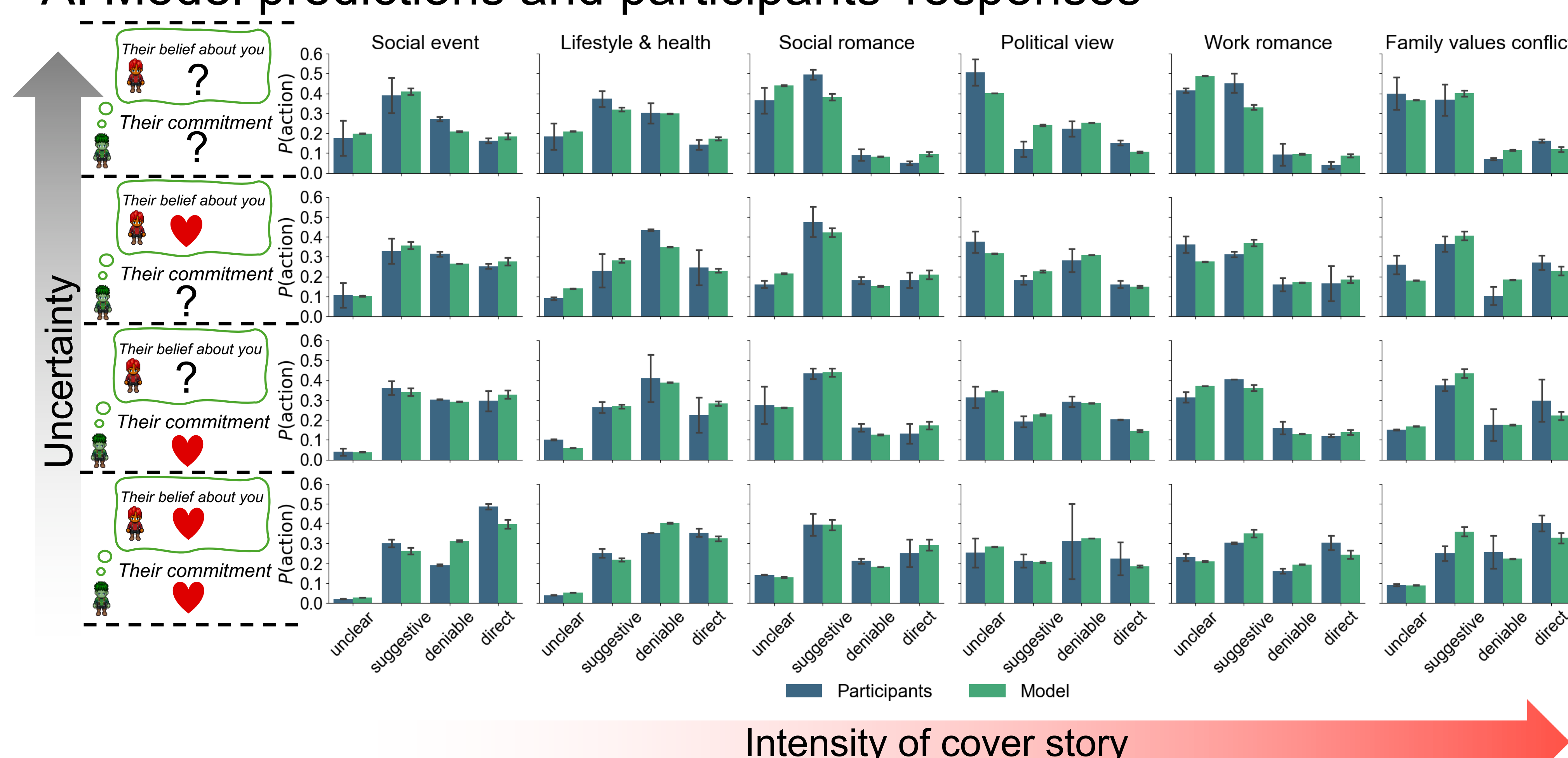
$$U_{S_2}(a, c; \rho) = \alpha_{\text{positive}} \log p(r=+ | a, c) + \alpha_{\text{ambiguity}} H[\rho(c), p_{L_1}(c | a)] - \alpha_{\text{align}} \mathbb{E}[D_{KL}[p(c_{L_1} | r) || \rho(c)]]$$

Actions endorse commitments to different degrees (Ψ)

