Journal of Law and Policy

Volume 19 | Issue 1 Article 6

2010

The Social and Moral Cognition of Group Agents

Bertram F. Malie

THE SOCIAL AND MORAL COGNITION OF GROUP AGENTS

*Bertram F. Malle**

INTRODUCTION

To better understand the possibility, scope, and limits of punishment for groups we must understand how humans conceptualize group agents, interpret their actions, and make moral judgments about them. In this article I therefore examine the social-cognitive foundations for human perceptions of groups and the moral evaluations of their conduct. Part I identifies the conceptual framework within which people perceive, interpret, and reason about individual agents. Part II examines whether people apply the same framework to cognitions of group agents. Part III introduces the psychological system that accomplishes people's moral judgments of individual agents. Part IV explores whether people equally apply this system to moral judgments of group agents. Finally, Part V discusses the limits of perceiving groups as moral agents—limits that the perceiver may feel more painfully than the agent.

I. SOCIAL COGNITION: THE CASE OF INDIVIDUALS

The ultimate goal of this article is to explore how groups can be targets of moral cognition. Moral cognition, however, is deeply embedded in social cognition, so my first concern is to identify

* Professor, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University. M.A. (Philosophy and Linguistics), M.Sc. (Psychology), University of Graz, Austria; Ph.D. (Psychology), Stanford University. I would like to thank the organizers of the Trager Symposium for inspiring this article and the editorial team at the Brooklyn *Journal of Law and Policy* for sharpening its delivery.

core constituents of social cognition—of both individuals and groups—that provide a foundation for moral cognition.[1] My focus will be on three such constituents: judgments of intentionality, behavior explanations, and ascriptions of mental states. These phenomena will reveal a fascinating complexity of social cognition that translates into a similar complexity of moral cognition. The key question will be whether this complexity extends from individual to group agents.

A. Fundamental Concepts: Agency and Intentionality

Humans perceive people, and interactions among them, through a framework that conceptualizes behavior as fundamentally linked with mental states. This framework, variously called *common-sense psychology*, *folk psychology*, or *theory of mind* consists of two parts: (a) systems that filter, organize, and integrate certain stimulus inputs into such concepts or categories as *agent*, *intention*, *belief*, and *reason*;[2] and (b)

---

[1] I should emphasize that social cognition encompasses a broader set of concepts, capacities, and activities than I will investigate here. For reviews see Bertram F. Malle, *Folk Theory of Mind: Conceptual Foundations of Human Social Cognition*, *in* THE NEW UNCONSCIOUS 225 (Ran Hassin et al. eds., 2005) [hereinafter Malle *in* THE NEW UNCONSCIOUS]; Bertram F. Malle, *The Fundamental Tools, and Possibly Universals, of Social Cognition*, *in* HANDBOOK OF MOTIVATION AND COGNITION ACROSS CULTURES 267 (Richard Sorrentino & Susumu Yamaguchi eds., 2008) [hereinafter Malle *in* HANDBOOK OF MOTIVATION AND COGNITION]. Among the capacities and processes I will set aside we find face recognition, gaze following, mimicry, automatic empathy, joint attention, imitation, simulation, and perspective taking. I will also say little about the vast literature on stereotyping because its focus is on the perception of individuals *as group members*, not the perception *of groups*. The actual social-psychological literature on group perception will be well covered by Steve Sherman and Elise Percy's article, *The Psychology of Collective Responsibility: When and Why Collective Entities are Likely to be Held Responsible for the Misdeeds of Individual Members*, 18 J.L. & POL'Y forthcoming (2010).

[2] *See generally* Roy D'Andrade, *A Folk Model of the Mind*, *in* CULTURAL MODELS IN LANGUAGE AND THOUGHT 112 (Dorothy Holland & Naomi Quinn eds., 1987); Yoshihisa Kashima, Allison McIntyre & Paul Clifford, *The Category of the Mind: Folk Psychology of Belief, Desire, and Intention*, 1 ASIAN J. SOC. PSYCHOL. 289 (1989); Alan M. Leslie, *A Theory of Agency*, *in* CAUSAL COGNITION: A MULTIDISCIPLINARY DEBATE 121 (Dan Sperber, et al. eds.,

assumptions about these categories and their relationships.[3]

This conceptual framework is distinct from the variety of psychological processes that solve social-cognitive tasks—such as action parsing, gaze following, simulation, or inference. These processes have distinct evolutionary and developmental paths, and they operate within the conceptual framework.[4] Gaze following, for example, exists in other primates, but its connection to inferences of mental states such as *seeing*, appears to emerge only in humans.[5] The entire package of conceptual framework plus the suite of psychological processes that operate within it make up the larger phenomenon of *human social cognition*.[6]

The central categories of the folk theory of mind are arguably *agent*, *intentionality*, and *mind*, and they are closely related to one another.[7] Agents are entities that can act intentionally; intentional actions require a particular involvement of the mind; and only agents have minds. This web of concepts is anchored in specific perceptual-cognitive processes emerging in the first year of life. Infants quickly identify *agents* from a few central cues, including faces, self-propelled motion, and contingent interaction with the perceiver.[8] Having identified an agent, human perceivers are

---

1995); Bertram F. Malle & Joshua Knobe, *The Folk Concept of Intentionality*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 101 (1997) [hereinafter *Folk Concept of Intentionality*]; BERTRAM F. MALLE, HOW THE MIND EXPLAINS BEHAVIOR: FOLK EXPLANATIONS, MEANING, AND SOCIAL INTERACTION (MIT Press 2004) [hereinafter HOW THE MIND EXPLAINS BEHAVIOR].

[3] *See generally* JERRY A. FODOR, MODULARITY OF THE MIND (1983); Malle *in* THE NEW UNCONSCIOUS, *supra* note 1; Adam Morton, *Folk Psychology is not a Predictive Device*, 105 MIND 119 (1996); Henry M. Wellman & Jacqueline D. Woolley, *From Simple Desires to Ordinary Beliefs: The Early Development of Everyday Psychology*, 35 COGNITION 245 (1990).

[4] HOW THE MIND EXPLAINS BEHAVIOR, *supra* note 2, at ch. 2.

[5] *See* Daniel J. Povinelli & Timothy J. Eddy, *What Young Chimpanzees Know About Seeing*, 61 MONOGRAPHS SOC'Y FOR RES. CHILD DEV. 122 (1996).

[6] For an introduction to this entire package, see generally Malle *in* HANDBOOK OF MOTIVATION AND COGNITION, *supra* note 1.

[7] *See generally* JAN SMEDSLUND, THE STRUCTURE OF PSYCHOLOGICAL COMMON SENSE (1997).

[8] Susan C. Johnson, *The Recognition of Mentalistic Agents in Infancy*, 4 TRENDS COGNITIVE SCI. 22, 25 (2000); *see generally* David Premack, *The Infant's Theory of Self-Propelled Objects*, 36 COGNITION 1 (1990).

sensitive to facial expression, gaze, and motion patterns that reveal the agent's specific intention underlying the observed behavior.[9]

Many linguists count the concepts of *agent* and *intentionality* as fundamental to the way humans see and talk about the world, and, indeed, linguistic forms of these concepts have been found across all known languages.[10] Developmental research, too, finds the concept of *intentional agent* across all studied cultures.[11]

The concept of intentionality guides social cognition from infancy on and, through development, reaches a remarkable complexity. In adults, the concept encompasses five components that refer to distinct mental states and capacities.[12] For children, acquiring such a complex concept is not an easy feat and is not done in one trial. Instead, the child builds up the concept from simple beginnings over many years of conceptual and social development.[13] During the first year of life, infants identify intentional behavior by paying close attention to self-propelled movement and especially object-directed movement such as

---

[9] *See generally* Jodie A. Baird & Dare A. Baldwin, *Making Sense of Human Behavior: Action Parsing and Intentional Inference*, *in* INTENTIONS AND INTENTIONALITY: FOUNDATIONS OF SOCIAL COGNITION 193 (Bertram F. Malle, et al. eds., 2001) [hereinafter INTENTIONS AND INTENTIONALITY]; Winand J. Dittrich & S.E.G. Lea, *Visual Perception of Intentional Motion*, 23 PERCEPTION 253 (1994); Ann T. Phillips, Henry M. Wellman & Elizabeth S. Spelke, *Infants' Ability to Connect Gaze and Emotional Expression to Intentional Action*, 85 COGNITION 53 (2002); Jeffrey M. Zacks, Shawn Kumar, Richard A. Abrams & Ritesh Mehta, *Using Movement and Intentions to Understand Human Activity*, 112 COGNITION 201 (2009).

[10] *See generally* JOAN BYBEE ET AL., THE EVOLUTION OF GRAMMAR: TENSE, ASPECT, AND MODALITY IN THE LANGUAGES OF THE WORLD (1994); Talmy Givón, *Cause and Control: On the Semantics of Interpersonal Manipulation*, 4 SYNTAX & SEMANTICS 59 (1975); ANNA WIERZBICKA, SEMANTICS: PRIMES AND UNIVERSALS (1996).

[11] *See generally* Henry M. Wellman & Joan G. Miller, *Developing Conceptions of Responsive Intentional Agents*, 6 J. COGNITION & CULTURE 27 (2006).

[12] The five components are desire, belief, intention, awareness, and skill. *See infra* Part I.B.

[13] *See generally* Alison Gopnik & Andrew N. Meltzoff, *The Child's Theory of Action*, *in* WORDS, THOUGHTS, AND THEORIES 125 (1997); INTENTIONS AND INTENTIONALITY, *supra* note 9.

grasping or putting.[14] At the end of the first year, they are able to parse streams of behavior, at just the right junctures, into units that correspond to initiated or completed intentional actions,[15] properly taking advantage of eye-gaze and verbal markers (e.g., "oops").[16] Through the second year, they refine their understanding of an agent's "object-directedness"[17] into the first truly mental concept of *desire*. Children recognize that another person can have desires different from their own,[18] and they infer an agent's goals even from incomplete action attempts.[19] Over the next years, children acquire the concepts of *belief* and *intention*—the latter of which is cleanly differentiated from desire only after the age of five.[20]

Developmental change in the folk-psychological framework thus occurs primarily as a differentiation of the intentionality concept, from a simple behavioral understanding to an increasingly rich mentalistic understanding. I now take a closer look at the endpoint of this development—the fully-fledged concept of

---

[14] *See* Amanda L. Woodward, *Infants Selectively Encode the Goal Object of an Actor's Reach*, 69 COGNITION 1 (1998).

