

<https://helda.helsinki.fi>

A model-based approach to social ontology

Sarkia, Matti

2022-06

Sarkia , M 2022 , ' A model-based approach to social ontology ' , Philosophy of the Social Sciences , vol. 52 , no. 3 , pp. 175-203 . <https://doi.org/10.1177/00483931211056952>

<http://hdl.handle.net/10138/345466>

<https://doi.org/10.1177/00483931211056952>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

A model-based approach to social ontology

Matti Sarkia 

Philosophy of the Social Sciences
2022, Vol. 52(3) 175–203
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00483931211056952
journals.sagepub.com/home/pos



Abstract

This paper argues for theoretical modeling and model-construction as central (but not necessarily the only) types of activities that philosophers of social ontology (in the analytic tradition) engage in. This claim is defended through a detailed case study and revisionary interpretation of Raimo Tuomela's account of the we-perspective. My interpretation is grounded in Ronald Giere's account of scientific models, and argued to be compatible with, but less demanding than Tuomela's own description of his account as a philosophical theory of the social world. My approach is also suggested to be applicable to many (but not necessarily all) other methodologically naturalist accounts of collective intentionality and social ontology.

Keywords

Social ontology, collective intentionality, methodological naturalism, philosophical methodology, scientific modeling

Introduction

The research field of social ontology has been described in various ways by philosophers.¹ For example, [Epstein \(2018\)](#) describes social ontology as “the

¹Social ontology has also been discussed by theoretically oriented economists (e.g., [Lawson 2019](#)) and sociologists (e.g., [Archer 2017](#); [Elder-Vass 2012](#); [Kaidesoja 2013](#)). This paper will focus on discussions of social ontology in the field of analytic philosophy.

Department of Political and Economic Studies, University of Helsinki, Helsinki, Finland

Corresponding Author:

Matti Sarkia, Department of Political and Economic Studies, University of Helsinki, Unioninkatu 40 A, Helsinki 00014, Finland.

Email: matti.sarkia@helsinki.fi

study of the nature and properties of the social world... concerned with analyzing the various entities in the world that arise from social interaction.” Tuomela (2013, ix) says that “social ontology is not only a study of the basic nature of social reality but at least in part a study of what the best-explaining social scientific theories need to appeal to in their postulated ontologies.” And Searle (2010, 5) describes social ontology as concerned with “the mode of existence of social entities such as governments, families, cocktail parties, summer vacations, trade unions, baseball games, and passports.” These philosophers describe central concerns of social ontology as including the explanation of social action, the metaphysics of social reality, and the nature of social facts. However, while most contemporary philosophers regard social ontology as deeply important and even foundational for the social sciences, many social scientists care more about research methods and empirical investigation of contingent social facts than about ontological questions concerning the fundamental nature or constituents of the social world. The relation between social ontology and social scientific methodology accordingly remains an important topic, which has received too little attention in the literature so far.

This paper defends a methodologically naturalist approach to social ontology, which is based on theoretical modeling and model-construction as central (but not necessarily the only) types of activities that philosophers working in this field of inquiry engage in. This conception of social ontology is grounded in naturalist approaches to the philosophy of science, which distinguish theoretical modeling and model-construction “in the laboratory of the mind” from empirical investigation by means of controlled observation and experimentation (e.g., Bechtel & Richardson 2010; Downes 2011; Giere 1988; 2004; Godfrey-Smith 2006; Hausman 1992; Mäki 2009; Weisberg 2013; Wimsatt 2007; Woodward 2003). They also recognize that even highly abstract and idealized theoretical models can serve as proper parts or constituents of empirical theories, when they are combined with *theoretical hypotheses* (Giere 1988) asserting that some part or aspect of the world is similar to the model in relevant degrees and respects. However, I will not insist that *all* parts of social ontology are concerned with theoretical modeling and model-construction, nor will I argue that the parts of social ontology that I will discuss *must* be understood in these terms, since there may sometimes be several alternative meta-theoretical descriptions of the same field of inquiry (Giere 2004). My focus will be especially on those parts of social ontology, which are concerned with collective intentionality and collective agency, and social facts that are supervenient on (see Kim 1984; Kincaid 1986; List & Pettit 2011) or grounded by (see Epstein 2015; Schaffer 2016) such forms of collective intentionality and collective agency (including, e.g., collective emotions in addition to overt collective behavior (e.g., Salmela & Nagatsu 2016)).

The idea that philosophical research on social ontology can be understood as analogous to theoretical modeling and model-construction in science has not been previously advanced (to my knowledge). To be sure, some philosophers have used the informal term “model” to refer to their accounts of social ontology (e.g., Bratman 2014; Epstein 2015). However, they have not elaborated in detail what they mean by this choice of terminology, nor have they explicitly described their accounts as similar in structure and function to *scientific* models. This being said, I take the methodological status of any theoretical account to be in part a function of the way in which it is treated in the context of further theory-construction, rather than an upshot of its intrinsic features alone (Giere 2004). For example, Newton’s principles of mechanics were once treated as universal laws of nature. However, this conception of the nature and scope of Newtonian mechanics changed as physical science progressed in subsequent centuries. Thus, I take myself to be in a position to argue that at least some philosophical research on social ontology *can* be understood as theoretical modeling and model-construction, and at least relative to some theoretical concerns, it *ought* to be understood in these terms due to (among other things) the important type of continuity that this can be used to establish between social ontology and social scientific research practices. To alleviate concerns that my approach cannot accommodate the distinctive characteristics of philosophical investigation, I will carry out a detailed case study of a particular account of social ontology, which illustrates how the view that I will defend functions in practice. My work in this paper is an extension of themes that I have earlier developed in relation to different methodological strategies in the philosophy of mind and action (Sarkia 2021) and in the study of minimal forms of shared intentionality (Heinonen 2016).²

