Philosophical Explorations, 2018
https://doi.org/10.1080/13869795.2018.1492007

"The Group Knobe Effect": evidence that people intuitively attribute agency and responsibility to groups

John Andrew Michael[a,b] and András Szigeti[c,d]*

[a]*Department of Philosophy, University of Warwick, Coventry, UK;* [b]*Department of Cognitive Science, Central European University, Budapest, Hungary;* [c]*Department of Philosophy, Linköping University, Linköping, Sweden;* [d]*Department of Philosophy, Lund University, Lund, Sweden*

(*Received 28 August 2016; final version received 4 May 2018*)

In the current paper, we present and discuss a series of experiments in which we investigated people's willingness to ascribe intentions, as well as blame and praise, to groups. The experiments draw upon the so-called "Knobe Effect". Knobe [2003. "Intentional action and side effects in ordinary language." *Analysis* 63: 190–194] found that the positiveness or negativeness of side-effects of actions influences people's assessment of whether those side-effects were brought about intentionally, and also that people are more willing to assign blame for negative side-effects of actions than they are to assign praise for positive side-effect of actions. Building upon this research, we found evidence that the positiveness or negativeness of side-effects of group actions influences people's willingness to attribute intentions to groups (Experiment 1a), and that people are more willing to assign blame to groups for negative side-effects of actions than they are to assign praise to groups for positive side-effects of actions (Experiment 1b). We also found evidence (Experiments 2a, 2b, 3 and 4) that the "Group Knobe Effect" persists even when intentions and blame/praise are attributed to groups non-distributively, indicating that people tend not to think of group intentions and group blame/praise in distributive terms. We conclude that the folk are collectivist about group intentions, and also about the blameworthiness and praiseworthiness of groups.

Keywords: collective responsibility; collective agency; Knobe Effect; blame; praise; collective intentions

1. Introduction

In recent decades, there has been a resurgence of the theoretical debate about the legitimacy of ascribing agency and responsibility to collectives. Collectivism, in our understanding, is the view that groups can be agents and can be morally responsible *qua* groups. For a responsibility-collectivist, it can be true that you have assigned all individual responsibility there is to assign for an outcome or an action to the individual members of the group (and also to non-members if relevant) – and yet there will be some responsibility left which is to be assigned directly to the group. In other words, collectivists are committed to group responsibility of a kind that is responsibility over and above the responsibility of individuals. Similarly, for an agency collectivist, talk about group action is not a convenient shorthand to

*Corresponding author. Email: andras.szigeti@liu.se

refer to a set of some number of individual actions. Agency-collectivists think that some actions are literally carried out by groups. In such cases, action-attributions to the group cannot be substituted by action-attributions to one or more individuals. Individualism rejects both responsibility-collectivism and agency-collectivism.[1]

Typically, the labels "individualism" and "collectivism" as defined above are used to refer to opposing views in the ethics and metaphysics of agency. Our concern in this paper, however, pertains to folk psychology. To what extent is folk psychology collectivist? We believe that this question is of interest because, despite the importance of legal and political judgments and decisions concerning group agents, relatively little is known about the psychological mechanisms underpinning people's moral evaluations of groups.

In the following, we address this gap in the literature by presenting and discussing a series of experiments in which we investigated people's willingness to ascribe intentions, as well as blame and praise, to groups. In order to do this, we drew upon the so-called "Knobe Effect": Knobe (2003) found that the positiveness or negativeness[2] of side-effects of actions influences people's assessment of whether those side-effects were brought about intentionally, and also that people are more willing to assign blame for negative side-effects of actions than they are to assign praise for positive side-effect of actions (see also Nadelhoffer 2005, 2006; Knobe 2006; Nichols and Ulatowski 2007; Malle 2010; Cova 2015). Building upon this body of research, we aimed to answer the question whether a "Group Knobe Effect" (GKE) also obtains, i.e. whether the positiveness or negativeness of side-effects of group actions influences people's willingness to attribute intentions to groups (Experiment 1a), and whether people are more willing to assign blame to groups for negative side-effects of actions than they are to assign praise to groups for positive side-effects of actions (Experiment 1b).

As we shall explain in a moment, the findings from Experiments 1a and 1b provide strong evidence that the GKE indeed obtains. But there is also a deflationary explanation of why the GKE obtains, namely that people attribute intentions and blame/praise *distributively*. According to this explanation, the GKE is just the aggregate of the Knobe Effect for individual members. People attribute intentions and blame/praise to every group member in accordance with what the Knobe Effect predicts, so it is hardly surprising that we observe the Knobe Effect for groups as well. Experiments 2a, 2b, 3 and 4 were therefore designed to test the plausibility of such a distributivist explanation of the GKE. The results of these experiments indicate that the GKE persists even when intentions and blame/praise are attributed to groups non-distributively, and thus provide evidence that people tend not to think of group intentions and group blame/praise in distributive terms. This suggests that the folk are collectivist about group intentions, and also about the blameworthiness and praiseworthiness of groups.

Why approach the issue of folk collectivism via GKE? In our view, the existence of non-distributive GKE is in itself an interesting finding. It provides further evidence of the robustness of the Knobe Effect. Feltz (2007) notes that the Knobe Effect has been replicated not only for various kinds of side-effects (Knobe and Mendlow 2004; Nadelhoffer 2004a, 2004b; Feltz 2007) but also for different cultures (Knobe and Burra 2006) and age groups (Leslie, Knobe, and Cohen 2006). We show in this paper that it can also be replicated for *different kinds of agents*. Moreover, we see at least three reasons why testing for GKE is a particularly useful way of investigating whether folk psychology is collectivist. First, non-distributive GKE is evidence that people's judgments of individual and group agency run parallel insofar as asymmetries of intention-attribution and blame/praise-attribution are not significantly affected when the target of the attribution is a group rather than an individual. The best explanation of the persistence of this asymmetry is that people are collectivist about agency and moral responsibility. Secondly, let us

consider for the sake of the argument a situation in which people (contrary to our results) did not show the double asymmetry characteristic of GKE, but nevertheless were found to attribute intention and moral responsibility to groups in some way.[3] This would show that different evaluative mechanisms shaped people's reasoning about individual and group actions. If that were the case, then it would not be true that people treat individual agency and group agency on a par. Moreover, it would indicate that people's reasons for ascribing intentions and moral responsibility to individuals and groups could be quite different. Consequently, this situation would allow individualists to help themselves to the popular argument that the reason why people sometimes go collectivist is epistemic or pragmatic: we address the group only because it is too difficult to ascertain which group member did what exactly and how much responsibility we should ascribe to each – but nobody really thinks that the group *qua* group can act or be morally responsible. But we did find the double asymmetry was quite robust and so this argument does not seem readily available to the individualist. Third, we will see below that focusing on GKE has also been heuristically useful in designing the experiments aimed at teasing out people's intuitions about groups. In particular, by tracking the double asymmetry, we could more confidently exclude the possibility that our participants were thinking about aggregates of individuals rather than about groups as such.