[15] *See* Dare A. Baldwin, Jodie A. Baird, Megan M. Saylor & M. Angela Clark, *Infants Parse Dynamic Action*, *72* CHILD DEV. 708 (2001); Megan M. Saylor et al., *Infants' On-line Segmentation of Dynamic Human Action*, 8 J. COGNITION & DEV. 113 (2007).

[16] *See* Malinda Carpenter, Nameera Akhtar & Michael Tomasello, *Fourteen- Through 18-Month-Old Infants Differentially Imitate Intentional and Accidental Actions*, 21 INFANT BEHAV. & DEV. 315 (1998).

[17] *See* Henry M. Wellman & Ann T. Phillips, *Developing Intentional Understandings*, *in* INTENTIONS AND INTENTIONALITY, *supra* note 9, at 139–40.

[18] *See* Betty M. Repacholi & Alison Gopnik, *Early Reasoning About Desires: Evidence From 14 and 18-Month-Olds*, 33 DEV. PSYCHOL. 12 (1997).

[19] Andrew N. Meltzoff, *Understanding the Intentions of Others: Re-enactment of Intended Acts by 18-Month-Old Children*, 31 DEV. PSYCHOL. 838, 842 (1995).

[20] *See* Jodie A. Baird & Louis J. Moses, *Do Preschoolers Appreciate That Identical Actions May be Motivated by Different Intentions?*, 2 J. COGNITION & DEV. 413 (2001). For a review of the developmental path, see generally Janet W. Astington, *The Paradox of Intention: Assessing Children's Metarepresentational Understanding*, *in* INTENTIONS AND INTENTIONALITY, *supra* note 9. For the distinction between desire and intention in adults, see Bertram F. Malle & Joshua Knobe, *The Distinction Between Desire and Intention: A Folk-Conceptual Analysis*, *in* INTENTIONS AND INTENTIONALITY, *supra* note 9, at 45.

intentionality in adults.

B. Intentionality: Concept and Judgments

Over the centuries, philosophers have offered countless analyses of the concept of intentionality. But whose concept is it? How can we decide, for instance, whether *intention* is truly different from *desire*? Malle and Knobe investigated empirically what ordinary people's concept of intentionality is.[21] After all, people use their own folk concept—not any philosopher's idealized model—to solve everyday interaction tasks and make moral judgments.[22] In a first study, participants read descriptions of twenty behaviors and rated them for their intentionality. About one half of the participants received no definition of intentionality before they made their ratings; the other half did receive such a definition ("it means that the person had a reason to do what she did and that she chose to do so"). Agreement of intentionality ratings across the twenty behaviors was high: on average, any two people's ratings correlated at $r = .64$, and any one person's ratings correlated at $r = .80$ with the remaining group. More important, the experimenter-provided definition had no effect on agreement. It appears that people share a folk concept of intentionality and spontaneously use it to judge behaviors.

Given that there is a shared folk concept of intentionality, what components does this concept have? Under what conditions do people consider a behavior intentional? As a first pass, Malle and Knobe asked people for explicit definitions of intentionality ("When you say that somebody performed an action intentionally, what does this mean?"). These definitions showed consensus, revealing four main components: for an agent to perform a behavior intentionally, the agent must have (a) a desire for an outcome; (b) a belief that the behavior will lead to that outcome; (c) an intention to perform the behavior; and (d) awareness of

---

[21] *See generally Folk Concept of Intentionality*, *supra* note 2.

[22] *See generally* Bertram F. Malle & Sarah E. Nelson, *Judging Mens Rea: The Tension Between Folk Concepts and Legal Concepts of Intentionality*, 21 BEHAV. SCI. & LAW 563 (2003); Bertram F. Malle, *Intentionality, Morality and Their Relationship in Human Judgment*, 6 J. COGNITION & CULTURE 61 (2006) [hereinafter *Judgments of Intentionality and Morality*].

fulfilling the intention while performing the behavior.

Malle and Knobe also postulated a fifth component of intentionality: skill—or the ability to control and replicate the behavior in question (rather than being lucky in somehow managing to perform it). They presented participants with stories about behaviors and experimentally varied whether evidence for certain components was present (among them the agent's skill). They found that for difficult actions, people indeed look for evidence of the agent's skill at controlling the behavior.[23]

Malle and Knobe thus proposed a five-component model of the folk concept of intentionality, displayed in Figure 1. According to this folk concept, the direct cause of an intentional action is the mental state of intention. For an intention to be ascribed, at a minimum a desire (for an outcome) and a belief (about the action-outcome link) must be present. For an action to be seen as *performed* intentionally, however, skill and awareness have to be present as well. Thus, people distinguish between intention as a mental state and intentionality as a property of an action. This two-layer structure was supported in an additional experiment in which belief and desire information was found to be necessary for intention ascriptions and, given an intention ascription, skill and awareness were found to be necessary for an intentionality ascription.[24] The five constituents of the folk concept of intentionality—belief, desire, intention, skill, and awareness—can be reliably found across many different languages and may be universal conceptual primitives of the mind.[25]

---

[23] This component was likely omitted from people's explicit definitions because people focused on social behaviors, for which skill is almost always assumed, rather than artistic, athletic or otherwise difficult behaviors, for which skill may not be as readily assumed.

[24] *Folk Concept of Intentionality*, *supra* note 2, at 109–11 (Study 3).

[25] Wierzbicka, *supra* note 10, at 36–38. Wierzbicka's list of such conceptual primitives does not distinguish intentions from desires, but other linguistic work provides evidence for the unique role of intentions. Bybee has shown that, across countless and diverse languages, *intending* and the future tense are tightly connected, something that we don't see for desire or goal concepts. *See generally* Bybee, *supra* note 10.

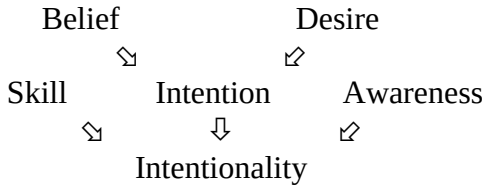

Figure 1. A model of the folk concept of intentionality.[26]

1. Two Forms of Intentionality Judgments

Even though the concept of intentionality consists of five components and people are sensitive to the presence or absence of each of these components, we should not expect people to deliberate about these five components each time they judge a behavior as intentional. They are likely to consider carefully each of the components if uncertainty or the weight of the judgment demands it, but we can expect people to use a more efficient path to assess intentionality in everyday situations. This more efficient path is configural, fast, and often unconscious; it develops earlier, and evolved longer ago. The former, more deliberate path searches for and weighs information; it is slower, often conscious, develops later, and evolved more recently.[27] Evidence for these two paths is largely indirect, but it is convincing nonetheless.

Many animals can quickly recognize predator actions. These naturally intentional actions *look* a certain way, and the prey makes a configural judgment about such an attacking action.[28] The animal

---

[26] Figure 1 adapted from *The Folk Concept of Intentionality*, *supra* note 2, © Lawrence Erlbaum Assoc., Inc., with permission.

[27] These two paths may constitute two distinct processes (and neural structures) or one process that is variably modulated by other processes (e.g., by higher-order executive control). I have no commitment to one or the other version, and there is no current evidence that could decide between the two.

[28] Among goslings, a silhouette that moves in one direction in the sky, resembling a hawk, is recognized as a threat whereas the same silhouette moving in the opposite direction, resembling a duck, is ignored. This is a configural judgment of the agent's identity but is nonetheless a powerful illustration of fast, configural judgments. *See generally* Nora Canty & James L. Gould, *The Hawk/Goose Experiment: Sources of Variability*, 50 ANIMAL BEHAV. 1091 (1995).

is unlikely to make a literal distinction between intentional and unintentional behavior, but a structure exists within the animal that responds to certain configurations of movement that *are* intentional actions and does not respond to others. We also know that apes distinguish between behavior patterns that reveal something about the agent's intentions and behavior patterns that are apparently accidental.[29] Further, the detection of mirror neurons in monkeys suggests that some of this capacity to recognize intentional action comes from the immediate interconnection between the perception of another's action and the perceiver's own motor program for this action.[30] Such mirror neurons fire both when the monkey sees someone crack a nut and when the monkey cracks a nut itself.[31] Thus, if the animal perceives a certain intentional behavior, then its own triggered motor program will be intentional as well, which provides a "tag" for the perceived behavior to be intentional. On this account, perceivers can easily recognize intentional actions that they themselves are capable of performing but not actions that they themselves are not yet capable of performing. Research on 3- to 6-month-old infants indeed shows this to be the case.[32] Languages also code for intentionality, in lexicon or morphology, and listeners can decode this feature in a split second. For example, we immediately know the difference between a person saying "I slept in" or "I overslept." Finally, recent work in our lab has shown that intentionality judgments of basic human actions are made very fast and are largely independent of other judgments (such as of goals, beliefs, and personality).[33]

---

[29] *See* Josep Call & Michael Tomasello, *Distinguishing Intentional From Accidental Actions in Orangutans (Pongo Pygmaeus), Chimpanzees (Pan Troglodytes), and Human Children (Homo Sapiens)*, 112 J. COMP. PSYCHOL. 192 (1998).

[30] *See generally* Giacomo Rizzolatti & Laila Craighero, *The Mirror-Neuron System*, 27 ANN. REV. NEUROSCI. 169 (2004).

[31] This is true for other sensory modalities as well, such as hearing a nut being cracked. *See* Evelyne Kohler et al., *Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons*, 297 SCI. 846 (2002).

[32] *See* Jessica A. Sommerville et al., *Action Experience Alters 3-Month-Old Infants' Perception of Others' Actions*, 96 COGNITION B1 (2005).

[33] Jess Holbrook, The Time Course of Social Perception: Inferences of Intentionality, Goals, Beliefs, and Traits From Behavior (2006) (unpublished Ph.D. dissertation, University of Oregon) (on file with author). Bertram F.

Evidence for the more deliberate path comes in part from adults' debates over ambiguous actions, such as in the jury box, but also in daily life, such as when a person ponders whether a colleague who didn't invite him to the conference did so intentionally or just failed to think of it. People seek evidence for specific components if they are ambiguous or missing,[34] which implies a refined conceptual network and probably requires language as well as school-age maturity. This we don't find in other primates, who can identify intentional action but, according to extant evidence, do not have mentalistic concepts of desire, belief, and intention.[35] This is consistent with the general assumption that chimpanzees are similar to 2-year-olds, who are at the transition from a sophisticated behavioral to a mentalistic understanding of intentional action.

Judgments of intentionality play a central role in many social cognitive activities. The following section reviews one of the most important ones: how people explain behavior. In particular, we will see how people ascribe the mental states of beliefs and desires to an agent in order to make sense of her behavior. The unique way in which beliefs and desires are seen as the agent's *reasons* constitute a hallmark of the folk theory of mind and behavior. Consequently, this hallmark will become a test case (in Part II of this article) for how similar the social cognition of groups is to the social cognition of individuals.