The decision to develop my naturalist approach to social ontology through a detailed case study may seem to require some further justification. This approach is motivated by the idea, which is itself familiar from naturalist philosophy of science, that applying general methodological ideas to particular cases may often yield more important insights than canvassing their consequences in greater generality, even when these ideas have broad applicability (Giere 1985; Kuhn 1962). The case study that I will focus on, Raimo Tuomela’s account of the *we*-perspective, is representative of what Francesco Guala (2007) has described as the *standard model of social ontology* (within analytic philosophy) by granting a particular form of performative and reflexive collective intentionality, which Tuomela describes as *we-mode collective acceptance*, a pivotal role in giving rise to central features of the social world. Thus, it can be used to identify general methodological

²My contribution to the latter topic was published under a different surname.

challenges, which many similar accounts of social ontology face, and widely applicable solutions to these challenges. Moreover, my case study can be used to pick out relevant differences to alternative approaches to social ontology, where the methodologically naturalist approach that I will defend may seem to have only limited applicability, such as those developed within the critical realist tradition (e.g., Archer 2017; Elder-Vass 2012; Lawson 2019; cf. Kaidesoja 2013). However, I will leave the precise scope of my approach, as well its relation to alternative approaches to social ontology, largely open to further investigation.

Here is a brief summary of what is to come. The next section discusses the idea of methodological naturalism and the (surprisingly controversial) standards that it imposes on philosophical practice. The third section introduces my case study of Raimo Tuomela's account of the we-perspective, focusing on the distinction between the (pro-group) I-mode and the we-mode, and Tuomela's three constitutive criteria of group reasons, collective commitment, and the collectivity condition for the we-mode. The fourth section turns to the roles that models and theories play in science, drawing primarily on the work of Ronald Giere and other philosophers in the naturalist tradition in philosophy of science. The fifth section argues in favor of understanding Tuomela's account of the we-perspective as a theoretical model of the social cognition and behavior of actual agents in the social world. The concluding section generalizes from my case study to some general features of a model-based and naturalist approach to social ontology.

Methodological naturalism and social ontology

The methodological approach that I will defend in this paper is intended to apply primarily to philosophical accounts of social ontology that are (or aspire to be) consistent with methodological naturalism (Giere 1985; Guala 2016; Kincaid 2012a; 2012b; Papineau 1993). To motivate the general thrust of this paper, I will in this section present a general challenge, the *naturalist's conundrum*, which I believe that all methodologically naturalist accounts of social ontology should be able to satisfy. To begin, I will discuss the general notion of methodological naturalism, and the constraints that it imposes on research practices in the field of social ontology. Then, I will argue that while many contemporary philosophers speak out in favor of methodological naturalism, the methodological standards that it imposes on philosophical practice are in fact far from trivial to satisfy. Thus, many self-professed naturalists find themselves failing to practice what they preach, or endorsing views that are inconsistent with methodological naturalism. To reach a balanced view of philosophical practice that is consistent with methodological naturalism, a new approach (one that is based on the idea of model-construction) is needed.

The practical implications of methodological naturalism for philosophical practice are deeply contested, although there is relatively broad agreement on the general tenets (and desirability) of methodological naturalism. For example, Harold Kincaid (2012b, 3) says that methodological naturalism “denies that there is something special about the social world that makes it unamenable to scientific investigation, and also denies that there is something special about philosophy that makes it independent or prior to the sciences in general and the social sciences in particular” (Kincaid 2012b, 3). The view that Kincaid expresses hardly seems controversial in the contemporary philosophical landscape, although it should be qualified by emphasizing that there are many different forms of scientific investigation, some of which may be closer to philosophical inquiry than others (of course, Kincaid is aware that methodological naturalism does not entail methodological monism, i.e. the idea that there is some unique method that is common to all of science). However, despite frequent lip service to the idea of methodological naturalism, it is disputable how many philosophers actually live up to the standards that it imposes on methodological practice. Giving voice to this skeptical view, Kincaid (2012a, 391) contends that “since naturalism can be interpreted weakly or strongly and since a great many philosophers would prefer not to be thought of as anti-naturalists, it is common to find contributors who would claim to be naturalists but at the same time provide arguments and analyses that are much closer to the anti-naturalist pole.”

Francesco Guala (2016, 44) echoes Kincaid’s concerns about the shallowness and easy dispensability of methodological commitments that are made under a naturalist veneer, when he describes as the most distinctive feature of contemporary philosophy of social science “the persistence and influence of anti-naturalism and its recent growth, especially in the area of ‘social ontology.’” Guala (2016) identifies (and ultimately rejects) two seasoned arguments for anti-naturalism in the philosophy of the social sciences: the reflexivity of human action (Hacking 1995; Soros 2013) and the constitutive dependence of social reality on representation and collective acceptance (Searle 2010; Thomasson 2003; Tuomela 2013). According to Guala (2016, 50–53), the reflexivity of human action poses little challenge for methodological naturalists, since reflexive expectations can be modeled by the tools of game theory (e.g., Elster 2007). Moreover, Hindriks and Guala (2015) argue that the constitutive dependence thesis can be debunked by conceiving of constitutive rules as regulative rules-in-equilibria, which are combined with theoretical terms that are introduced for reasons of cognitive economy.