While our findings support the hypothesis that folk psychology is collectivist, it must be acknowledged that this does not provide direct evidence in support of collectivism or against individualism: the folk, of course, may simply be wrong. But one important implication of our findings is that individualism is more revisionist than may otherwise have been assumed. Thus, individualists cannot appeal to folk intuitions to buttress their own position. Indeed, they will not only have to provide positive arguments in favor of their position but will have the additional burden of supplying an error theory to account for the divergence between folk psychology and their own philosophical position. We will address these issues in the final section of the paper.

2. The experiments

2.1. *The GKE*

Experiment 1 was designed to investigate whether the Knobe Effect generalizes to groups. Drawing upon the well-known experiments reported in Knobe (2003), we aimed to test whether the positiveness or negativeness of side-effects of group actions influences people's willingness to attribute intentions to groups (Experiment 1a), and whether people are more willing to assign blame to groups for negative side-effects of actions than they are to assign praise to groups for positive side-effects of actions (Experiment 1b).

Because we were concerned that asking participants whether a group agent had brought about a side-effect intentionally may bias their response to a subsequent question about whether the group deserved blame or praise for the side-effect, and vice versa, we opted to address these questions in separate experiments with different participants. First, in Experiment 1a, we asked whether the group agent had brought about the side-effect intentionally. Next, in Experiment 1b, we asked whether the group agent deserved praise/blame for the side-effect.

2.1.1. *Experiment 1a*

In the first experiment, we tested the hypothesis that the positiveness or negativeness of a side-effect influences people's judgments about whether a group agent brought about that

side-effect intentionally. With the help of *SurveyMonkey Audience*, we collected responses from 65 participants (36 female, mean age = 47.58, *SD* = 16.5, range = 18–75), each of whom received a small monetary reward for their participation.

As in Knobe's original experiments, participants were randomly assigned either to the "harm condition" or to the "help condition". Participants in the harm condition read the following brief vignette:

> 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 it will also harm the environment." The reply from the board was, "We don't care at all about harming the environment. We just want to make as much profit as we can. Let's start the new program." Sure enough, the program harmed the environment.

These participants then responded to the following question: "On a scale from 0–8, to what extent would you agree that the board intentionally harmed the environment? (0 = disagree strongly/8 = agree strongly)."

Participants in the help condition read a vignette that differed only slightly. Specifically, the word "harm" was replaced by the word "help":

> 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, and it will also help the environment." The reply from the board was, "We don't care at all about helping the environment. We just want to make as much profit as we can. Let's start the new program." Sure enough, the program helped the environment.

These participants were asked the same question as participants in the harm condition. As predicted, the responses given by participants in the two conditions differed dramatically: In the harm condition, the mean response was 7.48, with 96.8% of the participants judging that the board had brought about the side-effect intentionally (i.e. giving responses of 5–8); in the help condition, in contrast, the mean response was 1.29, with 79.4% judging that the board had not brought about the side-effect intentionally (i.e. giving responses of 0–4).[4]

2.1.2. *Experiment 1b*

Knobe (2003) also asked his participants to judge the blameworthiness or praiseworthiness of the agent described in the vignette. His results indicated that people were more inclined to attribute blame for negative side-effects than praise for positive side-effects.[5] To explore whether this pattern also holds for the GKE, we conducted Experiment 1b to test the hypothesis that people are more likely to assign blame to groups for negative side-effects than they are to assign praise for positive side-effects.

We recruited 75 different naive participants (40 female, mean age = 45.53, *SD* = 15.4, range = 20–69), again using *SurveyMonkey Audience.* As in Experiment 1a, participants were randomly assigned to a harm condition or a help condition. The vignettes with which they were presented were identical to those used in Experiment 1a. However, rather than being asked whether they thought the board had acted intentionally, participants were now asked to assign blame or praise to the board. Specifically, each were presented with one of the following questions:

> On a scale from 0–8, how much blame do you think the board deserves for harming the environment? [Harm condition]

or:

On a scale from 0–8, how much praise do you think the board deserves for helping the environment? [Help condition]

As in Experiment 1a, participants' responses varied dramatically according to the condition to which they had been assigned. In the harm condition, participants gave a mean response of 7.23, with 100% of participants judging that the board deserved to be blamed (i.e. giving responses 5–8). In the help condition, in contrast, participants gave a mean response of 3.45, with 58% judging that the board did not deserve to be praised (i.e. giving responses 0–4).[6] These results clearly indicate that participants are more willing to ascribe blame to groups for bad side-effects than they are to ascribe praise to groups for actions with good side-effects.

In sum, the findings from Experiments 1a and 1b indicate that the Knobe Effect generalizes to folk attitudes towards groups.[7] However, as we noted in the introduction, there could be an alternative, deflationary explanation, according to which the GKE would not provide evidence that the folk are collectivists. In the next section, we set out this alternative, deflationary explanation, and present a series of further experiments which we conducted in order to test it.

2.2. *The non-distributive GKE*

We often speak of groups acting one way or another. Statements such as "IBM successfully conquered the Japanese market", "the Allied bombers set Dresden on fire", "the committee awarded tenure to the candidate", "none of the bystanders stepped in to help" and "we painted the house together yesterday" all refer to the actions (or omissions) by groups of people. Some of these groups are large, others are small, some stable, others ad hoc, some hierarchically organized, others egalitarian (see Kutz 2000). But, according to collectivist philosophers, at least some such statements about at least some such groups are to be interpreted non-distributively (see List and Pettit 2011). They argue that it would be wrong to take some of these statements as just economical ways of summarily and perhaps somewhat loosely referring to conjunctive lists of individual actions, e.g. "Bomber_1's bombing_1, Bomber_2's bombing_2, Bomber_3's bombing_3 … , Bomber_N's bombing_n set Dresden on fire". Rather, some such statements attribute action to the group directly. When that is the case, any attempt to replace the statement referring to collective action by conjuncts of statements referring to individual actions would not only entail a loss of meaning (and be very cumbersome) but also constitute a category mistake. The same applies to pronouncements on the moral responsibility of groups.