Malle & Jess Holbrook, Is There a Hierarchy of Social Inference? Evidence From a New Experimental Paradigm (June 2008) (unpublished paper presented at the Society of Philosophy & Psychology 34th Annual Meeting, Philadelphia, PA) (on file with author).

[34] Angela Laurita, The Concept of Intentionality Underlying People's Judgments of Criminal Behavior (2006) (unpublished Honor's thesis, University of Oregon) (on file with author). For a summary, see *Judgments of Intentionality and Morality*, *supra* note 22, at 71–72.

[35] *See generally* Derek C. Penn & Daniel J. Povinelli, *On the Lack of Evidence That Chimpanzees Possess Anything Remotely Resembling a 'Theory of Mind,'* 362 PHIL. TRANS. OF ROYAL SOC. BAR 731 (2007). For a more optimistic view, see generally Josep Call & Michael Tomasello, *Does the Chimpanzee Have a Theory of Mind? 30 Years Later*, 12 TRENDS IN COGNITIVE SCI. 187 (2008).

C. Core Application of Intentionality: Behavior Explanations

The complex folk concept of intentional action is most clearly revealed when we examine how people explain such actions. As Heider noted, people's explanations look quite different for unintentional and intentional behaviors.[36] They explain unintentional behaviors by referring to "mechanical" causal factors (e.g., emotions, traits, others' behaviors, physical events), and we may label these *cause explanations*. In such explanations, people presuppose nothing but a straightforward cause-effect relation; there is no role for notions of intention or awareness. For example, "I almost failed my exams 'cause I didn't really prepare for them" (*not preparing* caused *failing*) or "A friend cried on the phone because she felt unloved" (*feeling unloved* caused *crying*).

In contrast, explanations of intentional behavior are far more complex, involving assumptions of awareness, rationality, and intentional control. In fact, people use three distinct "modes" of explaining intentional behavior, and two are important for our purposes here: *reason explanations* and *causal history of reason explanations*.[37] I now describe them in sufficient detail to explore their role in explanations of group behaviors in Part II of this article.[38]

---

[36] FRITZ HEIDER, THE PSYCHOLOGY OF INTERPERSONAL RELATIONS 100 (Wiley 1958). Heider's work has often been misinterpreted in the very literature that followed his lead. *See* Bertram F. Malle & William Ickes, *Fritz Heider: Philosopher and Psychologist*, *in* 4 PORTRAITS OF PIONEERS IN PSYCHOLOGY 193 (Gregory A. Kimble & Michael Wertheimer eds., 2000); *See also generally* Bertram F. Malle, *Fritz Heider's Legacy: Celebrated Insights, Many of Them Misunderstood*, 39 SOC. PSYCHOL. 163 (2008).

[37] The third, and relatively rare, mode of explaining intentional action refers to factors that enabled the action to come about as it was intended. *See* Bertram F. Malle et al., *Conceptual Structure and Social Functions of Behavior Explanations: Beyond Person–Situation Attributions*, 79 J. PERSONALITY & SOC. PSYCHOL. 309 (2000) [hereinafter *Conceptual Structure*]; John McClure & Denis Hilton, *For You Can't Always Get What You Want: When Preconditions Are Better Explanations Than Goals*, 36 BRIT. J. SOC. PSYCHOL. 223 (1997).

[38] For more detailed analyses, see Bertram F. Malle, *How People Explain Behavior: A New Theoretical Framework*, 3 PERSONALITY & SOC. PSYCHOL. REV. 21 (1999) [hereinafter *How People Explain Behavior*]; see also *Conceptual Structure*, *supra* note 37, at 310–15; HOW THE MIND EXPLAINS BEHAVIOR,

1. Reason Explanations

Reason explanations are the most frequently used mode, and they reflect the core of the intentionality concept—the reasoning process that leads from belief and desire to an intention. This process, according to people's folk theory, occurs when an agent decides to act *in light of* and *on the grounds of* those beliefs and desires, which makes those beliefs and desires the *reasons for which* she acted.

An agent decides to act *in light of* certain beliefs or desires if the agent consciously considered them when deciding to act. This "subjectivity" assumption[39] is the first defining feature of reason explanations: they are designed to capture the agent's subjective viewpoint, to reconstruct the agent's actual beliefs and desires that shaped her intention.

An agent decides to act *on the grounds of* certain beliefs or desires if the agent saw them as reasonable grounds for deciding to so act. This "rationality" assumption[40] is the second defining feature of reason explanations: they have to hang together so as to offer justification for the reasonableness and comprehensibility of the intention or action. For example, when an explainer claims, "Anne invited Ben to dinner because he had fixed her car," then the explainer must presume that Anne actually considered Ben's fixing her car when deciding to invite him and saw his fixing her car as reasonable grounds for inviting him.

2. Causal History of Reason (CHR) Explanations

Even though people explain most intentional behaviors by reference to the agent's reasons, they explain some of them by pointing to factors that lay in the causal history of those reasons but were not themselves reasons. These *causal history of reason* (CHR) explanations can cite the agent's unconscious mental states,

---

supra note 2, at ch. 4–5.

[39] *How People Explain Behavior*, *supra* note 38, at 36; HOW THE MIND EXPLAINS BEHAVIOR, *supra* note 2, at 92–93.

[40] *How People Explain Behavior*, *supra* note 38, at 36–37; HOW THE MIND EXPLAINS BEHAVIOR, *supra* note 2, at 93.

personality, upbringing, culture, and the immediate context.[41] Whereas reason explanations try to capture what the agent herself considered and weighed when deciding to act, CHR explanations take a step back and try to capture what led up to the agent's reasons in the first place. For example, when clarifying why Kim didn't vote, an explainer might say "She is lazy" or "Her whole family is apolitical." Both statements help explain Kim's action, but they do not pick out Kim's subjective reasons for not voting. Causal history of reason explanations explain an intentional action by citing causal antecedents to the agent's reasoning and her decision to act, but there is no assumption that the agent actively or rationally considered those antecedents in her reasoning process. Hence, when an explainer states that "Kim didn't vote because she is lazy," he does not imply that Kim reasoned: "I am lazy; therefore I shouldn't vote."

In Part II of this article I examine the social cognition of groups with a particular focus on the concept of intentional agency, the explanation of action by reasons, and the types of mental states that are (or are not) attributed to group agents.

II. SOCIAL COGNITION OF GROUP AGENTS

> *The Ventura County Community College District, the Ventura County Fire Protection District, and the Ventura County Sheriff's Department* ***intend*** *to pool their resources*
> (Los Angeles Times)

> *Giant has* ***offered*** *a buyout to its highest-paid workers in* ***an effort to*** *save on labor costs*
> (Washington Post)

> *Wimbledon simply did not* ***know*** *what had hit them and if this was Arsenal's* ***reminder*** *to Manchester United that they* ***intend*** *to fight to retain their title to the last*
> (Irish Times)

---

[41] *How People Explain Behavior*, *supra* note 38, at 32-35; HOW THE MIND EXPLAINS BEHAVIOR, *supra* note 2, at ch. 4, especially 102–09; *Conceptual Structure*, *supra* note 37, at 311–15.

> *Sea Bright officials initially opposed the sand-pumping program because they* ***thought*** *their town would be inundated with out-of-towners*
> (Washington Post)

What should we make of such statements? At least at the linguistic surface they are ascriptions of intentional actions, intentions, knowledge, beliefs, and desires to group agents. Such ascriptions can be found with ease in speech and print.[42]

But are they not just metaphors? A sharp critic of "collectivism" writes, "The fact that we attribute intentional qualities to groups does not imply that those groups have real intentions. The intention we attribute to groups is metaphorical."[43] This criticism, however, is flawed in two respects. First, it makes the empirical claim that "we" (people, I presume) attribute mental states to groups metaphorically, but no empirical evidence is offered for this claim.[44] Second, the author apparently distinguishes between the practice of (metaphorically) *ascribing* intentions (or other mental states) to groups and some objective way in which groups don't literally *have* intentions. But what constitutes this objective reality? If most humans see no problem in ascribing intentions to groups, what is the scholar to say? "You are all wrong!" Or perhaps "You don't know what the concept of

---

[42] Austen Clark, *Beliefs and Desires Incorporated*, 91 J. PHIL. 404, 404 (1994); Paul Bloom & Csaba Veres, *The Perceived Intentionality of Groups*, 71 COGNITION B1, B2 (1999); Bryce Huebner, Michael Bruno & Hagop Sarkissian, *What Does the Nation of China Think About Phenomenal States?*, 1 REV. PHIL. PSYCH. 225, 226 (2010).

[43] Manuel Velasquez, *Debunking Corporate Moral Responsibility*, 13 BUS. ETHICS Q. 531, 545–46 (2003).

[44] Adam Arico, Brian Fiala, Robert Goldberg & Shaun Nichols, Folk Psychology of Consciousness (unpublished manuscript under review) (manuscript at n.9) *available at* http://www.u.arizona.edu/~arico/FPC.pdf. Arico and colleagues report evidence from pilot studies that people do not mean to be "metaphorical" in their ascriptions of mind attribution. The researchers found that ordinary people take statements such as "Some corporations want lower taxes" to be *literally* true. On a scale from 1 (figuratively true) to 7 (literally true), their ratings averaged 6.1. E-mail from Adam Arico, Dep't of Phil., Univ. of Arizona, to Bertram F. Malle, Professor of Psychol., Brown Univ. (June 19, 2010) (on file with author).

intention means!" In actuality, it is the scholar who is confused about the concept of intention. Given that there is a folk concept of intention, scholarly work must determine when and why people ascribe mental states to groups *using that concept*. If a scholar would like to adopt a different concept of intention, based perhaps on an alleged neural substrate, he or she would have to argue for it but should better call it something else (e.g., *neurointention*). People would continue to ascribe *intentions* to groups, and if they learned what the scholar means by *neurointention*, they may refrain from ascribing *neurointentions* to groups.

I therefore continue to assume that when people ascribe a mental state to a group they *literally* ascribe that state to the group. The question is just what this literal meaning of mental states amounts to.

