

Introduction

Can groups perform actions and take responsibility for their consequences? If so, in what sense? Apart from being described as acting and responsible, can a group also be described as possessing beliefs, intentions, and desires? Answers to these questions determine philosophical positions in the debate on collective agency and intentionality and thus are central to social philosophy and metaphysics.

Two lines of division between the theoretical positions:

- **Realism** - groups are intentional agents, capable of acting, possessing beliefs, desires and knowledge and holding responsibility for their actions.
- **Irrealism** - groups are merely collections of individuals and should not be regarded as intentional agents.

Furthermore, **realist** views may be divided into two categories:

- **Distributivism** - groups are intentional agents *qua* its members.
- **Collectivism** - groups *qua* groups are independent intentional agents.

All these stances have been defended by appealing to commonly held philosophical intuitions about groups' agency and intuitive necessary conditions of intentional action and possession of propositional attitudes (Goldman 2002, Tollefsen 2002, Tuomela 2007). The use of such arguments calls for experimental investigation into the folk concept of group agency. There is a problem though how to properly measure the level of group intentionality or agency in folk ascriptions.

The Group Knobe Effect as a Measure of Group Agency

The Knobe Effect - an asymmetry in ascription of intentionality of bringing about negative or positive side-effects of an action first described by Joshua Knobe (2003). Negative side-effects are given a higher intentionality rating by the laypeople than the positive side-effects of the same action.

Michael and Sziget (2019) proposed using Knobe Effect to investigate the folk concept of group agency. They assumed the Knobe asymmetry might be an indicator of intentionality and tested whether it can be observed, when the story's protagonist is a group agent (**Group Knobe Effect, GKE**), which would align with realist intuitions. Furthermore, they hypothesized that GKE can be used to spot the difference between distributivist and collectivist intuitions, too. According to M&S this can be found by measuring GKE levels in participants' ascriptions of intentionality to group agents and to their individual members, in two separate settings: when the members supported the group decision or dissented from it.

However, according to Mizumoto (2018) the original Knobe Effect concerning intentionality ascriptions might be considered as a linguistic default and thus discredited as intentionality indicator in general. This threat does not concern knowledge and belief ascriptions, which exhibit analogous asymmetry in individual cases (Beebe and Buckwalter 2010, Beebe 2013). We decided to extend the original experiment to cover belief and knowledge ascriptions and seek for **Group Epistemic** and **Doxastic Knobe Effects (GEKE and GDKE)** as more robust evidence of group agency. Furthermore, we noticed that mere difference in intentionality ascriptions to group members in support and dissent conditions might be accounted for by what we called the **Hypocrisy Effect**, which consists in giving lower intentionality ratings to agent's actions undertaken despite agent's dissent. If the respondents were to be counted as collectivists, the Hypocrisy Effect should show on individual, but not on the group level of ascription. Only this difference is a proper measure of collectivism.

Hypotheses

(H1) We will find the Knobe asymmetry in folk intentionality attributions to group agents (i.e., we will corroborate M&S's initial findings).

(H2) We will find the Knobe asymmetry in folk knowledge and belief attributions to group agents (i.e., evidence of GEKE and GDKE).

(H3) Realist intuitions concerning group agents' intentions, knowledge, and beliefs exhibited by laypeople will be partly in line with the distributivist interpretation and partly with the collectivist interpretation (some will show collectivist intuitions, while some will show distributivist intuitions).

The Group Knobe Effect Revisited

Epistemic and Doxastic Side-Effect Effects in Intuitive Judgments Concerning Group Agents

Maciej Tarnowski, Adrian Ziółkowski, Mieszko Tałasiewicz

Study 1: Group Epistemic and Doxastic Knobe Effects

In our first study, following Michael and Sziget, we adopted a simple between-subjects design with one dichotomous independent variable (Harm/Help condition). The following story was presented depending on the condition:

ACME Inc. started a new program. When launching the new program, data suggested that the program would help ACME Inc. increase profits, but that it would also [harm/help] the environment. In line with ACME Inc.'s business policies and in the interest of maximizing profits, the new program was implemented. Sure enough, the environment was [harmed/helped].