At this point, however, it is still to be ascertained if and when people adopt such a collectivist perspective. Distributivists do not think this is the case in Experiments 1a and 1b. They interpret the GKE as merely the aggregative result of ascriptions of intentions and praise/blame to several individuals. In our case, these would be the *members* of the board in Experiments 1a and 1b as opposed to the board *qua* group. The objection can draw additional support from existing data to the effect that people tend to use plural pronouns when thinking of groups in distributive terms (Phelan, Arico, and Nichols 2013, 711–714).[8] And indeed the board's responses are paraphrased in the vignettes used in Experiments 1a and 1b with the pronoun "we". Consequently, it is possible that when asked about the blameworthiness or praiseworthiness of a group or about that group's intentions, participants in fact covertly reasoned about the blameworthiness/praiseworthiness or the intentions of individual members of the group. If so, then our findings only demonstrate the existence of distributive GKE.

Experiments 2–4 were therefore explicitly designed to explore the possibility that our participants may have been reasoning about aggregates of individuals rather than about groups *per se*.

2.2.1. *Experiment 2a*

In order to test the alternative hypothesis that the effects observed in Experiments 1 were due to participants making judgments about aggregates of individuals rather than about groups *per se,* Experiment 2a was set up so as to reduce distributive intentionality-ascriptions by making it explicit in the vignettes that each individual member intended to prevent the implementation of the program and that it was the board that decided to implement the program.

To this end, we made two changes to the vignettes used in Experiment 1. First, we removed the direct quote from the board, which the vignettes in Experiments 1 included ("We don't care at all about harming/helping the environment … "). As noted, this quote, especially the plural pronoun "we", may encourage participants to think as if a plurality of individuals were speaking, and thus to covertly evaluate them aggregatively. Second, we added the following two sentences in both conditions:

> For various reasons, each individual member of the board personally opposed the program and tried to prevent it from being implemented. Nevertheless, the interests of the company and the shareholders prevailed and the board decided to implement the new program.

The changes to the vignettes are important because it is reasonable to expect that the use of a collective noun "the board" (instead of the first person plural noun "we") encourages people to some degree to think of the relevant group non-distributively, i.e. *qua* board and not as a plurality of individuals.[9] So we reasoned that if participants tended to make ascriptions of intentions predominantly in a distributive sense, and were for this reason reluctant to attribute intentions to groups *qua* groups, then the change to the vignettes should disincline participants to ascribe intentions to the board in the harm condition.

We recruited 102 different naive participants (51 female mean age = 44.67, *SD* = 16.8, range = 18–76), again using *SurveyMonkey Audience.* As in Experiment 1, participants were randomly assigned to a harm condition or a help condition. The vignettes with which they were presented were the same as those used in Experiment 1, but with the aforementioned changes.

As in Experiment 1, participants' responses varied dramatically according to the condition to which they had been assigned. In the harm condition, participants gave a mean response of 6.94, with 75.93% of participants judging that the board had intentionally harmed the environment (i.e. giving responses 5–8). In the help condition, in contrast, participants gave a mean response of 3.81, with only 22.92% judging that the board had acted intentionally.[10]

As in Experiment 1, participants were highly inclined to view the side-effect as intentional in the harm condition and highly disinclined to view the side-effect as non-intentional in the help condition. These results are difficult to reconcile with the hypothesis that ascriptions of intentions are understood by participants in a distributive sense.

A skeptic may however object that neither using the noun phrase "the board" nor eliminating the first person plural report attributed to the board (used in Experiment 1) guarantees that people will be thinking of the board in non-distributive terms (*qua* board). This could be a problem for us because if people are thinking of the board distributively – i.e. interpret the noun phrase "the board" as a shorthand to refer to board members – then they may be ascribing intentions distributively as well.

We respond by granting that the use of the noun phrase "the board" does not ensure that people will be considering this group in non-distributive terms. However, recall that the vignette of Experiment 2a strongly suggests that each individual member intended to prevent the implementation of the program. Furthermore, the vignette explicitly states that it was the board that decided to implement the program. It is apparent that participants would have two options here when considering how to address the situation described in the vignette. First, they could decide that if the board members did not intend the harmful result *qua* individuals, then nobody intended the harmful side-effect. If our participants had reasoned this way, then they should not have attributed an intention to bring about the side-effect to the board in either condition, and they should not have been more inclined to do so in the harm condition than in the help condition. But this is not what we observed. Rather, participants appear to have favored the second option, which was to attribute intentions to the board. In fact, three-quarters of the participants were still willing to say that the board brought about the harmful side-effect intentionally.

Observe also that the characteristic asymmetry of intention-attributions obtained here just as it does in the standard Knobe Effect. This provides further evidence for the existence of non-distributive GKE. To see why, recall that the hypothesis that GKE is distributive generates the prediction that after a manipulation of the vignette such as that in Experiment 2a, GKE should be inhibited. This prediction is not consistent with our findings: participants were much more willing to ascribe intentions for a bad side-effect than for a good side-effect. These findings therefore support the hypothesis that participants are more willing to ascribe intentions for negative side-effects than for positive side-effects, *and* do so non-distributively to groups *qua* groups.

2.2.2. *Experiment 2b*

Naturally, the findings from Experiment 2a can be challenged. It may be conceded that we did indeed reduce distributive intentionality-ascriptions in Experiment 2a by making it explicit in the vignettes that each individual member intended to prevent the implementation of the program and that it was the board that decided to implement the program. Nevertheless, some participants may still have reasoned implicitly – so the objection – that while no individual member as a person *simpliciter* – with their non-member hats on, as it were – intended the implementation of the program (because they were indeed wary of the possible adverse side-effects on the environment), at least some members *qua members* must have individually intended the implementation of the program and contributed to the decision. After all, the decision was taken by the board, which must mean that, depending on the decision procedure used, that at least one person on the board (if the board used a dictatorial decision procedure), but perhaps even all members (if consensus was required) decided to implement the program adverse environmental effects notwithstanding. In other words, while there was a clash between what they would prefer to do privately and as board members, at least some board members ultimately came to the conclusion that in their role as board members, they could not choose according to their private preferences. So they were swayed by the company's and the shareholders' interests. If so, they can be thought to have intended the harmful side-effects individually and may be individually blameworthy for them.