As with most folk concepts of the mind,[45] people are, in philosophers' language, *functionalists*. They ascribe mental states not by looking for a physical substrate but by integrating perceptual cues (e.g., eyes, contingent behavior), categorical assumptions (e.g., agents can act intentionally and have minds), and context information. People envision "minds" in most animals, gods, aliens, and computers if at least some of those conditions are met: cues (e.g., biological motion triggering the expectation of mental states); assumptions (e.g., that gods and aliens must be like humans, just more so); and context (e.g., companies make decisions and so are likely to deliberate). Because of the flexibility of their concepts of agency, intentionality, and mentality, people certainly have no difficulty ascribing minds to groups.[46] However,

---

[45] *See generally Folk Concept of Intentionality*, *supra* note 2; Andrew E. Monroe & Bertram F. Malle, *From Uncaused Will to Conscious Choice: The Need to Study, Not Speculate About, People's Folk Concept of Free Will*, 1 REV. PHIL. & PSYCHOL. 211 (2010); D'Andrade, *supra* note 2.

[46] Some scholars reject the notion of group intentionality because collectives don't "exist" the way individuals do; they have no bodies, hence no minds. *See*, eg., John Hasnas, *Where is Felix Cohen When We Need Him?: Transcendental Nonsense and the Moral Responsibility of* Corporations, 18 J.L. & POL'Y (forthcoming 2010). All we see are individuals who act on "the group's behalf." But in this sense "I" don't really act either; all we see is my arm that moves the bottle and my legs that bring the arm close enough for it to move the bottle. My arm and my leg would then act "on my behalf." But individual agents get things done in a variety of ways, with the help of a variety of organic

they are not indiscriminate in those ascriptions. In fact, there is an important distinction between two kinds of "group agents" that people (and, increasingly, scholars) make.

A. Two Types of Groups

To adequately examine the social perception of group agents we must distinguish between two types of groups:[47] (a) *aggregate groups*, in which the members of a group or collective all perform the same action but do so independently (e.g., "Many New Yorkers went to the Kandinsky retrospective at the Guggenheim"); and (b) *jointly acting groups*, in which the members of a group act together as a single agent (e.g., "The Tribeca Art Club went to the Kandinsky retrospective at the Guggenheim").

The label *aggregate group*[48] captures well the fact that a perceiver literally aggregates the members of this group into a (linguistic) group category, such as "men," "high school seniors," or "Irish peasants." There is no assumption of interaction, communication, or planning among members of aggregate groups. Therefore, when we say, "With a sluggish labor market, college graduates seriously consider post-secondary education to avert unemployment," we mean that each graduate considers this option alone but that many of them do so. In this sense, an aggregate group is not an agent but a collection of agents.

Jointly acting groups engage together in deliberation, decision, planning, and (often) action. In this process, not only does each individual member undergo certain mental states, but the group decides and acts "as one." Examples include a department faculty deciding on a candidate's promotion, a design team settling on a product proposal, or a family planning a vacation. Here, interaction, communication, and cognitive sharing among members

---

and inorganic aids, so we have to be equally tolerant of groups getting things done in a variety of ways.

[47] Matthew J. O'Laughlin & Bertram F. Malle, *How People Explain Actions Performed by Groups and Individuals*, 82 J. PERSONALITY & SOC. PSYCHOL. 33, 38–39 (2002).

[48] French proposed a similar label for this type of group: "aggregate collectivities." PETER FRENCH, COLLECTIVE AND CORPORATE RESPONSIBILITY 5 (1984).

are critically implied. Therefore, when we say, "With a sluggish labor market, the board seriously considers furloughs to avert job loss," we mean that the board as a whole considers this option (even if some individual board members may be quiet or may not share the group's concerns).

This second group type meets Pettit's requirement of genuine intentional group agency—namely, that a group agent must *have reasons and form intentions to act.*[49] As we have seen, the key element in intentional agency is the belief-desire-intention (BDI) reasoning process—that is, the transition from beliefs and desires to intentions (and eventually actions). Normally these transitions occur at the individual level (in a person's mind), but for a group agent they must occur *at the group level*.[50] Thus, it is the *group* that has desires and beliefs, and forms intentions. Rather than allowing each group member to go through his or her individual reasoning process and then somehow aggregating the output of those processes into a collective intention to act, the group itself must identify a desired goal, collate the relevant beliefs, check for compatibility with other beliefs, desires, and already planned actions, and then form a proper intention to rationally pursue the goal. The individual members may not be unanimous in adopting the beliefs, desires, and intention; but whatever procedure (e.g., majority voting) licenses the group to instantiate those states, that procedure has to occur at each step of the reasoning process. Figure 2 exemplifies the process of joint group reasoning.

	Desire to hire somebody now?	Belief that P.P. is the best candidate?	Belief that P.P. is available?
Group	✔	✔	✔
Individuals	(5/6 majority)	(4/6 majority)	(4/6 majority)
Result: Group intention to offer the job to P.P.			

Figure 2. The reasoning process from desires and beliefs to intentions for a group-level agent

[49] Philip Pettit, *Collective Intentions*, *in* INTENTION IN LAW AND PHILOSOPHY 241, 241–42 (Ngaire Naffine, Rosemary Owens, & John Williams eds., 2001) [hereinafter Pettit *in* INTENTION IN LAW AND PHILOSOPHY].

[50] *See generally* MARGARET GILBERT, ON SOCIAL FACTS (1989).

At each point of the deliberation when there is room for disagreement, the group needs to say: "We believe, on the whole, that X" or "We want, on the whole, that Y." Specific procedures thus have to be put in place that *count as* the group wanting, believing, and intending something. Without such procedures, the group agent never emerges (there are only a number of individuals who try to figure out what others want and think), and any social perceiver will also be unlikely to treat the group as a joint agent.

Consider a Dean who wants to know whether the department endorses job candidate A or job candidate B. If the department chair said, "The faculty like A better," the Dean would retort, "Are you reporting on a general preference or have you voted in a faculty meeting?" That is, a mere distribution of individual preferences cannot count as the group's preference or the group's decision. Moreover, no member of jointly acting groups, not even the "leader," can form an intention *for the group* to do something; it has to be the group (according to its procedures) that makes the decisions.[51]

To illustrate the force and distinctness of such group-level intentional action, Pettit[52] discusses paradoxical situations in which a group jointly forms an intention to act even though an aggregation of individual reasoning (e.g., each faculty member making a decision in his or her office alone) would have led to rejecting the intention. This can occur when many group members lacked just one premise of the reasoning process and therefore wouldn't individually adopt the intention. Returning to Figure 2, if member 1 was lacking the desire, members 2 and 3 were missing

---

[51] In larger groups and with more complex actions, the decision process often requires distribution of partial subplans across individuals. Not all members have all the relevant desires, beliefs, and knowledge for implementing the overarching goal; even what constitutes the action may be distributed over time and space. But the same is true for individuals. Moving to the U.S. was an action I performed in 1990 (and not just at the moment the plane touched down at San Francisco International Airport), and deciding and planning to do so extended over almost two years and many locations. But this distribution does not disqualify me from being the author of that planning, that action.

[52] Pettit *in* INTENTION IN LAW AND PHILOSOPHY, *supra* note 49, at 244–47. *See also* Philip Pettit, *Groups with Minds of their Own*, *in* SOCIALIZING METAPHYSICS 167, 170–72 (Frederick F. Schmitt ed., 2003).

the first belief, and member 4 was missing the second belief, 4 of the 6 members would have individually rejected the intention to act. But in the case of genuine group reasoning, individuals subsume their doubts at any given stage under the properly licensed group reasons and therefore are bound to support the group action.[53]

The subordination of individual group members to the group intention (and action) is a key feature that distinguishes collective agency from merely in-step multiples of individual actions, or a set of individual intentions that are in agreement. Arnold, following French,[54] highlights the normative feature of decision rules that groups adhere to (and that individual group members abide by).[55] Once a department, board of directors, or Congress arrives at a decision "in a manner consistent with the procedural norms,"[56] the department, board, or Congress really has an intention to act a certain way even if not all members identify with the intention.

But when does subordination become noninvolvement? If most members of a group don't know and don't care about the issue and if the decision rules do not include a quorum, it becomes questionable whether the voting among 10% of the group should count as the intention of the whole group. Social perceivers may doubt the very fabric of this group agent, criticize the decision rules, and treat the intention not as one endorsed by the group but by an elite within the group. The tyranny of a ruling elite should not be taken as a nation's will, and when a "State" declares war on another, we should always ask who exactly had the aggressive intent and what the constituents and boundaries of that group agent

---

[53] Interestingly, many groups still vote on the last stage—the forming of the group intention. They could omit this step if they strictly followed the group reasoning process because the conclusion would logically follow from the premises. But there are other reasons why such a vote is helpful (e.g., establishing mutual knowledge, mutual commitment, and a public record of the intention).

[54] *See* PETER A. FRENCH, CORPORATE ETHICS (1995).

[55] Denis G. Arnold, *Corporate Moral Agency*, 30 MIDWEST STUD. PHIL. 279, 291 (2006).

[56] These norms will need to foster, among other things, information processing within the group that meets requirements of rationality and reasoning. For a detailed argument, see Philip Pettit, *Rationality, Reasoning and Group Agency*, 61 DIALECTICA 495 (2007).

are.[57]

The following sections introduce two test cases for the hypothesis that the social cognition of groups is fundamentally similar, though not identical, to the social cognition of individuals. The first case documents how people explain behaviors performed by groups compared with behaviors performed by individuals. The second case examines ascriptions of various mental states, particularly ascriptions of two basic classes—propositional attitudes (e.g., beliefs, desires, intentions) and phenomenal states (e.g., feelings, emotions, sensations).

B. Behavior Explanations

There is no better domain in which to demonstrate people's commitment to group-level agency than behavior explanations. The logic is straightforward. If people consider group agents to act intentionally they should explain those actions with reasons—which is, as we have seen earlier, the dominant mode by which people explain individual intentional behavior.[58] The use of reason explanations for group agents would thus provide evidence for both the application of intentionality and the use of mental state (belief and desire) ascriptions in making sense of group agent behavior.

1. Do People Use Reason Explanations When Explaining Groups' Behaviors?

O'Laughlin and Malle devised three experiments in which ordinary people explained a variety of behaviors, some performed by individuals, some performed by groups.[59] These explanations

---

[57] For a discussion, see Larry May, *State Aggression, Collective Liability, and Individual* Mens Rea, 30 MIDWEST STUD. PHIL. 309, 314–17 (2006). For a distinction between larger, amorphous groups, such as the U.S. citizenry as a whole, and more tightly organized and epistemically connected groups, such as the U.S. Supreme Court, a Board of Directors, or a university department, see J. ANGELO CORLETT, RESPONSIBILITY AND PUNISHMENT 162–63 (3d ed., 2006).

[58] *See* discussion *supra* Part I.C.1.

[59] O'Laughlin & Malle, *supra* note 47, at 36. The participants were asked to explain, in writing, a series of behaviors in the context of a fictitious

were classified into reasons and causal history of reason (CHR) explanations using the F.Ex coding scheme.[60] For ease of presentation I report relative percentages of reasons.