Questions

- On a scale from 0–8, to what extent would you agree that ACME Inc. **intentionally [harmed/helped]** the environment?
- On a scale from 0–8, how much **[blame/praise]** do you think ACME deserves for [harming/helping] the environment?
- On a scale from 0–8, to what extent would you agree that ACME Inc. **knew** that it was [harming/helping] the environment?
- On a scale from 0–8, to what extent would you agree that ACME Inc. **believed** that it was [harming/helping] the environment?

Participants

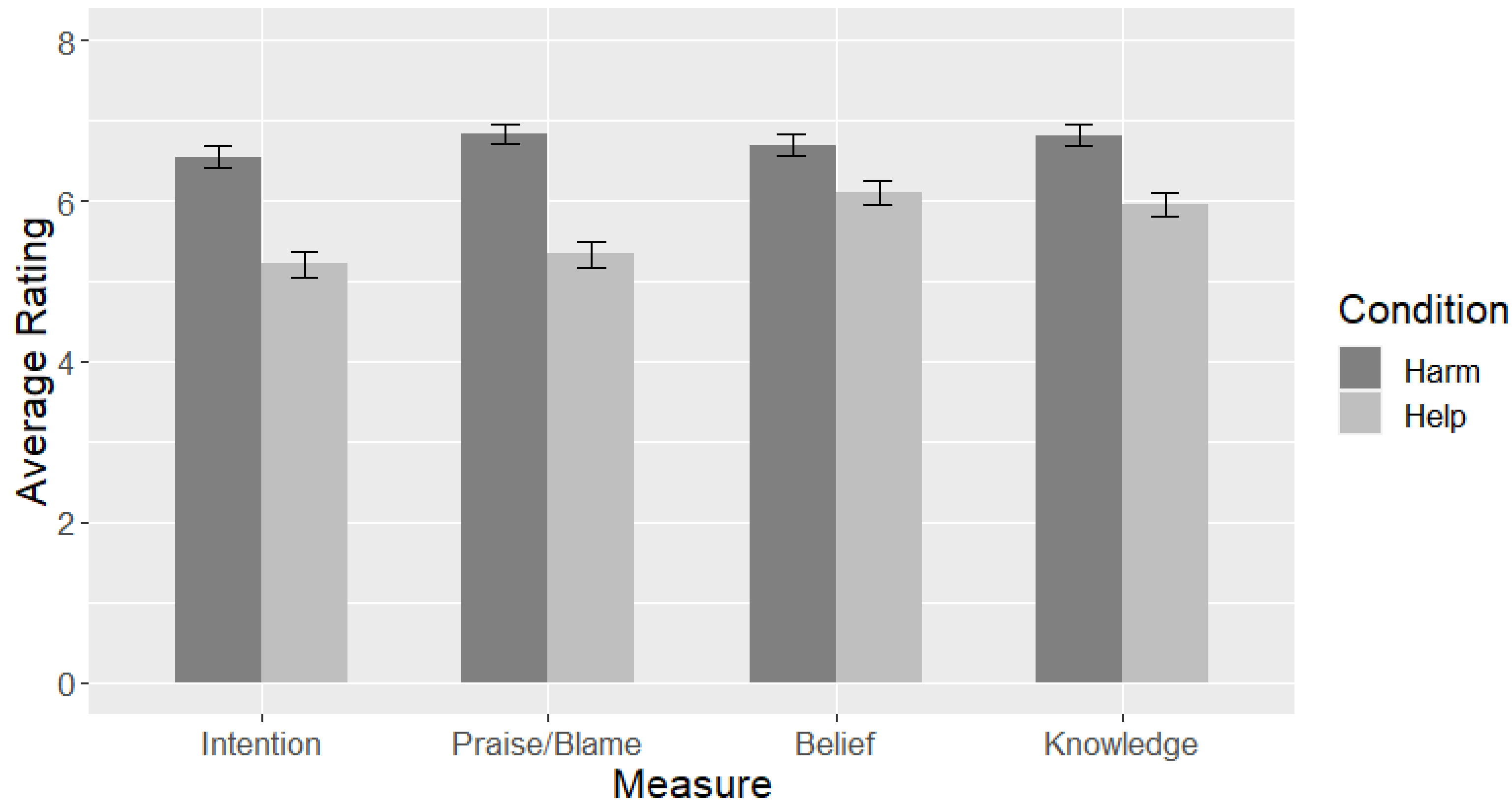
N = 289 (140 in the Harm condition and 149 in the Help condition). 55.7% of the participants were male, 41.5% were female, and 2.7% indicated they were non-binary or chose not to disclose their gender. Average age was 38.45, with SD = 12.53.