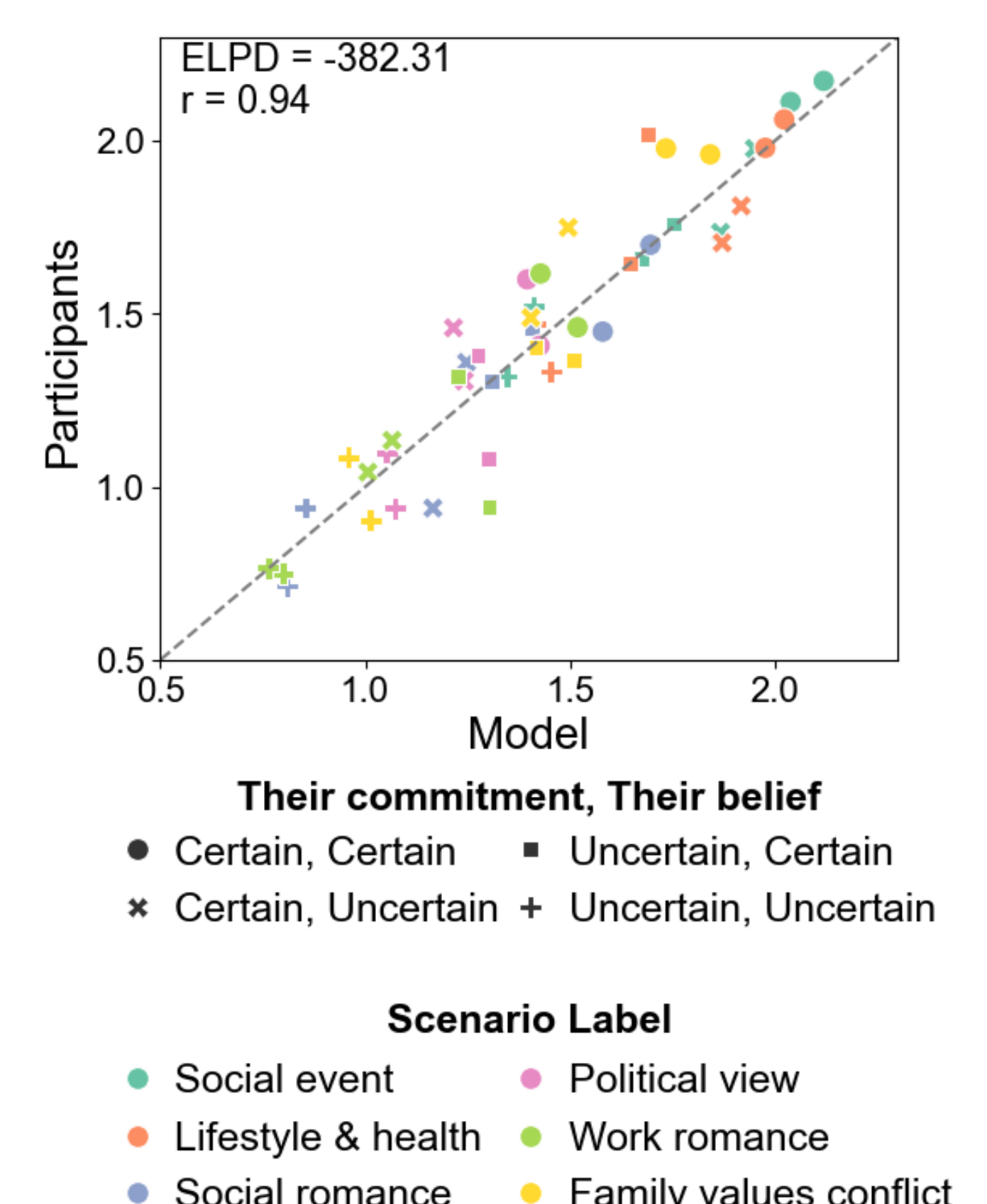
	Indifferent		Committed	
Silence	+1	+1	0	0
Suggestive	-1	+1	+1	0
Deniable	-1	-1	+1	+1
Direct	-1	-1	-1	+1

Results

A. Model predictions and participants' responses



B. Model fit



Discussion

- Across 6 cover stories, the model captures a graded shift from direct to indirect action as uncertainty rises – people track information-theoretic quantities over the listener's beliefs.
- But stakes aren't only epistemic. In romance or family conflict, one exchange reshapes the whole relationship (Lee Pinker, 2010). Beliefs cache longer-horizon goals.
- Future work should capture how the stakes of the relationship itself – not just the single exchange – shape these goals.



Scan for extended abstract