The stimuli were created such that each behavior could be presented as having been performed by either an individual or a group agent (e.g., "Why did Nina use drugs?" vs. "Why did the High School Seniors use drugs?").

In Study 1, actions performed by individuals elicited 71% reasons whereas actions performed by groups elicited 56% reasons. These results suggested two conclusions: first, that people do use reasons (i.e., mental states in light of which an agent decides to act) when explaining a group's action; but second, that they do so less than when they explain a group agent's action.[61] All plural agents in Study 1 were aggregate groups. We therefore interpreted the findings by noting that members of aggregate groups often have distinct reasons for acting the same way (e.g., the various reasons that "young people" may have for not voting). Citing all of those reasons would be cumbersome, so a social perceiver might do well by offering a CHR factor as a parsimonious (if general) explanation, pointing to a background that triggered the array of different reasons.

In Study 2, O'Laughlin and Malle made use of the distinction between aggregate groups and jointly acting groups to further explore both the general finding that group behaviors elicit a substantial rate of reason explanations and the more specific finding that this rate of reasons is lower than the one found in explanations of individual behavior. The authors' hypothesis was

---

conversation with a friend. The instructions emphasized that participants should formulate their explanations based on how they might actually respond within the context of such a conversation; "highly technical, or exam-like answers were discouraged." [In each] "conversation excerpt," the friend always elicited an explanation from the participant by asking why the agent performed the behavior at issue and participants answered the question by writing on three blank lines. *Id.*

[60] BERTRAM F. MALLE, F.EX: CODING SCHEME FOR PEOPLE'S FOLK EXPLANATIONS OF BEHAVIOR (2010), *available at* http://research.clps.brown.edu/SocCogSci/CodingSchemes.html.

[61] This reduction in reason explanations (and the corresponding increase in CHR explanations) for group agents was statistically reliable and explained 18% of the variance in explanations.

as follows. Jointly acting groups are united by their reasoning from beliefs and desires to intentions (see Fig. 2 above). A parsimonious (and informative) explanation for their behavior could therefore refer to those reasons, so the rate of reason explanations for jointly acting group behaviors should be similar to the rate for individual behaviors, and both should be higher than the rate for aggregate group behaviors. Indeed, the average of the jointly acting groups' reason rate and the individual agents' reason rate was 76% whereas the aggregate groups' reason rate was 62%.[62] However, one surprising finding emerged: the reason rate for jointly acting groups (81%) was even greater than the rate for individual groups (71%). This pattern was replicated in a follow-up study, which showed both the drop of reason explanations for aggregate group behaviors (46%) relative to the average of individual and jointly acting group behaviors (76%) and the even greater rate of reason explanations for jointly acting group behaviors (86%) than for individual behaviors (66%).

2. The Hyperagent Hypothesis

The reported studies document that people have no difficulties ascribing mental states *qua* reasons to whole groups, whether aggregate or jointly acting. Thus, social perceivers appear to use the same conceptual framework (their folk theory of mind and behavior) for explaining group behaviors as they do for explaining individual behaviors, and the claim that people credit "plural subjects" with minds can hardly be doubted.

But the reported findings raise the possibility that people see jointly acting groups as even more "agentic," even more driven by subjective and rational reasons, than they see individuals. I call this the *hyperagent hypothesis*. Why might groups acquire such a hyperagent status?

First, a jointly acting group's reasoning process may be particularly salient or easily imaginable because joint deliberation and joint decision making are the key features people use to identify this type of group. Second, a jointly formed group

---

[62] This difference was statistically reliable and explained 13% of the variance in explanations.

intention has presumably overcome the different interests of various group members and might therefore be seen as stronger than an individual's intention. In fact, the subordination of individual group members to the group's intention signals the unifying force that the group exerts on individuals. Third, the high degree of organization and procedure in jointly acting groups may be interpreted as a forceful determination to act.

These inferred features of salient deliberation, strength of intention, and organized action readiness can be exploited when social perceivers describe jointly acting groups as menacing and powerful. Reason explanations will therefore be particularly likely in propaganda against jointly acting groups.[63] For example, in justifying the impending attack on Poland in 1939, Hitler reportedly characterized the Eastern neighbor in the following way:

> In spite of treaties of friendship, Poland has always had the secret intention of exploiting every opportunity to do us harm.[64]

Similarly, an excerpt from an anti-homosexual document alleges the tight organization of the "homosexual movement":

> The homosexual movement is extremely well organised, and has made powerful allies and friends who lobby on its behalf . . . .
>
> . . . .
>
> In recent months, the organised homosexual movement has been lobbying vociferously, and sometimes violently, for a reduction in the age of consent.[65]

In legal contexts, too, a prosecutor may want to emphasize coordination ("conspiracy") among individual defendants committing a series of crimes. In a powerful example from history,

---

[63] For a more detailed exposition of this topic, see HOW THE MIND EXPLAINS BEHAVIOR, *supra* note 2, at ch. 8, especially section 8.4.

[64] 1 OFFICE OF U.S. CHIEF OF COUNSEL FOR PROSECUTION OF AXIS CRIMINALITY, NAZI CONSPIRACY AND AGGRESSION 391 (1946), *available at* http://www.loc.gov/rr/frd/Military_Law/pdf/NT_Nazi_Vol-I.pdf.

[65] Alexander Baron, *Baron's Guide to "Gay" Sex: A Primer for Children and Young People*, THE WEBSITE OF ALEXANDER BARON, http://www.infotext manuscripts.org/barons_guide_1.html (last visited Oct. 3, 2010).

the Allied prosecutors in the Nürnberg indictments treated the principal group of twenty-four Nazi war criminals as *conspirators*—as one jointly acting group

> In order to accomplish their aims and purposes, the Nazi conspirators prepared to seize totalitarian control over Germany to assure that no effective resistance against them could arise within Germany itself.[66]
>
> . . . .
>
> Implementing their "master race" policy, the conspirators joined in a program of relentless persecution of the Jews, designed to exterminate them.[67]

The writers liberally use linguistic expressions of planning ("aims and purposes," "prepared to seize," "program") and of reasons and goals ("in order to . . .," "designed to"), underscoring the motivation, resolution, and intentionality of the collective atrocities.

C. Mental State Ascriptions

I now turn to the second test case for the hypothesis that the social cognition of groups is fundamentally similar to the social cognition of individuals. I examine what kinds of mental states people ascribe to groups and whether these states differ in kind from the ones people ascribe to individuals.

We know from the previous section that people explain group behaviors with reasons. Reasons are typically desires and beliefs—what philosophers of mind call *propositional attitudes*. But what about other mental states? Do people feel comfortable ascribing such states as fear, love, hearing, and tasting to a group's mind? Note that in this case, ascriptions to aggregate groups (as argued earlier) are not much of an issue. For example, "The men felt embarrassed when they lost the game" can be easily interpreted as saying that each man felt embarrassed; no group mind felt embarrassed. Likewise, "The people standing in line turned

---

[66] 2 INT'L MILITARY TRIBUNAL, TRIAL OF THE MAJOR WAR CRIMINALS BEFORE THE INTERNATIONAL MILITARY TRIBUNAL 33 (1947), *available at* http://www.loc.gov/rr/frd/Military_Law/NT_major-war-criminals.html.

[67] *Id*. at 35.

frustrated and impatient because only one counter was open" says only that each of the people turned frustrated and impatient. By contrast, jointly acting groups are candidates for ascriptions of such affective states: Did the BP corporation feel embarrassed over its handling of the Gulf oil spill in 2010? And was the U.S. government angry at BP?

1. Different Mental States?

In classic philosophy of mind work, we find a distinction between two main classes of mental states: *propositional states* (e.g., beliefs, desires, intentions) and *nonpropositional* states (e.g., pain, feeling sad, smelling fresh coffee).[68] Propositional states are understood by many to be "computational"—that is, they could in principle be implemented in other media besides the human brain, such as in computers or aliens. The other, nonpropositional class of states has often been characterized as *phenomenal*—as having a certain experiential quality. Ascriptions of phenomenal states to groups may be more restrictive; they may require a certain brain/body or a unitary mind for implementation.

Robbins and Jack proposed that people take a *phenomenal stance* to certain other creatures, and certainly to other human beings.[69] That means they regard them as a locus of experience and ascribe to them a variety of phenomenal states (emotions, moods, pains, visual sensations, etc.). This stance is contrasted with the intentional stance,[70] according to which people ascribe to others a variety of propositional states (e.g., belief, desire, intention). One key feature of Robbins and Jack's proposal is that the phenomenal stance comes with a consideration of the other creature as having moral standing. Ascribing to another the capacity for distress comes with a desire to prevent such distress and, in particular, to shield the other from potential harm. Thus, the phenomenal stance

---

[68] For an overview of this distinction, see generally David Pitt, *Mental Representation*, STAN. ENCYC. PHIL. (last updated July 21, 2008), http://plato.stanford.edu/entries/mental-representation.

[69] Philip Robbins & Anthony I. Jack, *The Phenomenal Stance*, 127 PHIL. STUD. 59, 69–70 (2006).

[70] *See generally* DANIEL C. DENNETT, THE INTENTIONAL STANCE (1987).

is "morally compelling."[71] I will return to this potential link between ascriptions of phenomenal states and considerations of moral standing.

What evidence is there for the claim that ordinary people make a distinction between propositional and phenomenal states? Gray, Gray, and Wegner asked participants to ascribe a variety of mental states and capacities to different agents.[72] The pattern of these ascriptions suggested a two-dimensional space: One axis, which the authors labeled "Experience," was constituted primarily by phenomenal states (e.g., hunger, fear, pain, pleasure); the other axis, which the authors labeled "Agency," was constituted primarily by states of higher cognition (e.g., self-restraint, moral judgment, memory, and planning). However, very few propositional states were included in the study, and those that were (desire, planning) did not cluster together. The results also allow for alternative interpretations. The items defining the first axis can be understood as capturing unintentional states, and the items defining the second axis can be understood as capturing intentional states. Likewise, the first axis can be considered a dimension of affect and the second a dimension of cognition.

2. Different Mental States For Groups?

Knobe and Prinz conducted a series of studies that explore both the distinction between propositional and phenomenal states and

---

[71] Robbins & Jack, *supra* note 69, at 70.