The arguments that Guala (2016) expresses are unlikely to convince committed anti-naturalists about social ontology, since they often draw on some form of transcendental argumentation to justify the alleged priority of social ontology relative to social scientific investigation. For example, Searle (2009, 9) writes “where the social sciences are concerned, social ontology is

prior to methodology and theory... in the sense that unless you have a clear conception of the nature of the phenomena you are investigating, you are unlikely to develop the right methodology and the right theoretical apparatus for conducting the investigation.” In a different tradition, but expressing a similar spirit, Bhaskar (1979, 7) writes “philosophy is distinguished by the kinds of considerations and arguments that it employs... it considers... that world [studied by the sciences] from the standpoint of what can be established about it by a priori argument.” These types of transcendental claims seem inconsistent with Kincaid’s (2012b, 3) denial of the claim that “there is something special about philosophy that makes it independent or prior to the sciences in general and the social sciences in particular.” However, in line with Kincaid’s prediction of false naturalist pretensions, neither Bhaskar (1979) nor Searle (2010) consider themselves anti-naturalists.

Tuomela (2007, 2013) is another philosopher, whose work Guala (2016) identifies as being closer to the anti-naturalist pole in contemporary philosophy of social science. Nevertheless, Tuomela frequently speaks out in favor of scientific investigation of his “a posteriori” philosophical ideas, and characterizes his methodological approach as being “based on a science-friendly philosophical naturalism” (Tuomela 2013, 4–5). Tuomela does not seem to be referring to mere ontological naturalism here, or the idea that his account does not posit supernatural entities, since such a trivial commitment would hardly be worth mentioning—even granting that we now recognize as natural, some capacities that were previously (in the history of science) thought of as supernatural (Newton’s law of gravity serves as one useful example).³ While almost all contemporary philosophers claim to be ontological naturalists in the sense that their accounts are compatible with natural science, not all philosophers advocate methodological naturalism in the more interesting Kincaidian sense according to which “philosophy of social science and social science itself are continuous, philosophy has no special knowledge or tools that only it can provide, philosophy of social science has to be intimately connected with real social research... and social phenomena are susceptible to the broad methods of science in general” (Kincaid 2012a, 391). To assess whether Tuomela’s allegiance to naturalism is merely an instance of an anti-naturalist putting on a naturalist veneer, which Kincaid warned us about, we must look more closely at the types of theoretical goals that Tuomela formulates for his philosophical account and the ways in which he intends to pursue them. In the preface of a seminal book, Tuomela (2007) describes his theoretical goals as follows:

“This book systematically and analytically develops the basic group notions for the study of the social world, thinking and acting as a group member being a

³For a classic discussion of different varieties of naturalism, see Papineau (1993).

most central underlying notion here. In a way, this book can be said to present almost a philosophical “theory of everything” in the social world relying on the we-perspective.” (Tuomela 2007, vii–viii)

Tuomela seems to suggest two contrasting (and on the face of it, incompatible) roles for social ontology relative to empirical research in the cognitive and social sciences. The first role presents philosophers as constructing conceptual frameworks for empirical science, with little claim to empirical theorizing on their own account.⁴ In this sense, philosophy is presented as pursuing an analytic task that serves as a necessary precondition, as well as often a chronological precedent, to successful empirical research. Thus, Tuomela “systematically and analytically develops the basic group notions for the study of the social world.” Only with the aid of clear and unambiguous concepts can successful empirical research be pursued, the suggestion goes. The second role presents philosophy as being almost an equal partner and collaborator to the empirical sciences in the serious business of understanding and explaining the social world that we live in. Thus, Tuomela says that he wants to “present almost a philosophical ‘theory of everything’ in the social world” (sic). Tuomela does not tell us how a “philosophical theory” differs from a theory *simpliciter*. However, he clearly has some concrete and tangible contribution to our understanding of the social world in mind. This bolder conception of what philosophy aspires to contrasts with the disengaged observer and analyst of concepts expressed by the first role.

The two roles that Tuomela articulates for his philosophical account of the we-perspective might each be put under considerable pressure by contemporary naturalist philosophers of social science, such as Kincaid (2012a) and Guala (2016), who have frequently emphasized the need for philosophers to pay attention to the fine details of scientific practice. First, one might ask why philosophers should have any special warrant for legislating over the concepts that scientists use in their theories of the social world. Are social scientists not the best experts of their own trade, who hold the authority to decide which concepts to use in their theorizing? After all, the social sciences are not lacking in conceptually oriented researchers in their own right, who are often (even if not necessarily always) more familiar with the day-to-day research practices and topics of inquiry in their own disciplines than philosophers of social ontology (see, e.g., Coleman 1990; Goertz 2006; Joas & Knöbl 2009; Swedberg 2014). Second, one might be

⁴By saying that philosophers are not engaged in empirical theorizing *on their own account*, I mean that philosophers typically do not have the aim of generating *novel* empirical theories about the world of their own (even when their accounts are informed by empirical science, or their goal is to elucidate the conceptual or ontological pre-suppositions of current scientific research).

puzzled about how one could possibly make a serious pass at constructing a theory of the social world from the comfort of the proverbial philosophical armchair, without getting one's hands dirty in the laborious and often inconvenient grind of day-to-day empirical research. Would it not be best to leave theory-construction for properly trained scientists, while philosophers may concentrate on the activity of conceptual analysis, which they are more familiar with? Thus, Tuomela seems forced to choose between a task of conceptual analysis that he *can* perform, but which social scientists frequently choose to ignore (see Wettersten 2009; 2010), or a task of theory-construction, which philosophers seem less well suited to perform than practicing social scientists, who are experts in the local and specific research questions that their disciplines seek to address.