Results

- **GKE for intentionality ascriptions:** we found a significant difference in the predicted direction between the Harm and Help conditions; $t(287) = 6.15$; $p < 0.001$; $d = 0.73$.
- **GKE for blame/praise ascriptions:** $t(287) = 7.16$; $p < 0.001$; $d = 0.85$.
- **GEKE (knowledge ascriptions):** we found an epistemic counterpart of the Knobe Effect; $t(287) = 4.38$; $p < 0.001$; $d = 0.51$.
- **GDKE (belief ascriptions):** we found a doxastic counterpart of the Knobe Effect; $t(287) = 3.02$; $p = 0.003$; $d = 0.36$.

Conclusion

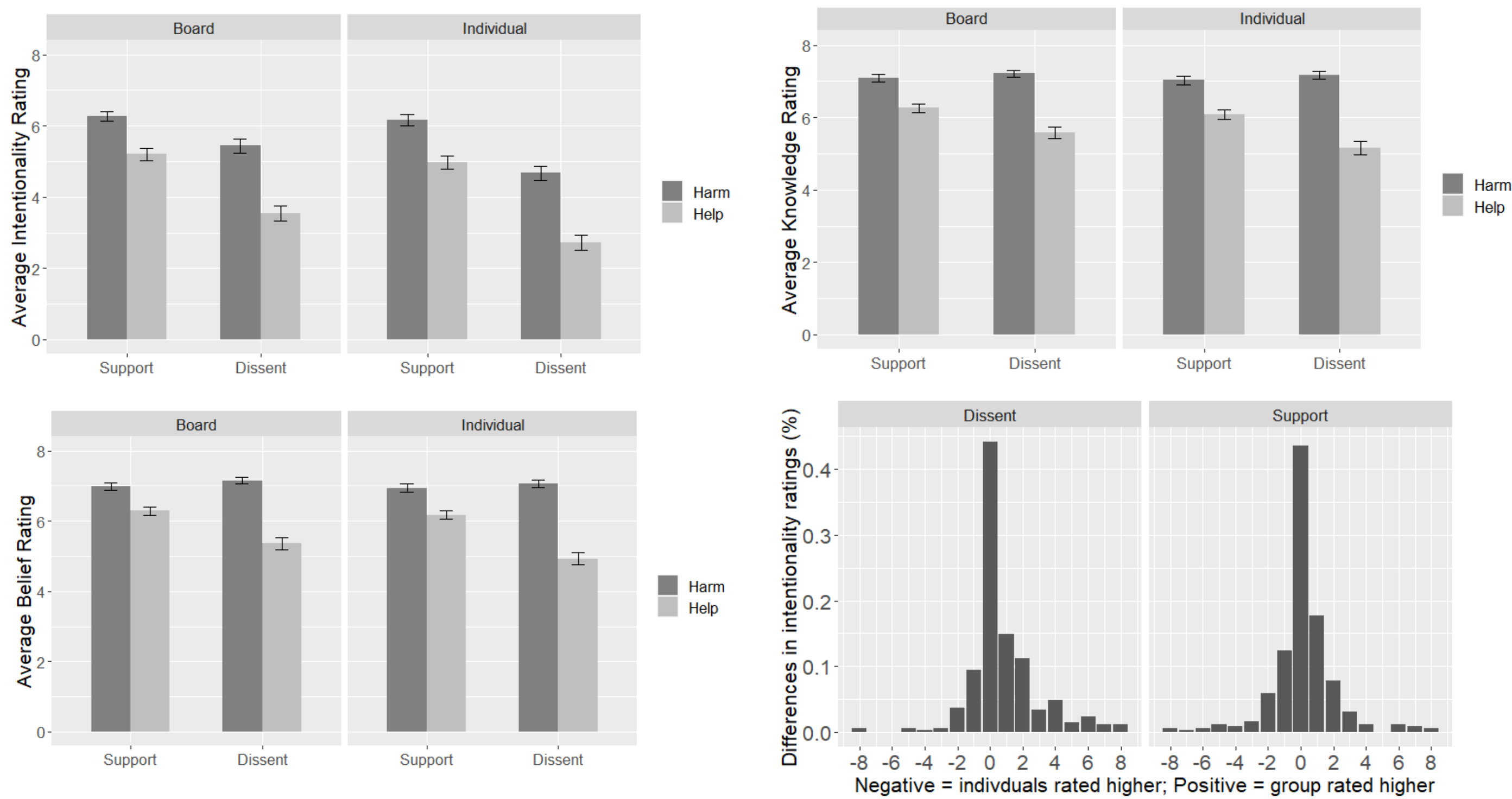
We corroborated original M&S findings regarding the existence of GKE for intentionality and blame/praise ascriptions (H1), as well as found evidence of its epistemic and doxastic counterparts (H2).