We therefore designed Experiment 2b to ensure that the differences observed between the harm and help conditions of Experiment 2a were not driven by covert judgments about the individual members of the board. Moreover, we adapted the vignettes to identify each individual member of the board by name, and asked participants the same questions as in

Experiment 2a about the intentions and blame- or praiseworthiness of the board *as well as* about the individual board members. We reasoned that if the GKE is distributive, then participants should not be more inclined to attribute intentions or blame- or praiseworthiness to the board than to the individual members, whereas if the GKE is non-distributive, they should be.

We recruited 411 naive participants (214 female mean age = 45.73, *SD* = 18.6, range = 18–85), again using *SurveyMonkey Audience*.[11] In a 2 × 2 between-subjects design, participants were randomly assigned to a harm/support condition, a harm/dissent condition or a help/support condition, and a help/dissent condition. The vignette in the harm/dissent condition read as follows:

> 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 it will also harm 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, 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 result, in line with 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.

In the harm/support condition, we replaced the passage "For various reasons, each of them personally opposed the program and tried to prevent it from being implemented. However … " with "Each of them personally supported the program and did not object to its being implemented. In any case … " In the two help conditions, we substituted only the word "help" for "harm," as in the previous experiments.

We performed a two-way ANOVA for responses to the question about group intentionality (i.e. the board's intentions), which revealed a main effect of the harm/help factor, with participants making significantly higher attributions of intentionality to the board in the harm condition (*M* = 7.01, *SD* = 2.65) than in the help condition (*M* = 4.61, *SD* = 2.46).[12] This replicates the pattern observed in the earlier experiments. The ANOVA also revealed that participants' judgments about the board's intentionality were significantly affected by the information they received about the individual board members' dissent (*M* = 5.61, *SD* = 2.90) or support (*M* = 6.00, *SD* = 2.74) for the board's decision.[13] There was no significant interaction between the factors harm/helm and support/dissent.[14]

We then performed a two-way ANOVA for responses to the question about group blame/praiseworthiness (i.e. the board's blame/praiseworthiness), which revealed a main effect of the harm/help factor, with participants making significantly higher attributions of blame to the board in the harm condition (*M* = 7.54, *SD* = 2.25) than praise in the help condition (*M* = 4.86, *SD* = 2.45).[15] Interestingly, however, participants' judgments about the board's blame- or praiseworthiness were not significantly affected by the information they received about the individual board members' dissent (*M* = 6.22, *SD* = 2.78) or support (*M* = 6.19, *SD* = 2.63) for the board's decision.[16] There was no significant interaction between the factors harm/helm and support/dissent.[17]

Next, we performed a two-way ANOVA for responses to the question about individual intentions (i.e. the individuals' intentions), which revealed a main effect of the harm/help factor, with participants making significantly higher attributions of intentionality to the board in the harm condition (*M* = 6.66, *SD* = 2.62) than in the help condition (*M* = 4.43, *SD* = 2.42).[18] The ANOVA also revealed that participants' judgments about the individuals' intentionality were significantly affected by the information they received about the

individual board members' dissent ($M = 5.08$, $SD = 2.86$) or support ($M = 6.01$, $SD = 2.53$) for the board's decision.[19] There was no significant interaction between the factors harm/helm and support/dissent.[20]

We then performed a two-way ANOVA for responses to the question about individual blame/praiseworthiness, which revealed a main effect of the harm/help factor, with participants making significantly higher attributions of intentionality to the board in the harm condition ($M = 6.63$, $SD = 2.53$) than in the help condition ($M = 4.49$, $SD = 2.44$).[21] The ANOVA also revealed that participants' judgments about the individuals' blame/praiseworthiness were significantly affected by the information they received about the individual board members' dissent ($M = 5.22$, $SD = 2.87$) or support ($M = 5.92$, $SD = 2.49$) for the board's decision.[22] There was no significant interaction between the factors harm/helm and support/dissent.[23]

In order to ensure that the significant differences observed between the harm and the help conditions for the questions about the board's intentions were not driven by the differences in participants' judgments about the individual members' intentions, we also performed a two-way ANCOVA with responses to the question about the members' intentions as covariate, which revealed a significant effect of the factor harm/help upon judgments about the board's intention even after controlling for judgments about the members' intentions.[24]

Next, in order to ensure that the significant differences we observed between the harm and the help conditions for the questions about the board's blame- or praiseworthiness were not driven by the differences in participants' judgments about the individual members' blame- or praiseworthiness, we also performed a two-way ANCOVA with responses to the question about the members' blame- or praiseworthiness as covariate, which revealed a significant effect of the factor harm/help upon judgments about the board's blame- or praiseworthiness even after controlling for judgments about the members' blame- or praiseworthiness.[25]

In sum, participants were more inclined to judge that the group intended the side-effect in the harm condition than in the help condition, and more inclined to blame the group in the harm condition than to praise it in the help condition – irrespective of whether the individual members supported the group's decision or dissented from it. Crucially, neither of these effects can be explained by appealing to participants' judgments about whether the individual members intended the side-effect or deserved blame or praise. These are important findings because they support the conclusion that the willingness to ascribe an intention and blame/praise to the board does not seem to be generated by way of a bottom-up process: whether participants are willing to ascribe an intention and blame/praise at the higher level of the board is relatively independent from whether they are willing to ascribe intentions and blame/praise at the bottom level of individual members. At the same time, note also that the same double asymmetry of intention-ascriptions and of blame/praise-ascriptions, which was predicted by the Knobe Effect, showed up at both levels. So what the focus on GKE helps us see is that the same evaluative processes were triggered by questions about the board and questions about individual members. Moreover, it appears that this is *not* because people thought of the board distributively as a mere aggregation of individuals. These findings therefore support the existence of non-distributive GKE.

We do not think that it constitutes a problem for this claim that the double asymmetry did show up at the individual level or that intentions and blameworthiness/praiseworthiness were attributed to the individual board members even when they opposed the board decision. It is important to see that a group's non-distributive intention to X or a group's non-distributive blameworthiness/praiseworthiness for X is not inconsistent with ascribing

individual members intentions and blameworthiness/praiseworthiness for X as well. The collectivist's point is merely that the group's non-distributive intention to X or blameworthiness/praiseworthiness for X is not an aggregation of the intentions and blameworthiness/praiseworthiness we ascribe to individual members of the group. Therefore, it does not undermine our hypothesis about non-distributive GKE that participants continued to ascribe intentions and blameworthiness/praiseworthiness to board members even when they ascribed non-distributive intentions and blameworthiness/praiseworthiness to the board as a whole. What we needed for the non-distributive hypothesis is only to show that ascriptions to the board are not driven by ascriptions to individual members. And we believe Experiments 2a and 2b support this – as do Experiments 3 and 4, to which we now turn.