The above two-pronged challenge to Tuomela's theoretical aspirations can be framed in terms of a more general dilemma that I will call the *naturalist's conundrum*. The naturalist's conundrum emerges from two contrasting desiderata for social ontology, which are difficult (but not impossible, or so I will argue) to satisfy simultaneously: the challenge of relevance and the challenge of relative⁵ autonomy. The *challenge of relevance* is to show how research on social ontology can contribute to our overall, interdisciplinary understanding of the social world. To do so, it cannot consist in mere studious analysis of fabricated concepts, which often bear little resemblance either to the concepts that are used in ordinary language or to those that are used in social science. The *challenge of relative autonomy* is to show how social ontology can achieve this goal by the analytical tools that philosophers have at their disposal, without stepping over the boundaries of what philosophical methodology can realistically be expected to deliver. Despite recent forays into the field of experimental philosophy, most philosophers have not been trained to gather and interpret empirical data, nor to formulate and test counterfactuals by means of controlled experiments.⁶ Nor do most philosophers seem to have requisite subject knowledge (even if some do) to articulate broad synthesizing theories about the social world in a manner that would be both informative and

⁵The qualifier "relative" is there to indicate that there are of course numerous interconnections between social ontology and empirical social science. Indeed, one of the goals of this paper is identify important *continuities* between social ontology and social scientific research practices, rather than to argue that the two fields should be regarded as autonomous in any stronger sense than that they may perform different and complementary roles (moreover, there may be significant back-and-forth motion between these two roles, as I will later argue).

⁶The latter point may also be true of many forms of social theorizing that are relatively detached from empirical research (perhaps in the Habermasian, Luhmannian, or neo-Marxist traditions). One might argue that they face something akin to the challenges of relevance and relative autonomy that I have formulated for analytical social ontology. However, I will not pursue this issue further in this article.

insightful to practicing social scientists with a deep empirical and theoretical understanding of their subject matter. If philosophy is to flourish in the naturalistic and interdisciplinary age that we live in, one should be able to solve the challenge of relevance, without demanding that philosophers surrender their relative autonomy, and become empirical scientists, instead of philosophers.

This paper argues that the naturalist's conundrum can be solved, but doing so requires striking a fine balance between the relevance and relative autonomy of social ontology relative to the broader range of disciplines studying social reality. More specifically, I will argue that philosophers can (and sometimes do) use similar ways of studying the world as scientists, by constructing theoretical models of abstract systems that are presumed to be similar to some parts or aspects of reality. [Giere 1988; 2004; Godfrey-Smith 2006; Hausman 1992; Mäki 2009; Weisberg 2013; Wimsatt 2007; Godfrey-Smith 2005; Sarkia 2021](#). At the same time, I will also argue that the activity of model-construction needs to be kept distinct from the construction of theories and conceptual frameworks, which [Tuomela \(2007\)](#) originally formulated as the goals of his philosophical account. The former task is much too ambitious, since philosophers are often not the best placed to formulate theories of the social world. The latter task risks a philosophical arrogance of a different variety, as scientists do not generally speaking need philosophers to advise them about which concepts or methods to use in their theorizing. Standing between these two alternatives, I will argue that the activity of model-construction represents a modest methodological proposal for the study of social ontology, yet one that ensures philosophy a secure footing among the broader range of disciplines studying social reality.

Tuomela's account of the we-perspective

Tuomela's account of social ontology is representative of what [Guala \(2007\)](#) has described as the *standard model of social ontology* by granting a particular form of performative and reflexive collective intentionality an important role in giving rise to central features of the social world.⁷ Tuomela characterizes

⁷[Guala \(2007\)](#) distils the *standard model of social ontology* in terms of the ideas of performativity, reflexivity and collective intentionality. *Performativity* refers to certain types of social facts being made the case by asserting them to be (or acting as if they were) the case, as when a priest proclaims "I hereby declare you man and wife." *Reflexivity* refers to the idea that "social entities are constituted by *beliefs about beliefs*" [Guala \(2007, 961\)](#)—for example, the convention of driving on the left or right hand side of the road is in part constituted by the expectation that others will drive on the right or left hand side of the road. *Collective intentionality* involves the idea that certain types of collective attitudes are not mere aggregates of individual attitudes, for example, "when the Tory Party leader says that 'we believe that Britain should not enter the European Monetary Union'... it is the party's 'official policy', so to speak, to have certain collective beliefs and desires" ([Guala 2007, 963](#)).

the relevant form of collective intentionality in terms of what he calls the *we-perspective*. According to Tuomela (2007, 52–53), there are both I-mode and we-mode forms of the we-perspective. The *plain I-mode we-perspective* has to do with functioning as a private person in a group context in order to satisfy one's own interests and preferences. The *pro-group I-mode we-perspective* has to do with functioning as a private person in a group context in order to satisfy the interests and preferences of a social group. The *we-mode we-perspective* has to do with functioning as a group member in a group context in order to satisfy the interests and preferences of a social group that one strongly identifies with. Tuomela (2007, 3) characterizes the we-mode we-perspective as the “full we-perspective,” which is of crucial importance in explaining the existence of social facts and social institutions. Given this pivotal role in his philosophical theory, I will follow Tuomela's example in concentrating on the we-mode we-perspective, while relying on a partly pre-analytic conception of the I-mode we-perspective.

Tuomela (2013, Ch. 2) elucidates the we-mode we-perspective by appeal to three criterial features, which he calls group reasons, collective commitment, and the collectivity condition. According to *group reasons*, the members of a we-mode group must set aside their private views and preferences in the group context and take their reasons for action from what the group believes, desires, and intends (2013, 38). The group members are required to “give part of their natural authority to act to the group” (Tuomela 2013, 116) and take the *ethos* of the group (involving its constitutive goals and attitudes) as providing authoritative reasons to act within the group context. For example, the members of a departmental hiring committee might be required to set aside their private preferences about who they would like to work with in order to choose the person that is best suited for a vacant position. According to *collective commitment*, the members must be committed “as a group” to perform as the satisfaction of the group's attitudes requires them to perform (Tuomela 2013, 43). They must also be *group-socially committed* to one another to perform their parts of what constitutes the group's performance of the tasks in question (Tuomela 2013, 252). This can include filling in for other members, who are unable to perform their parts, instructing or pressuring others to perform their parts, and so forth. According to *the collectivity condition*, the group's attitudes can, on “conceptually necessary” grounds, only be satisfied for the group members together, rather than for each of them separately (Tuomela 2013, 41).