[72] Heather M. Gray, Kurt Gray & Daniel M. Wegner, *Dimensions of Mind Perception*, 315 SCIENCE 619, 619 (2007). The states included, among others, desire, embarrassment, hunger, memory, and morality; the agents included, among others, a robot, a frog, a human infant, adult humans, and god. Each participant evaluated just one state and indicated for all possible pairs of agents whether one agent had more of that state than the other agent (a procedure that facilitates between-agent differentiation rather than between-state differentiation). The comparative ratings were then aggregated for each agent and averaged across people who considered the same state. This resulted in mean ratings for thirteen agents on eighteen states. The data analysis was unconventional because a principal components analysis (PCA) was run on this matrix of thirteen rows and eighteen columns, even though a common requirement of PCA is to ensure 5–10 times as many rows as columns.

the distinction between individual and group agents.[73] The results suggested that people are comfortable ascribing propositional states (e.g., decide, want, intend, believe, know) to group agents,[74] but that they are reluctant to ascribe phenomenal states to them (a sudden urge, great joy, vividly imagining, getting depressed, feeling excruciating pain).[75] The authors favored a strong interpretation of these findings, according to which people consider groups to uniquely lack the capacity for "phenomenal consciousness." The evidence does not strongly support this interpretation, however.

In Knobe and Prinz's studies, even though people did not deem natural such statements as "Acme Corp. is *feeling upset*," they found it perfectly natural to say "Acme Corp. *is upset* about the court's recent ruling."[76] It would be a rather unusual conception of phenomenal states if one could *be in* those states but not *feel* them. Fortunately, however, people's conception does not appear to be that unusual. Sytsma and Machery were unable to replicate the difference between "feeling upset" and "is upset" (and similarly for "regret").[77] More important, for both formulations people on average endorsed the midpoint between "sounds clearly weird" and "sounds clearly natural"—not exactly evidence for a refusal to ascribe phenomenal states to groups. Arico, too, failed to replicate the importance of the "feeling" verb, and his participants similarly rated the naturalness of phenomenal group states at just above the midpoint of the scale.[78] These findings support only the contention

---

[73] *See* Joshua Knobe & Jesse Prinz, *Intuitions About Consciousness: Experimental Studies*, 7 PHENOMENOLOGY & COGNITIVE SCI. 67 (2008).

[74] The group agents in point were Microsoft and Acme Corp. *Id*. at 74–75.

[75] On a scale of 1–7, where 1 meant "sounds weird" and 7 meant "sounds natural," the propositional states received average ratings between 5 and 7, and the phenomenal states received average ratings between 2 and 5. *Id*. at 75.

[76] Knobe & Prinz, *supra* note 73, at 77–78. Note that the compared pairs differ not only in the presence of the "feeling" verb but also in the provision of actual content (being upset *about* something in particular). *Id.* Arico did not find any difference in naturalness ratings for either of the two differing features. Adam Arico, *Folk Psychology, Consciousness, and Context Effects*, 1 REV. PHIL. & PSYCHOL. 371, 379 (2010) [hereinafter Arico, *Folk Psychology*].

[77] Justin M. Sytsma & Edouard Machery, *How to Study Folk Intuitions About Phenomenal Consciousness*, 22 PHIL. PSYCHOL. 21, 28–30 (2009).

[78] Arico, *Folk Psychology*, supra note 76, at 379. Arico did find a modest

that people are unsure about whether phenomenal states can be ascribed to groups, not that they reject them outright or find them conceptually incoherent.

3. Drawing the Right Conclusions

The current state of evidence is thus highly mixed. On the one hand, one might conclude that people treat group agents quite similarly to individual agents, all the way to the ascription of many (if not all) phenomenal states. There appears to be a subtle difference in comfort when ascribing phenomenal states to individuals or groups, but that difference is not well understood at this point. On the other hand, one might conclude that there is really something remarkable about the lesser comfort people have when ascribing phenomenal states to groups. It tells us something about the uniqueness of group agents, even if this uniqueness has fluid boundaries.

It comes down to the question of how we should explain the small but perhaps consistent comfort difference. One hypothesis is that it results from the partial semantic fit of mental state terms with the physical substrate of groups. For example, people have no trouble with group agents wanting or desiring something but are more reluctant to ascribe a sudden urge to them. This may be due to specific semantic components being unmet (e.g., urges have physiological components whereas desires do not), and the overall semantic fit is therefore somewhat reduced when particular verbs are ascribed to groups. But if we learned that groups cannot do certain things because they lack a (single) body,[79] should we consider this to be a disappointing insight?

Such disappointment may be premature. Even though people provide intermediate judgments of "naturalness" for verbal

---

difference in perceived naturalness of group vs. individual mind ascriptions (4.5 vs. 5.6 on the familiar 1–7 scale). But a variety of nonconceptual factors may account for such a modest difference (e.g., frequency of encountering one kind of descriptions). *Id*. at 380.

[79] Sytsma and Machery indeed show that people are highly reluctant to ascribe to groups such behavioral attributes as "being murdered," "napping," or "having insomnia," all requiring a body. Sytsma & Machery, *supra* note 77, at 26.

descriptions of phenomenal states, they still fail to spontaneously and frequently attribute such states to group agents. Knobe and Prinz claimed, on the basis of a Google search, that virtually no instances exist of phrases such as "Microsoft feels angry" or "depressed" or "scared," whereas thousands of instances exist of phrases such as "Microsoft decides" or "wants" or "hopes."[80] If people are indeed rarely exposed to verbal ascriptions of phenomenal states to group agents, the mere mental association between such states and groups will be weak, and people may be unlikely to search for or infer such states in groups. I call this the hypothesis of a reluctant inferential stance.

4. The Reluctant Inferential Stance

If people are reluctant to infer a family of affective and phenomenal states to groups, an intriguing possibility emerges: a group agent who rarely has worries and does not feel pain, who cannot be intimidated and has no regrets, will be seen as a perfectly self-interested, calculating agent. Even if individual members of a group may experience all these states, they are dispensable and exchangeable—the group's structure and organization can still maintain its calm rationality. Such an image

---

[80] Knobe & Prinz, *supra* note 73, at 73–74. The authors don't provide details about their search procedure, so the representativeness of their test is difficult to evaluate. I conducted a very brief search with the word pairs "Microsoft decide[]" as well as "Microsoft angry" and "Microsoft happy." Already the first results page of each search contained hits. Two sources ascribed anger to Microsoft: Trent Nouveau, *Microsoft Angry at Google Over Vulnerability Disclosure*, TG DAILY (June 11, 2010), http://www.tgdaily.com/security-features/50181-microsoft-angry-at-google-over-vulnerability-disclosure; *Microsoft Gets Angry at Yahoo-Google Tie Up*, THE INQUIRER (July 28, 2010), http://www.theinquirer.net/inquirer/blog-post/1725295/microsoft-angry-yahoo-google-tie. Three sources ascribed happiness to Microsoft, including Roy Schestowitz, *Microsoft Happy About Apple's Invocation of Software Patents Against GNU/Linux*, TECHRIGHTS (Mar. 16, 2010), http://techrights.org/2010/03/16/mobile-linux-victory. However, occurrences of "decide" were clearly more frequent overall in my search (about 8/10 per page). So people are not conceptually resistant to ascribing phenomenal states to group agents, but it is safe to say that people are far less frequently exposed to verbal ascriptions of group agents' phenomenal states.

would further contribute to the notion of jointly acting groups being hyperagents—agents that display more commitment, exert more power, and ultimately pose a greater threat than any individual agent could.

Such an image would have a further consequence. Agents who do not feel worry, regret, or pain are unlikely to be responsive to social censoring—to warnings, threats, and punishment. Such censoring is of course part and parcel of community members' moral treatment of one another. Therefore, we must ask: How do people morally treat a group agent? Are group agents proper targets of moral expectations? Moral evaluation? And moral punishment? These questions will occupy us in Part III of this article. First I consider people's moral judgment of individual agents.

III. MORAL JUDGMENT OF INDIVIDUAL AGENTS

A. Morality Is Embedded in Folk Psychology

When people make moral judgments about an agent, they evaluate the agent's behavior by considering (a) norms that the behavior may have violated and (b) what was in the agent's mind before, while, and even after performing the behavior. The latter considerations are a direct reflection of people's folk theory of mind (described in Part I of this article). We can thus say that this folk theory lies at the heart of moral judgment.[81]

Humans do not normally make moral judgments about earthquakes or hurricanes. What makes moral judgments genuinely moral is that they are directed at *agents* who are presumed to be capable of following socially shared standards of conduct. Hence, the agency concept is a crucial ingredient of moral judgments. Specifically, blame is assigned in consideration of an agent's principled capacity to reason about various paths of action and the capacity to intentionally pursue one such path. And even when

---

[81] I will focus on blame judgments, which can be considered the paradigmatic moral judgment. *See generally* Steve Guglielmo, Andrew E. Monroe & Bertram F. Malle, *At the Heart of Morality Lies Folk Psychology*, 52 INQUIRY 449 (2009).

harm occurs unintentionally, if the person could have and should have chosen a harm-avoiding alternate path but didn't, blame applies. Hence, the concept of intentionality is a crucial ingredient of blame judgments.[82] My colleagues and I have developed a working model of blame that prominently features the concept of intentionality and also integrates a number of well-supported features of blame.[83]

B. A Step Model of Blame[84]

People can be blamed for outcomes or behaviors. However, even in the case of outcomes, the agent is ultimately blamed for the behavior of causing, allowing, or failing to prevent the outcome. Thus, to streamline the presentation, I designate as the first step in the blame process the perceiver's detection of a norm-violating behavior.[85]

---

[82] Fiery Cushman, Liane Young & Marc Hauser, *The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm*, 17 PSYCHOL. SCI. 1082, 1083 (2006); Guglielmo, Monroe & Malle, *supra* note 81, at 451; *see generally* ANTHONY KENNY, *The History of Intention in Ethics*, *in* THE ANATOMY OF THE SOUL: HISTORICAL ESSAYS IN THE PHILOSOPHY OF MIND 129 (1973).

[83] Guglielmo, Monroe & Malle, *supra* note 81, at 450–52.

[84] This is a "step" model of blame because several information processing elements build on each other and will often be temporally ordered. However, as with all complex information processes, backward loops, premature processing, and omissions can occur. Moreover, the specific level of awareness that perceivers exhibit at each of these steps may vary. At least some of the time some of the steps will be conscious.

[85] When detection is direct perception of behavior ("This is wrong!" "This is forbidden!"), almost no cognitive work is needed; instead, we may see the operation of moral "intuitions" and "moral grammar rules." See references to Haidt and to Mikhail, *infra* note 86. Such detection may come with automatic negative affect (e.g., anger, disgust) that may influence subsequent processing steps. Claims about such influence (and especially of undue bias) have been frequent but often rely on highly indirect evidence—researchers almost never distinguish the early norm-breach affect from the later blame-accompanying affect.