2.2.3. *Experiment 3*

As a further test of the alternative hypothesis that participants' judgments in Experiments 1 and 2 were driven by covert judgments about aggregates of individuals rather than about groups *per se,* Experiments 3 was designed to give participants the opportunity to make judgments about a salient individual (i.e. "the Chairman of the board") rather than about the group (i.e. "the board"). We reasoned as follows: perhaps in Experiment 2 participants were influenced by the fact that the vignette mentioned *three* people, whereas participants might be prevented from attributing intentions to groups if they are encouraged to make judgments about individuals who are distinct from the groups in question. To this end, we added the following sentence to the vignettes used in Experiment 2:

> The decision was announced by the Chairman of the board, Donald Franklin, whose primary role is to "guide and mediate board actions with respect to organizational priorities and governance concerns."

We recruited 103 different naive participants (59 female, mean age = 47.45, *SD* = 15.5, range = 18–75), again using *SurveyMonkey Audience.* As in Experiments 1–2, participants were randomly assigned to a harm condition ($N = 58$) or a help condition ($N = 45$). The vignettes with which they were presented were the same as those used in Experiments 1 and 2 but with the aforementioned changes.

As in the previous experiments, participants' responses varied dramatically according to the condition to which they had been assigned. In the harm condition, participants gave a mean response of 7.10, with 77.59% of participants judging that the board had intentionally harmed the environment (i.e. giving responses 5–8). In the help condition, in contrast, participants gave a mean response of 4.38, with only 31.11% judging that the board had acted intentionally.[26]

The results indicate that directing the spotlight on a salient individual did not have an effect on people's willingness to attribute intentions to the group. In other words, Experiment 3 tested the possibility that people only attribute intentions perhaps because (given the lack of information) they are unsure precisely which members of the group intended the bad side-effect. While some ascriptions of intentions to groups may be of this sort (e.g. when one calls to task the entire graffiti gang for defacing the façade of the school, not knowing which member actually did it), this does not seem to be the case here. The identification of a salient individual who surely does incur some of the responsibility (if anyone does) had no effect on the willingness to ascribe intentions to the group in the harm condition.

And note once again that the characteristic asymmetry of intention-attributions obtained here just as it does in the standard Knobe Effect. People's willingness to attribute intentions to the board in the harm condition for the harmful side-effect was still as high as in Experiment 2 and not much lower than in Experiment 1. Furthermore, participants were found to be much more willing to ascribe intentions for a harmful side-effect than for a positive side-effect. The highlighting of a salient individual does not diminish GKE, which is what you would expect if GKE were distributive. Thus, Experiment 3 provides further evidence for the existence of a non-distributive GKE.

2.2.4. *Experiment 4*

We designed Experiment 4 to further minimize the likelihood that participants' judgments about the intentions of groups were driven by covert judgments about the individual members of the groups in question. To this end, we used new vignettes in which the group agent was a (fictitious) corporation.

We expected that the reference to a corporation would encourage non-distributive interpretations by our participants for the following reasons. There is considerable evidence that corporations are consistently given high entitativity ratings by people and that in general corporations are particularly likely to be perceived in non-distributive terms by ordinary folk (as well as by lawyers and philosophers), as we will see shortly.

As in Experiment 1, we aimed to probe not only the asymmetry of intention-attribution for positive and negative side-effects but also to investigate whether participants were more inclined to attribute blame for negative side-effects than praise for positive side-effects. Because the results of Experiments 1a and 1b indicated that asking participants whether a group agent had brought about a side-effect intentionally did not bias their response to a subsequent question about whether the group deserved blame or praise for the side-effect, we opted to ask all participants both questions in Experiment 4 (whether the side-effect had been brought about intentionally and whether the group deserved praise/blame).

We recruited 206 participants (104 female, mean age = 45.13, *SD* = 14.72, range = 19–75) using *SurveyMonkey*. In a between-subject design, we randomly assigned participants either to a harm condition (101) or a help condition (105), and presented them accordingly with one of the following two vignettes:

> 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 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.

These participants then responded to the following question: "On a scale from 0–8, to what extent would you agree that ACME Inc. intentionally harmed the environment? (0 = disagree strongly/8 = agree strongly)." [Harm condition]

Participants in the help condition read a vignette that differed only slightly; specifically, the word "harm" was replaced by the word "help":

> ACME Inc. started a new program. When launching the new program, data suggested that the program would help ACME Inc. increase profits, and that it would also 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 helped. [Help condition]

These participants then responded to the following question: "On a scale from 0–8, to what extent would you agree that ACME Inc. intentionally helped the environment? (0 = disagree strongly/8 = agree strongly)."

Next, participants were asked to assign blame or praise to ACME Inc. Specifically, each were presented with one of the following questions:

> On a scale from 0–8, how much blame do you think ACME deserves for harming the environment? [Harm condition]

or:

> On a scale from 0–8, how much praise do you think ACME deserves for helping the environment? [Help condition]

For the "intention question", participants' responses varied dramatically according to the condition to which they had been assigned. In the harm condition, the mean response was 6.79, with 87.2% of the participants judging that ACME had brought about the side-effect intentionally (i.e. giving responses of 5–8); in the help condition, in contrast, the mean response was 4.10, with only 43.8% judging that ACME had brought about the side-effect intentionally (i.e. giving responses of 5–8).[27]

For the "blame/praise-question", participants' responses also varied dramatically according to the condition to which they had been assigned. In the harm condition, participants gave a mean response of 7.05, with 91.1% of participants judging that ACME deserved to be blamed (i.e. giving responses 5–8). In the help condition, in contrast, participants gave a mean response of 4.11, with only 43.8% judging that ACME deserved to be praised (i.e. giving responses 5–8).[28]

We believe that Experiment 4 goes a long way towards answering potential challenges to Experiments 2 and 3. One important objection was that, despite our precautions, the use of the noun phrase "the board" may have been interpreted by some people distributively in these studies. We did argue that that was not likely to be the case, but Experiment 4 should give pause even to those who were unconvinced by our arguments concerning Experiments 2 and 3.