Tuomela (2013, 39) argues that the we-mode we-perspective involves both a change of agency and a change of mode of reasoning relative to the I-mode we-perspective. To begin with the change of agency, Tuomela (2013, 5) argues that the *intentional agent* of collective action in the we-mode is the group agent that the individuals together constitute, even if “ontologically, in the causal realm, individuals are the only initiating ‘causal motors’.” Tuomela

(2013, 15) says that the attitudes of the group agent are *collectively constructed* through the performative and reflexive collective acceptance of attitudes by the members *for the group*. This gives rise to the *ethos* of the group, a subset of its *realm of concern*, which provides the group members with authoritative directives about how to act *as group members* (Tuomela 2013, 27). To move on to the change in mode of reasoning, Tuomela (2013, 7, 181) says that individuals acting in the we-mode must *we-reason* from the point of view of what is best for the group as a whole, rather than from the point of view of what is best for them as individuals. While I-mode social groups are susceptible to disintegration because of free-riding and weakness of will, Tuomela (2013, 40) says that we-mode group members are required to “be in the same boat” and to obey the three musketeer’s principle of “all for one and one for all.” In joint work with Raul Hakli and Kaarlo Miller (Hakli et al. 2010), Tuomela argues that we-mode we-reasoning may decrease the amount of equilibria (relative to the pro-group I-mode) in game-theoretically modeled strategic interactions, but it may not increase the number of such equilibria (see also Tuomela 2013, 212). They conclude that we-mode reasoning can lead to Pareto-optimal solutions in coordination games, such as the Hi-Lo, and prevent social dilemmas, such as the Prisoner’s Dilemma (PD), from occurring (Hakli et al. 2010, 292; Tuomela 2013, 193–194):

Hakli et al. (2010, 300–306) argue that standard individualistic approaches to game theory cannot guarantee Pareto-optimal solutions to these problems (i.e., to reach an outcome where no player can be made better off without making some player worse off), without the introduction of additional assumptions, such as the controversial principle of *pay-off dominance* (Harsanyi & Selten 1988, 355–357; see also Bardsley 2007). This is because they are based on best response-reasoning, and what counts as the best response to the actions of the other player depends on what she chooses. For example, in the Hi-Lo, L is the best response to L, although HH gives a better outcome. Hakli et al. (2010) suggest that this challenge can be overcome by framing the Hi-Lo as a decision problem for a we-mode collective agent, and that the Prisoner’s Dilemma can be transformed into a Hi-Lo by way of preference transformation. They draw on Bacharach’s (1999) mathematical model of *unreliable team interactions* to argue that this combination of *agency transformation* and *preference transformation* yields different and Pareto-optimal results compared to an alternative approach that is based only on preference transformation (which is also compatible with pro-group I-mode reasoning). However, the precise degree of overlap between Tuomela’s philosophical framework and Bacharach’s mathematical model remains somewhat elusive, as Hakli et al. (2010, 311) grant that

“...our we-mode approach has not been formalized, partly because of its conceptual richness that goes beyond the narrow framework of game theory.

| | | |
|---------|------|------|
| (Hi-Lo) | H | L |
| H | 3, 3 | 0, 0 |
| L | 0, 0 | 1, 1 |

| | | |
|------|------|--------|
| (PD) | C | D |
| C | 3, 3 | 1, 4 |
| D | 4, 1 | -1, -1 |

Even our central criteria of the we-mode, viz. group reason, collectivity condition and collective commitment, do not seem to be game-theoretically formalizable...”.

My suggestion is that instead of taking Bacharach’s mathematical model as providing direct support to the validity of Tuomela’s conceptual distinction between the (pro-group) I-mode and the we-mode (as Tuomela sometimes seems inclined to do), we should treat [Hakli et al.’s \(2010\)](#) discussion as an attempt to establish a *model-to-model mapping*, where certain conceptual features of Tuomela’s philosophical framework are argued to perform similar functional roles as corresponding features of Bacharach’s mathematical model. Consistently with this interpretation, [Hakli et al. \(2010, 301\)](#) propose a set of “assumptions that partially translate our framework into the game-theoretic one,” where “group utility in Bacharach’s account plays the role of group reason, the team-reasoning pattern corresponds to the collectivity condition, and his ‘group identification’ to the notion of collective commitment.” [Hakli et al. \(2010\)](#) go on to argue that Tuomela’s philosophical framework augments Bacharach’s mathematical model by making it possible to pay attention to the stages of information sharing and negotiation that precede the formation of a collective preference matrix, in addition to we-reasoning within a game-theoretically modeled strategic interaction. However, the somewhat stipulative nature of their enterprise is highlighted by the fact that Bacharach’s mathematical results have been recruited by other philosophers to support accounts of collective intentionality that are at odds with Tuomela’s strongly antireductionist account (e.g., [Bardsley 2007](#); [Gold & Sugden 2007](#); [Pacherie 2011](#)).⁸

⁸For more discussion of the relation between Tuomela’s philosophical account of the we-perspective and game theory, see [Gold & Sugden \(2007\)](#); [Paternotte \(2011\)](#); [Tuomela \(2009\)](#).