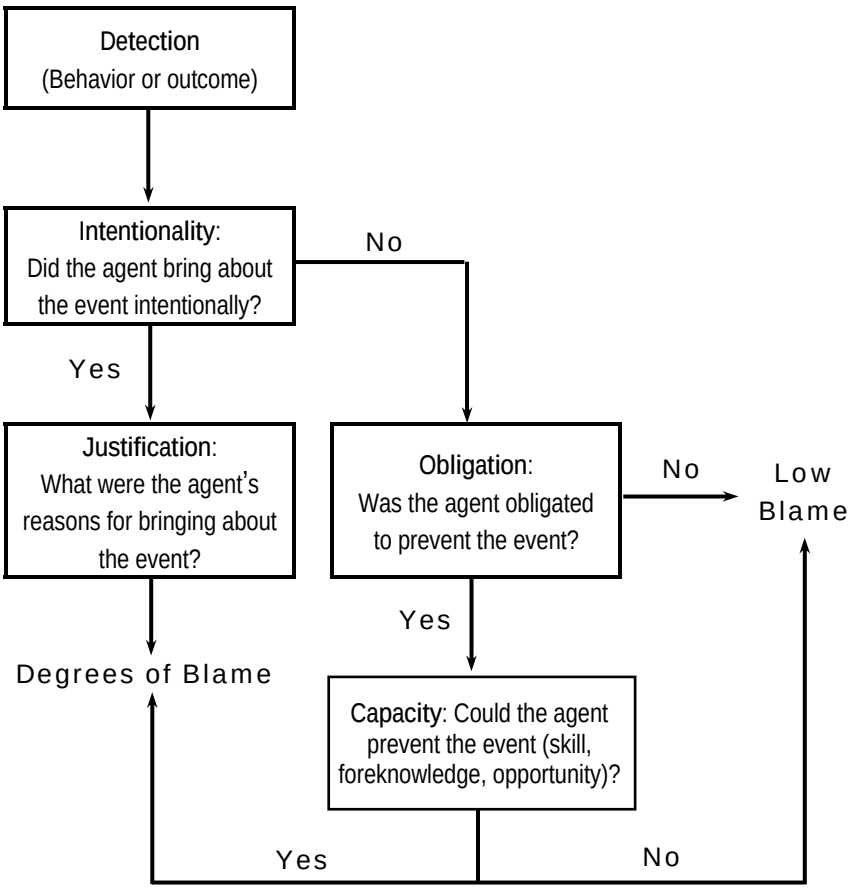

Figure 3. Step model of ordinary assessments of blame.[86]

[86] Guglielmo, Monroe & Malle, *supra* note 81, at 450–51.

The detection step may or may not be considered a genuine moral judgment yet,[87] but it certainly is a necessary requirement of a genuine blame judgment. During the detection phase the perceiver focuses on evaluating the "badness" of the behavior; during the judgment phase the perceiver focuses on the blameworthiness of the agent. Put simply, people don't blame behaviors, they blame agents.

Blaming an agent requires assessing the agent's mental states involved in the behavior. The key assessment here is the familiar question of whether the behavior was intentional or unintentional, which constitutes the next step in the model. This step is pivotal because it bifurcates the perceiver's further information processing into two paths. In the case of intentional behavior (the left path in Fig. 3), the perceiver considers the agent's particular reasons to act. In the case of unintentional behavior (the right path in Fig. 3), the perceiver considers the combination of the agent's obligation and capacity to prevent the event in question.

The reasons that the perceiver considers along the left path are the agent's beliefs and desires that lead to her intention to act, indicating that once people determine that an agent intentionally breached a norm they want to know why she acted this way.[88] These reasons will strongly increase or decrease blame by way of *justification*. An agent who hurt someone intentionally may have had justified reasons (e.g., a dentist trying to extract a child's unhealthy tooth) and will be blamed less than the one who had unjustified reasons (e.g., a schoolboy trying to provoke a fight). What reasons provide justification is of course a manner of community norms.

---

[87] Many treatments of moral judgment do not clearly distinguish the norm-violating phase from the blaming phase. Those that do vary in how much affect they postulate is involved in the early phase. Haidt argues there is much affect, Cushman as well as Mikhail argue that cognitive processes dominate. *See* Fiery Cushman, *Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment*, 108 COGNITION 353, 375–78 (2008); Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCHOL. REV. 814, 817–19 (2001); John Mikhail, *Universal Moral Grammar: Theory, Evidence and the Future*, 11 TRENDS IN COGNITIVE SCI. 143, 143 (2007).

[88] *See supra*, Part I.C.1 on reason explanations.

The information that perceivers consider along the right path is quite different. When people regard the agent as having acted unintentionally, they examine whether the agent *should have* prevented the event (obligation)[89] and *could have* prevented it (capacity).[90] Both of these considerations are tied to the intentionality concept. Social communities impose obligations on individuals because they expect them to intentionally act (or at least intend to act) in accordance with these obligations. If the agent lacks the requisite capacities to meet the obligations—cognitive capacities such as knowledge or reasoning and noncognitive capacities such as skill or opportunity—little to no blame will apply. If the agent has the necessary capacities and is subject to the obligation, then a failure to prevent the negative event will trigger substantial blame. This blame will vary by how much the failure is considered imprudent, negligent, or reckless behavior, or possibly even an intentional, deliberate breach of one's obligation.[91]

This model has a fair amount of overlap with legal decision making in history and current practice—and this should not be too surprising given that the law codifies some basic human features of

---

[89] For related discussion of these two elements, see *The Significance of Intentionality*, *in* INTENTIONS AND INTENTIONALITY, *supra* note 9, at 19–23. Obligation varies primarily with role, such as position within a hierarchy—see generally V. Lee Hamilton, *Chains of Command: Responsibility Attribution in Hierarchies*, 16 J. APPLIED SOC. PSYCHOL. 118 (1986)—and type of relationship (e.g., stranger vs. friend). Jonathan Haidt, & Jonathan Baron, *Social Roles and the Moral Judgment of Acts and Omissions*, 26 EUR. J. SOC. PSYCHOL. 201, 202–04 (1996).

[90] This roughly corresponds to Weiner's concept of controllability, which in his model is the critical precursor to responsibility judgments, which in turn foster blame (or sympathy). *See generally* BERNARD WEINER, JUDGMENTS OF RESPONSIBILITY: A FOUNDATION FOR A THEORY OF SOCIAL CONDUCT (1995).

[91] An intentional breach of one's obligation is itself a norm violation and will be evaluated along the left arm of the step model: the perceiver now considers the agent's justified or unjustified reasons for the breach. Perceivers may thus initially form a moral judgment about an agent's unintentional causing of a negative outcome and end with a moral judgment about the agent's intentional failure to prevent the outcome. For further discussion, see Steve Guglielmo & Bertram Malle, *Can Unintended Side-Effects be Intentional? Resolving a Controversy over Intentionality and Morality*, 36 PERSONALITY & SOC. PSYCHOL. BULL. (forthcoming 2010).

moral judgment. For example, murder is typically defined as intentional killing, which requires both intentional "body movements" and the mental state of "intent to kill."[92] More generally, criminal responsibility is often defined as the pairing of a harmful act and the "corresponding mental state or intent."[93] But the overlap ends when the courts and the literatures of philosophy and law engage in disputes about the exact meaning of *intention*.[94] These disputes are often fanned by individual scholars' intuitions about what intentionality is and they reflect historical and geographical variation in restricting or expanding the meaning of relevant terms (e.g., *intentionally*, *willingly*, and *knowingly*).[95] By contrast, the advantage of an empirical-based model of moral judgment such as the step model is that the terms don't have to be debated repeatedly, as they are fixed by ordinary people's folk concepts of intentionality, intention, and the like.

The main difference between the science of moral judgment and legal scholarship is that science typically focuses on a descriptive mission—clarifying the concepts and processes that guide people's behavior—whereas the law focuses on a reformative mission (changing people's thinking and behavior). Many legal and philosophical discussions, however, conflate the two missions by using ordinary words (e.g., *intention*) when they propose to reform people's thinking or attempt to sharpen the meaning of terms in legal texts. When using ordinary words, the reformed or sharpened meaning will often clash with the words' ordinary meanings (at least for those words that stand for a deeply ingrained folk concept, such as intention). Far better would be to introduce terms of art that, by steering clear of familiar folk concepts, may be more successful in shifting people's thinking or

---

[92] Stephen J. Morse, *Craziness and Criminal Responsibility*, 17 BEHAV. SCI. & LAW 147, 148 (1999).

[93] Alan R. Felthous, *Introduction to Mental Illness and Criminal Responsibility*, 17 BEHAV. SCI. & LAW 143, 143 (1999).

[94] *See, e.g.*, R. ANTHONY DUFF, INTENTION, AGENCY AND CRIMINAL LIABILITY (1990); KENNY, *supra* note 82, at 129; Nicola Lacey, *A Clear Concept of Intention: Elusive or Illusory?*, 56 THE MOD. L. REV. 621 (1993).

[95] *See generally* Justin D. Levinson, *Mentally Misguided: How State of Mind Inquiries Ignore Psychological Reality and Overlook Cultural Differences*, 49 HOW. L.J. 1 (2005); Malle & Nelson, *supra* note 22, at 569–70.

attitudes.[96]

Because I follow a descriptive mission here, the critical question now is this: Having in place a model of how people make blame judgments of individual agents, what do people do when they face a *group agent's* immoral behavior?

IV. MORAL JUDGMENT OF GROUP AGENTS

There is broad agreement in the literature that a group's capacity for intentional action is a prerequisite for the group's status as a moral agent. As Isaacs put it, "showing that collectives are capable of intentional action is necessary for showing that they are appropriate objects of praise and blame."[97] The capacity to exemplify intentionality, mental states, and reason-based choice (rationality) is also what French postulates as central in rendering a corporation a "moral agent."[98] He argues that corporations are moral agents *because* they are capable of intentional action. These are claims about the metaphysics of corporations; however, they are in accordance with ordinary social perception. As we have seen in Part II, Section B, people regard as agents groups who act intentionally and have reasons for their actions.

But the status of corporations and other group agents as intentional agents makes them only *eligible* for moral evaluation. What does such evaluation look like in detail? Does it work the same way as for the moral evaluation of individuals? We need not automatically assume that collective moral judgment operates the same way, but if there is no evidence to the contrary, we may continue to accept it as a working hypothesis. This equal-operation hypothesis is also strengthened by a basic theoretical argument. If people's powerful folk psychology is unflappably applied to group agents, and if moral judgment is deeply drawing on folk psychology, then moral judgment, too, should be applied to group

---

[96] Malle & Nelson, *supra* note 22, at 565.