Experiment 4 is interesting also because we asked participants to judge the blameworthiness/praiseworthiness of the group. The results reveal that our participants were quite willing to attribute blame to the group for the harmful side-effect (and to a much lesser extent praise for the good side-effect).[29] Once again, in view of the evidence that people do tend to think of corporations in non-distributive terms, the high willingness to make blame-attributions indicates that people ascribe blame non-distributively in these cases (the same for praise when they ascribe praise).[30]

3. Discussion

We approached the question regarding people's perception of group agency in two steps. Experiment 1 has shown that the same evaluative processes were at work for group agents as for individual agents. The same double asymmetry of intention-ascriptions and of blame/praise-ascriptions, which was predicted by the Knobe Effect, showed up whether people were asked about individuals or whether they were asked about groups. Experiments 2–4 are significant because they establish that these parallels between evaluative processes are not due to the fact that people think of the group in the relevant cases as mere pluralities of individuals. So, when combined, these findings lend support to the claim

that folk psychology is at least sometimes collectivist: people perceive and evaluate the behavior of individuals and groups in similar ways when it comes to questions of intentionality and blameworthiness/praiseworthiness because they (sometimes) regard (some) groups as autonomous agents. We can also better appreciate now why focusing GKE has been a good way to test whether folk psychology is collectivist. The robustness of the double asymmetry characteristic of GKE indicates that people's judgments of individual and group agency run parallel as far as intentionality and blameworthiness/praiseworthiness are concerned.

These findings cast some doubt on alternative approaches in the literature. Tyler and Mentovich (2010) write that "people have greater trouble holding entities responsible for wrongdoing and punishing them than they do making judgments of responsibility and endorsing punitive actions for individuals" (Tyler and Mentovich 2010, 203). Our findings are difficult to reconcile with this view; they indicate that people have no trouble blaming groups for their actions, and that when they do so, they at least sometimes embrace a non-distributive interpretation of group blame. It is also worth adding that, *pace* Tyler and Mentovich, there is no indication that people had *greater* trouble holding groups responsible than individuals. Thus, it is interesting to note that in Experiment 1 the difference between the mean responses in the two conditions was even more pronounced than in Knobe's original experiments (see Knobe 2003).[31]

Future work is required to determine the extent to which GKE can be modulated by the type of group in question and the degree of entitativity of the group agent. But the studies reported above, especially Experiment 4, offer evidence that the conclusion drawn by Tyler and Mentovich (2010) is quite probably too general: people at least in certain situations are willing to ascribe blame/praise and intentionality to at least certain kinds of groups non-distributively.[32]

Note as well that the results also undermine a particular explanation that has been offered to account for people's presumed reluctance to attribute blame to collectives. This explanation is that people are reluctant to do so *because* they generally find it difficult or awkward to ascribe intentional agency to collectives – in particular to corporations (Tyler and Mentovich 2010, 203). Our results directly contradict this claim. They show that people are highly willing to attribute intentions to at least some kinds of collectives under certain circumstances, and that they are highly willing to do so embracing a non-distributive interpretation of what it is for a group to have intentions. Again, using GKE, we could provide support for the claim that in fact people judgments about group intentions and group blame/praise run parallel since we could observe the crucial asymmetry both for intention-attribution and blame/praise-attribution.

Now, if an entity (1) has the capacity to act intentionally, and (2) is an appropriate target of blame and praise, then it is safe to say that that entity comes very close to qualifying as a morally responsible agent. In short, (1) and (2) are jointly sufficient (or very close to sufficient) conditions of morally responsible agency. Therefore, if people are willing to attribute (1) and (2) to groups – and our findings seem to show that they are – then people can be said to treat at least some kinds of groups at least some of the time as morally responsible agents. And if people are willing to attribute (1) and (2) to groups in a non-distributive sense – and our findings seem to show that they are – then people can be said to be collectivist about agency and moral responsibility in the sense of collectivism defined in the introduction. That is, they are willing to treat at least some kind of groups at least some of the time as morally responsible agents, and they regard both group agency and the group's moral responsibility as non-distributive properties of the group.

4. Conclusion and outlook

In these concluding remarks, we would like to briefly summarize our main conclusions and raise a few more general points about their importance for the areas of research this paper touches upon. First, we have provided evidence that the Knobe Effect generalizes to groups. Secondly, the results provide support for the hypothesis that the GKE obtains non-distributively. Thirdly, we have also argued that this finding provides support for the claim that the folk are collectivist about the agency and moral responsibility of groups – at least in some situations and with regard to some kind of groups. That is, people appear to attribute the capacity to act directly to the group, and do not treat these ascriptions as a mere shorthand for a conjunct of ascriptions of agency to individuals. By the same token, people appear to accept that the group can be blameworthy and praiseworthy over and above the blameworthiness and praiseworthiness of individual group members.

This conclusion has two important broader implications for the debate between collectivists and individualists about the possibility of non-distributive group agency and group responsibility rehearsed at the beginning of our paper. The first implication is that individualism about agency and moral responsibility turns out to be more revisionist than is sometimes believed to be the case. That is of course neither here nor there as regards the correctness of the position, but it does put pressure on claims about its *prima facie* intuitive plausibility and psychological feasibility.

There may also be theoretical objections to taking revisionism too far. We are gesturing here towards a broadly Strawsonian conception of moral responsibility according to which certain philosophical challenges to existing practices are rendered philosophically (and not just psychologically) implausible by virtue of being too radically divorced from human needs and interests (Strawson 1974). If indeed individualism turns out to be revisionist relative to the default folk position, then it needs to be shown that it nevertheless does not violate the Strawsonian constraint (or else it needs to be shown why the Strawsonian constraint should be abandoned).

A second implication concerning the debate between individualism and collectivism is the possible need to buttress one's preferred position with an error theory. Again, if individualism is embraced on philosophical grounds, it could be important to explain why people nevertheless "go collectivist" about collective agency and moral responsibility in their everyday perceptions of groups.

Finally, we also believe that the GKE raises interesting new questions that bear upon our understanding of folk psychology. What sort of metaphysical assumptions underlie their attitudes towards groups? Is folk collectivism rooted in functionalism about agency?[33] Or is it perhaps ultimately to be traced back to semantic or pragmatic factors?[34] We think that the ultimate interest of the GKE lies in directing our attention to such questions which we hope will be taken up in future research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

John Michael wishes to acknowledge support for this research by a Starting Grant from the European Research Council (n 679092, SENSE OF COMMITMENT) and by a seed funding grant from the Interacting Minds Centre at Aarhus University (26167). András Szigeti would like to acknowledge

support for this research from the Lund Gothenburg Responsibility Project (LGRP) funded by the Swedish Research Council (Principal Investigator: Prof. Paul Russell).