To return to the topic of methodological naturalism, Tuomela's own assessment of the social scientific usefulness of his philosophical account is strikingly optimistic, and he even goes so far as to claim that "the social world can be adequately understood and rationally explained only with the help of we-mode concepts in addition to I-mode concepts" (Tuomela 2013, 7). However, many naturalist critics of Tuomela's research have remained unconvinced about the indispensability of his conceptual framework for social science. We have already seen that Guala (2016) describes Tuomela as a prominent proponent of *anti*-naturalism in the philosophy of social science, despite Tuomela's explicit assertions to the contrary. In even harsher terms, Wettersten (2009, 533) describes Tuomela's philosophical framework as essentialist and largely irrelevant to social science, because social scientific frameworks "have to prove themselves by their ability to lead to interesting empirical theories and not by their analytic 'proof'." In response to Tuomela's contention that the we-mode leads to the different action predictions than the I-mode, Wettersten (2010, 521) retorts that "this would only begin to be explanatory if we could predict when humans behave one way and when in another way and explain why they do so." Although Wettersten formulates his criticism in unnecessarily polemical terms, I believe that he is posing a legitimate challenge for Tuomela to elaborate in more detail how his abstract and idealized framework can be used by scientists for understanding and explaining particular phenomena in the social world that we live in. Before providing my own perspective on this challenge, I will digress on the roles that models and theories play in scientific practice.

Theories and models in science

The terminology of conceptual frameworks, models, and theories is used in a relatively relaxed manner by many contemporary philosophers of social ontology to refer to closely related methodological ideas, but without necessarily elaborating what these methodological ideas are or the types of empirical commitments that they entail. However, despite their seemingly innocuous character, these terminological choices arguably bring about important repercussions for how the empirical commitments of philosophical accounts of social ontology ought to be understood, and what types of normative standards they ought to be evaluated by. While almost any collection of interconnected concepts can be described as a conceptual framework, theories are generally speaking taken to involve substantive empirical commitments about the way the world is, and theoretical models have often been described as ontologically non-committal in the sense that they make empirical claims about the world only by the mediation of *theoretical hypotheses* (Giere 1988) or *ontological construals* (Godfrey-Smith 2005; 2006) connecting those models to the world in specific degrees and respects. To

serve as the basis for further discussion, I will in this section introduce a distinction between scientific models and theories, which is familiar from contemporary naturalist philosophy of science (e.g., [Giere 1988](#); [2004](#); [Godfrey-Smith 2006](#); [Hausman 1992](#); [Mäki 2009](#); [Weisberg 2013](#); [Wimsatt 2007](#)).

To begin, we may note that the notion of “theory” in ordinary language is subject to numerous ambiguities, as it is sometimes taken to bear an assumption of arbitrariness or uncertainty about whether the theory under consideration is in fact true or not. This sense of “theory” is in use when creationists claim that natural selection is “only a theory,” and when journalists refer to the outlandish scenarios concerning the world order that have been conceived by “conspiracy theorists.” By contrast, scientific theories are at least in their mature phase typically taken to be supported by large amounts of empirical evidence, which have been arrived at by systematic data-gathering methods, ideally from procedurally distinct but evidentially convergent sources ([Boyd 1991](#); [Kaidesoja 2013](#); [Niiniluoto 1999](#)). Thus, mature scientific theories are on a more secure epistemic and evidential footing than theories in the colloquial sense of the term. Here, I will abstract away from many of the details of the scientific processes, which make theories such reliable guides to the causal structure of the world ([Hacking 1983](#); [Woodward 2003](#)). Rather, I will concentrate on one central feature that is arguably central to scientific theories in any relevant sense of the term: they make empirical claims about the world, which can be either true or false ([Godfrey-Smith 2003](#), 176).⁹

Theories in the present sense are not the only kinds of epistemic constructs that arise in the context of scientific practice. Scientists also conduct experiments, formulate generalizations, carry out simulations, construct models, and much more ([Wimsatt 2007](#)). Many of these activities and their proximate epistemic products (e.g., the data from an experiment) may serve as important precursors or constituents of scientific theories. However, they may also be carried out at a certain remove from the construction of theories in the sense of systematically interconnected and evidentially supported claims about what the world is like. Nor need their proximate epistemic outputs always lead to or be eventually subsumed as proper parts of scientific theories. Multiple experiments are typically required in order to arrive at unambiguous data ([Hacking 1983](#)), some generalizations are uninformative or trivially true

⁹This claim might be disputed on the grounds that use of the term “theory” in the social sciences is highly permissive, encompassing such things as conceptual frameworks, normative approaches, and overall worldviews in addition to explanatory theories ([Abend 2008](#)). The usefulness of a concept that encompasses so many different types of things can reasonably be doubted, making a somewhat more regimented conception of the model/theory distinction in this section necessary. See also footnote 12 below.

(Cummins 2000), and the proximate results of many simple “toy models” and simulations only seem to teach us about possibilities, not about the way the world actually is (Fumagalli 2016; Grüne-Yanoff 2009). Given that social ontology rarely involves conducting experiments or carrying out simulations (however, see Michael & Szigeti 2019), I will not discuss these scientific practices in more detail in this paper. Rather, I will concentrate on the roles that models and theories play in scientific practice.

Giere (1988; 2004) has presented an insightful discussion of the relation between scientific theories and scientific models, which I will anchor my discussion to. According to Giere (1988, 79), *theoretical models* are abstract objects, which have all and only the properties that are attributed to them by a model-description.¹⁰ A *model-description* may come in many different forms—for example, in the form of a mathematical equation, a linguistic description, a picture, or a graph. Many different model-descriptions may pick out one and the same theoretical model. For example, the sentence “Demand is a function of price” and the mathematical formula “ $D = f(p)$ ” can be used to pick out a relation of interdependence between price and demand (although these statements do not yet involve detailed empirical commitments about the products for which this relation holds, the elasticity of the demand function, or how the relation between price and demand is influenced by the budget constraints of consumers). *Theoretical hypotheses* are linguistic statements, which connect such models to the world in various degrees and respects (Giere 1988, 85). For example, a theoretical hypothesis mapping the relation between price and demand to a particular market might specify that demand is an inverse function of price in the market for used cars, while demand is a direct function of price (up to a certain threshold) in the market for luxury goods. The same theoretical model can often be used to represent a wide range of different target systems by connecting it to different sets of theoretical hypotheses. Thus, there is an important pragmatic dimension to the activity of theoretical modeling, which can be captured by the following formula¹¹

S uses *X* to represent *W* for purposes *P* (Giere 2004, 743)

To summarize, what sets apart theoretical models from other epistemic constructs that are used in scientific practice is that they make claims about the world only *indirectly*, by the mediation of theoretical hypotheses, rather than

¹⁰There are also concrete models in science, such as model organisms and scale models (Downes 2011). However, I will only be concerned in this paper with non-concrete, theoretical models.