[97] Tracy Isaacs, *Collective Moral Responsibility and Collective Intention*, 20 MIDWEST STUD. PHIL. 59, 62 (2006).

[98] *See* Peter A. French, *The Corporation as a Moral Person*, 16 AM. PHIL. Q. 207 (1979). A revision of the postulate of corporations as moral agents is described in Peter A. French, *Integrity, Intentions, and Corporations*, 34 AM. BUS. L.J. 141 (1996).

agents.

To test the equal-operation hypothesis in more detail, we need to examine the elements of the step model of blame and examine whether judgments of group action could be shuttled through a cognitive apparatus with those elements.

A. Applying the Step Model of Blame to Group Agents

A brief look into any newspaper reveals that people easily and often detect norm-violating group behaviors—performed, for example, by teams, gangs, corporations, parties, governments, or nations. The specific norms for groups may differ from those for individuals, but for the norms that do apply, moral breaches are certainly recognized and formulated.

People also have no trouble distinguishing between intentional and unintentional group behavior.[99] Unintentional collective behaviors may be less frequent than intentional ones,[100] but acts of negligence (by definition unintentional) are commonplace in accusations of objectionable corporate behavior.

Following the left path of arriving at blame in the step model, we know that people ascribe reasons to group agents, so we can expect people to consider reasons as possible blame moderators for norm-violating actions. A corporation or government will certainly offer such justifying reasons in order to mitigate potential blame for its actions.

Following the right path of arriving at blame, the existence of norms for group agents implies that there are obligations for group agents as well, for being subject to a norm means being obligated to conform to it, and if there is a norm of prevention (especially of

---

[99] It might seem that in Pettit's model, Pettit *in* INTENTION IN LAW AND PHILOSOPHY, *supra* note 49, no unintentional behavior can be a genuine group action because the very decision process from reasons to intention are the prerequisites of his model of collective action. However, a group decision process may have unintended consequences that one can ascribe to the group, which is then an unintentional behavior performed by the group. Once the group has demonstrated, through undergoing their appropriate decision process, that it is a genuine group agent, it can perform both intentional and unintentional behaviors.

[100] O'Laughlin & Malle, *supra* note 47, at 34.

harm), the obligation to prevent will fully apply.

Furthermore, groups arguably vary in their capacities to prevent possible negative outcomes. They vary in their knowledge of certain facts as well as in skills and opportunities to execute certain intentions (to bring about or prevent outcomes). Ascriptions of knowledge (or lack thereof) to groups have been documented in Part II, Sections B and C, and variations in skills and opportunities are surely uncontroversial.

Thus we arrive, without making contentious assumptions, at a picture according to which group agents can be blamed through the operation of the same cognitive apparatus through which individuals are blamed. We have no direct evidence that social perceivers form group blame following only these steps, but there are at least no apparent obstacles for them to do so.

B. Ongoing Research

In a recent experiment, Dillon and Malle gathered some evidence for people's judgments of intentionality and blame in response to group agent behavior.[101] University student participants considered a number of actions performed by individuals and groups and made a variety of judgments about those actions: "Was the behavior intentional?" "Did you detect what the agent was thinking?" "Does the agent deserve praise or blame?" The key measurements were participants' rates of affirmative responses (indicating an inference of intentionality, thinking, or blame, respectively) and the response latency for those responses. Importantly, all actions were formulated in three ways: as performed by an individual, by an aggregate (e.g., inner city kids across the nation; students in the psychology class), and by a jointly acting group (e.g., the Latin Student Organization; the senior design team).[102] The results suggested that both individual and group agents elicited similar and differentiated rates of inference (i.e., intentionality was inferred more frequently than

---

[101] Kyle Dillon & Bertram F. Malle, Ease and Speed of Social Inferences from Individual and Group Behaviors (Sept. 2010) (unpublished data, Brown University) (on file with author).

[102] Each participant saw actions performed by all three groups, but never the three versions of the same action.

thinking, which was inferred more frequently than blame) and that speed of inferences were also remarkably similar across agents (with intentionality the fastest). The only noteworthy difference due to agent type emerged in the absolute frequency of intentionality inferences, where individual and jointly acting group agents elicited more affirmations of intentionality than did aggregate groups. However, when people did make those inferences, they accessed them equally fast.

These results are only preliminary until we demonstrate consistency across stimulus properties and task demands. However, they do provide an indication that it is quite easy and natural for people to make mental inferences and moral judgments about group actions.

V. THE MORAL LIFE OF GROUP AGENTS, LTD.

A. Problems for Blaming Groups

In light of the current evidence, we can be confident that people assign blame to group agents and do so with essentially the same psychological apparatus that they engage when blaming individual agents. But blame has two faces: the cognitive, on which I have focused, and the social, which so far I have set aside. This social face of blame consists of verbal or physical acts that express the perceiver's moral judgment and are typically directed at the target agent, presumably for purposes of behavior regulation. Punishment may follow, and expressed blame is one kind of social punishment.[103]

But here is the first problem: How well can social perceivers express blame toward group agents? People do not actually encounter nations, governments, or corporations; even teams or committees are rarely seen face to face. In modern life, people can write letters to a group agent, sue them, or publicly denounce them. But these expressions will be rare, limited in scope, and come with

---

[103] A more intense social punishment is exclusion. For a review, see generally Rorbert Kurzban & Mark R. Leary, *Evolutionary Origins of Stigmatization: The Functions of Social Exclusion*, 127 PSYCHOL. BULL. 187 (2001).

little assurance that the addressee actually notices the blame.

The second problem is this: If blame is rarely expressed and even more rarely heard, regulation of group agents' behavior runs idle. Of course, a social perceiver can vote against a government or refuse to buy from a company; but here she alters her own actions more than the group agent's actions. Only when individual social perceivers aggregate or join together can social blame and punishment become an effective regulator. Thus, it often takes a group agent to fight or put in its place another group agent.

A third problem was presented in Part II, Section C: If group agents lack affective mental states, they will also be unlikely to feel guilt, regret, or remorse.[104] As a result, groups will have fewer moral scruples, which further blocks social regulation as well as deterrence. If groups are rational, solely cognitive agents, potential punishment becomes part of the utility calculation for their actions; anticipated guilt or regret lies outside these calculations.

Fortunately there are boundaries to this bleak picture. As long as individual group members feel moral emotions and fear punishment, their influence on other members can alter group action. In this case, the locus of moral emotions is not the group agent but the individual who alters the decision making of the group agent. In the frame of Pettit's model, individual members' moral emotions can sway votes and lead to jointly adopted or rejected group reasons and intentions.[105] Such a process requires an accessible and transparent structure and a manageable group size so that individuals can express their emotions and thereby influence others' preferences and beliefs. Then the group can decide to forgo certain opportunities because they are morally objectionable. The group itself never instantiates guilt or raises a moral objection, but individuals do.

---

[104] Strictly speaking we cannot currently determine whether group agents actually lack such feelings or are only perceived to lack them. For simplicity I will continue to use the objective formulation.

[105] *See* Pettit *in* INTENTION IN LAW AND PHILOSOPHY, *supra* note 49. Within individuals, the power of moral emotions is not radically different: the emotions sway the person's decision making by changing desire and beliefs, making costs salient, or prospects unattractive. That the person *feels* the emotions ensures that those changes are executed.

B. Blaming Individual Group Members

If blaming and punishing group agents is difficult or ineffective, people have the option of punishing individual group members.[106] However, matters do not necessarily become easier here as individual group members may be viewed as differentially responsible for a group's undesirable action and should therefore be differentially punished.[107] But what guides such differentiation? The step model of blame may track the relevant factors.

What will matter first is the degree to which the individual was causally contributing to the group action. After the team's loss, substitute players are not subject to the same sanctions as starting players. The individual's motivation or justification for committing the act will be assessed as well. Was he under pressure from the group or intrinsically motivated to achieve the specific outcome? Did she vote with the group majority or against it when the group action was adopted?

The obligation of the individual in preventing the act will also be questioned: Is the citizen obliged to prevent the nation's war? Is the worker obliged to stop the company's discrimination?

Capacities, too, will be of importance, both cognitive ones (how much did the person know?) and physical ones (how much ability did he have to actually avert the group action?).

A final element is the level of identification or distancing that the group member displayed. If group membership is a choice (less so for nations than teams), then maintaining group membership counts as an endorsement of the group's collective actions. Attempts to protest, to distance oneself from those actions, even if these attempts were squelched, will count in the individual's favor. Conversely, happily going along with the group action when distancing was possible (but endorsement brought personal benefits) will count against the individual. A classic case of such complex assessments is the involvement of soldiers and citizens in the Holocaust. How much did the German and Austrian (and

---

[106] *See* Tracy Isaacs, *Individual Responsibility for Collective Wrongs*, *in* BRINGING POWER TO JUSTICE?: THE PROSPECTS OF THE INTERNATIONAL CRIMINAL COURT 167 (Joanna Harrington et al. eds., 2006).

[107] Paul Sheehy, *Holding Them Responsible*, 30 MIDWEST STUD. PHIL. 74, 89–90 (2006).

Polish and French . . .) citizenry know about the Nazi command's genocide? What could they have reasonably done to prevent it? What signs of endorsement or distancing can be documented?

These are exceptionally difficult questions in their own right, but they are made even more difficult because people's moral judgment apparatus has a single agent node and cannot easily handle multiple agents in intertwinement. Individuals are relatively easy to judge; perhaps group agents, too. But individuals *as* members of a group agent are not.

CONCLUSION

A considerable number of group agents are perceived as powerful, sometimes threatening, perhaps morally unregulated. Why? Humans have no trouble reasoning about the actions and minds of groups and have the desire to blame and punish them when they act immorally. But the ability and effectiveness to blame and punish groups is limited, particularly for institutionalized group agents (e.g., governments, corporations, committees). Moreover, such group agents rarely show the common emotions of fear, guilt, and remorse. All that fans people's perceptions of group agents as threatening, powerful, and morally unalterable. It also causes deep frustration, which in turn may explain people's growing distaste for corporations and the substantial damage awards in legal cases that punish large companies. Perhaps it also explains the political see-saw in most democratic countries, where the ideological tide switches in every other election, right after the last group agent was blamed for all the society's ills.

Thus, the modern world presents the human mind with group agents that trigger all the familiar social-cognitive and moral responses but that leave little room to act on those responses. Perhaps the law can provide some room for appropriate action—by codifying norms, obligations, and punishment not just for individuals but also for groups. In doing so, however, the law must heed the concepts and criteria by which ordinary people recognize group agents and judge their moral conduct.