Average ratings for each measure included in Study 1 depending on the experimental condition (error bars represent mean SE).

Agency and Intention in Language 2 Workshop Harvey Mudd College (Online) 14.01.2022

Study 2: Distributivism vs. collectivism



In the second study we employed a 2x2 between-subjects design with the following factors: Harm/Help (the standard Knobe-style manipulation) and Support/Dissent. The following story was presented depending on the condition:

Representatives from the research and development department of a company reported to the board and said, “We are thinking of starting a new program. It will help us increase profits, [but/and] it will also [harm/help] the environment.” The board consisted of three members: Benson, Franklin and Sorel. [For various reasons, each of them personally opposed the program and tried to prevent it from being implemented. However/Each of them personally supported the program and did not object to its being implemented. In any case], they were obliged to follow the board's standard decision-making protocol, which left no opportunity for their personal views to influence the decision. As a result, in line with the company's business policies and in the interest of maximizing profits, the new program was implemented. Sure enough, the program was highly profitable and the environment was [harmed/helped].

Questions

- On a scale from 0–8, to what extent would you agree that **the board** [knew/believed] that the environment will be [harmed/helped] / intentionally [harmed/helped] the environment?
- On a scale from 0–8, to what extent would you agree that **Benson, Franklin and Sorel** [knew/believed] that the environment will be [harmed/helped] / intentionally [harmed/helped] the environment?

Participants

N = 650 (Harm: n = 335; Help: n = 315; Dissent: n = 329; Support: n = 321). Fifty-two percent of respondents were male, 47.1% were female, and 0.9% chose the option “non-binary” or “other”. Average age was 40.08 (SD = 12.49).

Results

- We obtained further evidence of **GKE** ($F(1, 646) = 64.69$; $p < 0.001$; $\eta^2 = 0.091$), **GEKE** ($F(1, 646) = 96.78$; $p < 0.001$; $\eta^2 = 0.13$) and **GDKE** ($F(1, 646) = 107.48$; $p < 0.001$; $\eta^2 = 0.143$) for folk evaluations of the board as a group agent.
- **The Hypocrisy Effect** (difference of intentionality ratings between the Support and Dissent conditions) appeared for both the ratings concerning the group agent: $\eta^2 = 0.065$ and the individual members of the board: $\eta^2 = 0.129$. However, the group HE was smaller than the individual HE.
- Nearly half of the participants rated the intentions of the group and individuals **identically** (both in Support and Dissent conditions - see the graph on the lower right).

Conclusion

We corroborated the existence of GKE, GEKE and GDKE (H1 & H2) which contributes to the evidence of generally realist intuitions concerning group agency. As to whether this realism is of collectivist or merely distributivist nature, our results came out ambiguous. HE on individuals was stronger than on groups (which speaks for collectivism), while on groups it was still significant (which speaks for distributivism). Moreover, any difference between a group and its members in intentionality ratings, required for evidence of collective interpretation, was acknowledged only by half of the participants, which corroborates our H3 - that intuitions about collectivism/distributivism are mixed. There might be individual differences between laypersons when it comes to distributive and collectivist intuitions regarding group agency.