¹¹Uskali Mäki (2009, 32) formulates a more elaborate account of how theoretical models are used to represent the world. According to his formulation, “Agent A / uses object M as / a representative of some target system R / for purpose P, / addressing audience E, / prompting genuine issues of resemblance to arise; / and applies commentary C to identify and align these components.”

by directly describing law-like regularities or concrete particulars in the world (Godfrey-Smith 2006; Weisberg 2013). Indeed, Giere (1988, 93) goes so far as to suggest that theoretical models do not as such involve any empirical claims about the world at all, in the absence of theoretical hypotheses, and therefore cannot be true or false on their own. However, theoretical models can be subsumed as proper parts or constituents of scientific theories, when they have been complemented by theoretical hypotheses, and suitable additional conditions, such as systematicity and evidential support, have been satisfied (Giere 1988). Theoretical models can also be worth investigating in their own right, independently of the formulation of any theoretical hypotheses at all (Weisberg 2013; Wimsatt 2007). This may teach us about possibilities, or facilitate the understanding of a complex system that we know little about. Thus, the construction of theoretical models can be understood as a distinctive type of epistemic activity, which may, but need not lead to the formulation of scientific theories in the sense of evidentially supported and systematically interconnected sets of empirical claims about the world.

The account of scientific models and theories put forth by Giere and his followers is not the only one that is available in contemporary philosophy of science (cf. Craver 2002; Suppe 1989; Suarez 2004; Suppes 1960; van Fraassen 1987). The reason why I have chosen to draw on Giere's account in putting forth my model-based and naturalist approach to social ontology is that Giere's account has itself been developed in a methodologically naturalist manner with close attention to the details of scientific practice (see Giere 1985; 1988). Close relatives of Giere's account of theoretical models have been successfully applied to numerous different scientific disciplines, from physics (Giere 1988; 2004) to economics (Hausman 1992; Mäki 2009; Sugden 2000) and evolutionary biology (Godfrey-Smith 2006; Weisberg 2013). While some authors (e.g., Abend 2008) have claimed that the notions of "model" and "theory" in the social sciences are hopelessly ambiguous, the success of a Gierean approach in other scientific fields suggests that it is a promising candidate for capturing (and regimenting) certain important methodological practices in the social sciences as well.¹² This being said, I do not think that a Gierean approach can accommodate all forms of "theorizing" (in a broad sense) within the social sciences, nor should all of social science be understood to be concerned with theoretical modeling and model-construction at

¹² Abend (2008) distinguishes as much as seven different senses of the term "theory" in sociology, from a "logically connected system of general propositions, which establishes a connection between two or more variables" (Abend 2008, 177) or "an explanation of a particular social phenomenon" (Abend 2008, 178) to an "overall perspective from which one sees and interprets the world" (Abend 2008, 178) and "accounts that have a fundamental normative component... which usually reject the fact/value dichotomy" (Abend 2008, 180).

all (cf. Brante 2010). Rather, many social sciences are concerned with empirical investigation of the world in a much more direct sense, which does not involve *indirect* representation of the world by the mediation of a model (see Godfrey-Smith 2006; Weisberg 2013). This claim to local as opposed to global applicability (see Kincaid 2021) is also one important respect in which my model-based approach to social ontology differs from many anti-naturalist approaches, which view social ontology as necessarily preceding all of social science in a conceptual or logical sense (see Bhaskar 1979; Elder-Vass 2012; Lawson 2019; Searle 2010).

The we-perspective as a theoretical model

The goal of this section is to use my case study of Raimo Tuomela's philosophical account of the we-perspective to elucidate one way in which social ontology can contribute to theory-formation in the social sciences, without claiming that it is the only possible way, nor that all of social science depends on social ontology in the sense that I will explicate. With respect to the content of my case study, I will argue in favor of a revisionary interpretation of the we-perspective as a theoretical model of the social cognition and behavior of actual agents in the social world, which contrasts with Tuomela's own description of his account as a "philosophical theory of the social world." However, I will not attempt to show that the we-perspective *must* be understood as a theoretical model, nor that it *cannot* be understood as a philosophical theory, as there may be other senses of the notions of "theory" and "model" apart from the Gierean notions that I have considered. To begin, consider how Tuomela describes the empirical commitments of his account:

"...[the] major claims of the theory are at least indirectly and partially empirically testable...I will argue that neither individualism nor collectivism is capable of expressing the whole truth, so to speak: both individualistic (I-mode) and collectivistic (especially, we-mode) concepts and mental states are needed to explain, for example, social action and social institutions." (2013, 4)

My suggestion is that Giere's account of theoretical models can be used to provide one plausible perspective on what it means for Tuomela's account to be *indirectly* and *partially* testable (Tuomela does not elaborate what he means by these formulations). Remember that according to Giere, theoretical models do not involve any empirical commitments about the world on their own. Rather, they take on ontological commitments only when they are associated with theoretical hypotheses, which assert that some part or aspect of the world is similar to the model in relevant degrees and respects. To take up Giere's (1999, 95) example, a mathematical model of a simple harmonic oscillator can be used to represent the motion of a real mass on a string only when numerous

additional complications have been taken into account, such as the effect of air resistance or the angle of the swing. By formulating different theoretical hypotheses, the same mathematical model can be used to represent very different types of physical systems, such as bouncing springs or vibrating strings (Giere 1999, 110–111). Thus the representational roles that the model serves are mediated by the interests and purposes of the modeler, resulting in *indirect* representation. And typically, only a subset of the elements in the model, or only some of their properties, are mapped onto the world, resulting in *partial* representation. Accordingly, Giere's account of theoretical models can be used to unpack the type of indirect and partial representation that Tuomela himself highlights as central for how the ontological commitments of his philosophical account are to be understood.