Notes

1. One of the most prominent contributions to collectivism in recent years is List and Pettit (2011). For individualist criticisms of collectivism, see for example Miller and Mäkelä (2005) and Miller (2007). In what follows, we will refer to collectivism as a combination of agency-collectivism and responsibility-collectivism. It is an interesting question, but one we will not discuss here, whether it is possible to be one without the other. Even if it is, in practice, most collectivists embrace both agency-collectivism and responsibility-collectivism.
2. We wanted to use the most comprehensive and least controversial characterization possible of the Knobe Effect in this paper. Knobe's original hypothesis regarding the effect he described was that the observed asymmetry of intentionality-ascriptions is to be *explained* by the correlated asymmetry in blame/praise judgments. The definition we use throughout this paper does not take a stand on this specific hypothesis. In other words, in our characterization of the Knobe Effect, we accept that (1) whether a side-effect is considered intentional depends on whether the side-effect is negative or positive (in a very broad, quite possibly non-moral sense of negative and positive), (2) that people are more willing to assign blame for negative side-effects of actions than they are to assign praise for positive side-effect of actions. Incidentally, we did obtain data in line with Knobe's third claim, namely (3) that the asymmetry of intention-ascriptions and the asymmetry of blame/praise-ascriptions are correlated, see footnotes 7 and 24. But we take notice of criticisms of Knobe's original hypothesis and recognize that the psychological mechanisms underpinning the Knobe Effect remain a topic of considerable controversy. We also allow for the possibility that the correlation as in (3) only obtains in some cases but not in others (see Cova 2015). Therefore, we do not take a stand on the fourth claim in Knobe's original hypothesis: (4) that the asymmetry of blame/praise-ascriptions as in (2) *explains* the asymmetry of intention-ascriptions as in (1). See footnotes 5 and 25 for additional discussion. We thank an anonymous reviewer for urging us to clarify this point.
3. We thank an anonymous reviewer for pressing us to say more about the relevance of GKE to the collectivism debate in general and for asking us to consider the above hypothetical situation in particular.
4. An independent-samples *t*-test revealed a highly significant difference in the responses between the harm ($M = 7.48$, $SD = 2.13$) and help ($M = 1.33$, $SD = 1.60$) conditions; $t\ (63) = 15.58$, $p < .0001$, $d = 3.30$.
5. As noted above, Knobe originally proposed that the asymmetry of intention attribution is driven by the praise/blame asymmetry regarding the side-effect since he found the two asymmetries to be correlated. As more experimental data was forthcoming, many were beginning to find this hypothesis too narrow. It was replaced by the idea that people's moral evaluations in general (i.e., not just those to do with blameworthiness and praiseworthiness) have an impact on whether they understand an action as intentional or not. For discussion of various hypotheses concerning the role of moral judgment in such cases, see Nadelhoffer (2004a, 2004b); Knobe and Mendlow (2004); Feltz (2007). Some recent research even suggests that the Knobe Effect can be completely "explained without appeal to normative and evaluative considerations" (Cova 2015, 136). In this paper, we refrain from speculating about the explanation of the asymmetry of intention-ascription, in general, and about the potential impact of moral evaluations on this asymmetry, specifically.
6. An independent-samples *t*-test revealed a highly significant difference in the responses between the harm ($M = 7.23$, $SD = 0.69$) and help ($M = 3.45$, $SD = 3.00$) conditions; $t\ (73) = 9.11$, $p < .0001$, $d = 2.05$.
7. For the record, we did run an additional experiment to check whether participants' responses to the intentionality-question and to the blame/praise-question were correlated. We found a very significant correlation in the case of groups as well. We nevertheless excluded this experiment from this paper for reasons explained in footnote 5.
8. While people tended to use the singular pronoun "it" when referring to groups non-distributively (Phelan, Arico, and Nichols 2013).
9. Evidence for this can be found, among others, in Phelan, Arico, and Nichols (2013). They found that when people were thinking of groups in distributive terms, they were more inclined to use

plural pronouns, whereas when they were thinking of the group non-distributively, they used the *singular* ("it"). So it can be expected that the use of singular non-phrase "the board" will make a non-distributive interpretation at least somewhat more prevalent.

10. An independent-samples *t*-test revealed a highly significant difference in the responses between the harm ($M = 6.94$, $SD = 2.67$) and help ($M = 3.81$, $SD = 2.69$) conditions; $t\ (100) = 5.88$, $p < .0001$. $d = 1.17$.
11. Five participants completed the questions about intentions but did not complete the questions about blame/praiseworthiness. We include their data for the former questions.
12. $F\ (1,410) = 92.88$, $p < .001$, $\eta_p^2 = 0.186$.
13. $F\ (1,410) = 4.08$, $p = .044$, $\eta_p^2 = 0.01$.
14. $F\ (1,410) = 0.16$, $p = .692$.
15. $F\ (1,405) = 131.00$, $p < .001$, $\eta_p^2 = 0.246$.
16. $F\ (1,405) = 0.09$, $p = .77$.
17. $F\ (1,405) = 0.12$, $p = .728$.
18. $F\ (1,410) = 87.02$, $p < .001$, $\eta_p^2 = 0.177$.
19. $F\ (1,410) = 18.28$, $p < 0.001$, $\eta_p^2 = 0.043$.
20. $F\ (1,410) = 0.53$, $p = .469$.
21. $F\ (1,405) = 79.26$, $p < 0.001$, $\eta_p^2 = 0.165$.
22. $F\ (1,405) = 10.21$, $p < 0.01$, $\eta_p^2 = 0.025$.
23. $F\ (1,405) < 0.001$, $p = .993$.
24. $F\ (1,410) = 17.78$, $p < 0.001$, $\eta_p^2 = 0.042$.
25. $F\ (1,405) = 48.67$, $p < 0.001$, $\eta_p^2 = 0.108$.
26. An independent-samples *t*-test revealed a highly significant difference in the responses between the harm ($M = 7.10$, $SD = 2.60$) and help ($M = 4.38$, $SD = 2.35$) conditions; $t\ (101) = 5.50$, $p < .0001$, $d = 1.10$.
27. An independent-samples *t*-test revealed a highly significant difference in the responses between the harm ($M = 6.79$, $SD = 1.88$) and help ($M = 4.10$, $SD = 2.21$) conditions; $t\ (204) = 9.38$, $p < .0001$, $d = 1.32$.
28. An independent-samples *t*-test revealed a highly significant difference in the responses between the harm ($M = 7.05$, $SD = 1.68$) and help ($M = 4.11$, SD = 2.01) conditions; $t\ (204) = 11.36$, $p < .0001$. Incidentally, once again we found that participants' responses to the two test questions were significantly correlated, $r\ (206) = .812$, $p < .0001$.
29. As already stated, we cannot speculate in this paper about why one can consistently observe in the two conditions (positive and negative side effects) both an asymmetry of intentionality-ascriptions and an asymmetry of blame/praise-ascriptions. Although we did find that the two asymmetries were correlated, it is important to be cautious in interpreting this correlation (see Cova 2015). On the one hand, the correlation is consistent with Knobe's original conjecture that the asymmetry of intention-ascriptions is driven by the asymmetry of blameworthiness/praiseworthiness ascriptions. On the other hand, there also exist explanations of the Knobe Effect which, while predicting a correlation, reject the idea that the asymmetry in intentionality judgments is driven by the blame/praise asymmetry (Hindriks 2014). What our data do establish, however, is that the same evaluative processes are at work for group agents as for individual agents.
30. It is worth emphasizing once again that this conclusion would not be undermined by people's willingness to also attribute intentions and blameworthiness/praiseworthiness for employees of ACME Inc. As noted in the discussion of Experiment 2b above, non-distributivist ascriptions of intentions and blameworthiness/praiseworthiness are not inconsistent with ascribing individual members intentions and blame/praise as well.
31. That said, it must be acknowledged that no direct statistical comparison is possible between Knobe's original study and ours. So without further empirical research, no conclusions can be drawn from this observation regarding folk's *relative* willingness to attribute intentionality and blame/praise to groups as opposed to individuals.
32. As such our findings provide further empirical confirmation of claims made by Sherman and Percy (2010), who argue that people are inclined to ascribe intentionality, blame and responsibility to collectives with high entitativity.
33. Functionalism is preferred by most collectivists today since it enables one to argue for collectivism without having to make unpalatable dualistic or emergentist ontological commitments. But it is debatable whether folk metaphysics of agency and of the mind is really functionalist (see Knobe and Prinz 2008).

34. See Cova (2015) for a discussion of the suggestion that the Knobe Effect is at bottom a linguistic phenomenon.

Notes on contributors

John Michael is Assistant Professor of Philosophy at the University of Warwick and Affiliated Researcher at the Department of Cognitive Science of the Central European University in Budapest. His research interests include the sense of commitment, self-control, joint action, perspective-taking and other issues at the intersection between philosophy and cognitive science. He currently holds an ERC starting grant investigating the sense of commitment in joint action.

András Szigeti is Senior Lecturer in Practical Philosophy at Linköping University and Associate Director of the Lund Gothenburg Responsibility Project (LGRP). He specializes in action theory, emotion theory, and the ethics and metaphysics of individual and collective responsibility.

References

Cova, F. 2015. "The Folk Concept of Intentional Action: Empirical Approaches." In *A Companion to Experimental Philosophy*, edited by J. Sytsma and W. Buckwalter, 121–141. Chichester: Wiley-Blackwell.

Feltz, A. 2007. "The Knobe Effect: A Brief Overview." *Journal of Mind and Behavior* 28: 265–277.

Hindriks, F. 2014. "Normativity in Action: How to Explain the Knobe Effect and its Relatives." *Mind and Language* 29: 51–72.

Knobe, J. 2003. "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63: 190–194.

Knobe, J. 2006. "The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology." *Philosophical Studies* 130: 203–231.

Knobe, J., and A. Burra. 2006. "The Folk Concepts of Intention and Intentional Action: A Cross-Cultural Study." *Journal of Cognition and Culture* 6: 113–132.

Knobe, J., and G. S. Mendlow. 2004. "The Good, the Bad and the Blameworthy: Understanding the Role of Evaluative Reasoning in Folk Psychology." *Journal of Theoretical and Philosophical Psychology* 24: 252–258.

Knobe, J., and J. Prinz. 2008. "Intuitions About Consciousness: Experimental Studies." *Phenomenology and the Cognitive Sciences* 7: 67–83.

Kutz, C. 2000. *Complicity: Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press.

Leslie, A. M., J. Knobe, and A. Cohen. 2006. "Acting Intentionally and the Side-Effect Effect: Theory of Mind and Moral Judgment." *Psychological Science* 17: 421–427.

List, C., and P. Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.

Malle, B. F. 2010. "The Social and Moral Cognition of Group Agents." *Journal of Law and Policy* 19: 95–136.

Miller, S. 2007. "Against the Collective Moral Autonomy Thesis." *Journal of Social Philosophy* 38: 389–409.

Miller, S., and P. Mäkelä. 2005. "The Collectivist Approach to Collective Moral Responsibility." *Metaphilosophy* 36: 634–651.

Nadelhoffer, T. 2004a. "On Praise, Side-Effects, and Folk Ascriptions of Intentionality." *Journal of Theoretical and Philosophical Psychology* 24: 196–213.

Nadelhoffer, T. 2004b. "Blame, Badness, and Intentional Action: A Reply to Knobe and Mendlow." *Journal of Theoretical and Philosophical Psychology* 24: 259–269.

Nadelhoffer, T. 2005. "Skill, Luck, Control, and Intentional Action." *Philosophical Psychology* 18: 341–352.

Nadelhoffer, T. 2006. "Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality." *Philosophical Explorations* 9: 203–219.

Nichols, S., and J. Ulatowski. 2007. "Intuitions and Individual Differences: The Knobe Effect Revisited." *Mind and Language* 22: 346–365.

Phelan, M., A. Arico, and S. Nichols. 2013. "Thinking Things and Feeling Things: On an Alleged Discontinuity in Folk Metaphysics of Mind." *Phenomenology and the Cognitive Sciences* 12: 703–725.

Sherman, S. J., and E. J. Percy. 2010. "The Psychology of Collective Responsibility: When and Why Collective Entities are Likely to be Held Responsible for the Misdeeds of Individual Members." *Journal of Law and Policy* 19: 137–170.

Strawson, P. 1974. "Freedom and Resentment." In *Freedom and Resentment*, 28–21. London: Methuen.

Tyler, T. R., and A. Mentovich. 2010. "Punishing Collective Entities." *Journal of Law and Policy* 19: 203–230.