The indirect nature of model-based representation can also be used to accommodate Tuomela's ambivalence about how literally (in a psychological sense) his distinction between the I-mode and the we-mode we-perspectives should be taken. Consistently with an instrumentalist approach to science, Tuomela (e.g., 2013, 197) occasionally describes the I-mode and the we-mode as "ideal types" and defends his top-down approach to social ontology by saying that "in the case of large and hierarchically structured groups, group-level descriptions typically involve pragmatic simplicity and are instrumentally useful" (Tuomela 2013, 90). However, Tuomela equally often seems tempted by a psychologically realistic interpretation of the we-perspective. For example, Tuomela (2007, 57) writes that "...it would seem that we-mode actions and attitudes do not always require I-mode intermediaries but can, so to speak, be mapped directly onto neural states and events. There seem to be at least no in-principle barriers to the we-mode "going all the way down"—not only conceptually but also in this factual psychological sense." This strongly realist interpretation of the we-perspective is at odds with the rather instrumentalist alternative that he seems to favor on other occasions. This makes it extremely difficult to evaluate what precisely are the truth-makers for Tuomela's (2007, viii) "philosophical theory of the social world."

The ontologically non-committal nature of theoretical models can be used to avert this challenge. Let us call an *ontological construal* of a theoretical model a collection of theoretical hypotheses, which specify in what ways and to what extent (if at all) each element in the model is similar to some part or aspect of the real world (Godfrey-Smith 2005). Often, there may be many different ontological construals of one and the same theoretical model. Accordingly, Tuomela's account of the we-perspective *might* be interpreted as a relatively literal description of the psychological states and dispositions of actual human agents, if ontologically construed in association with one set of theoretical hypotheses. *Or* it might be interpreted as an instrumental framework for predicting some aspects of their collective behavior, if ontologically construed in association with another set of theoretical hypotheses.

Thus, even if it were to turn out that individuals do not have “we-mode mental states with specific neural bases” (Tuomela 2013, 5), one could not yet infer that Tuomela’s account is a poor model, since it might still serve some useful theoretical roles on a greater level of abstraction.

The indirect nature of model-based representation can also be used to accommodate the numerous abstractions and idealizations that Tuomela’s account of the we-perspective involves.¹³ Generally speaking, Tuomela assumes that individuals function in one mode at most relative to each type of activity at a time, and that functioning in one mode precludes functioning in the other. For example, Tuomela (2013, 38) writes that “paradigmatic or genuine we-mode functioning excludes conflicting individualistic motives: An ideal we-moder can think and act only for group-centered motives, regardless of whether they conflict with individualistic motives.” When we-mode group members fail to satisfy this requirement, they are assumed to “lapse into intentionally satisfying their private interests”—in other words, to function as I-moders (Tuomela 2013, 199). Tuomela (2013, 156) even speaks of a “switch” between the I-mode and the we-mode. However, it seems like a more plausible psychological hypothesis that the I-mode and the we-mode we-perspectives should exist on a continuum with one another, if the distinction between the two modes is regarded as psychologically real in the first place: one would not expect to find a categorical switch between the I-mode and the we-mode in the human brain, although human beings are sensitive to being in a group context (see Brewer 2003), and their attitudes and actions are influenced by the presence of other individuals (see Hari et al. 2015).¹⁴

The features of abstraction and idealization as such are neither positive nor negative epistemic attributes in a theoretical model, but features that need to be evaluated with respect to the representational or inferential goals that the model is supposed to achieve. Given that the world is complex and scientific models need to be sufficiently tractable in order to be theoretically feasible, abstraction and idealization are both ubiquitous and inevitable features of many scientific models (Teller 2001). By contrast, scientific theories should at least ideally aspire to do away with such abstractions and idealizations in order to arrive at a true description of the world as it is (Boyd 1991; Kaidesoja 2013; Niimiluoto 1999). To be sure, one might describe Tuomela’s account of the we-perspective as a theory that is “approximately true” (Niimiluoto 1999) or true in a “*ceteris paribus*”—sense (Pietroski & Rey 1995). However, this would cast the numerous abstractions and idealizations that his account involves as

¹³For the notions of abstraction and idealization, see Thomson-Jones (2005, Weisberg (2013).

¹⁴For an interesting discussion of whether cognitive science supports the idea of an irreducible “we-mode” in the brain, see Gallotti & Frith (2013). For a critical review of some of the data that they refer to, see Dolk & Prinz (2016).

problematic features to be dispensed with, rather than as *constructive* features that have been deliberately introduced in order to increase the epistemic tractability and simplicity of his theoretical model at the expense of its realism. While truth provides the overriding epistemic norms for scientific theories, theoretical models can aspire to satisfy many different non-representational desiderata, such as simplicity, elegance, fertility in generating new hypotheses, and epistemic tractability (Wimsatt 2007). Arguably, many of these desiderata have at best a tenuous relationship to truth.

The numerous qualifications that Tuomela makes in informal discussion about the categorical distinction between the I-mode and the we-mode can be used to further support a Gierean interpretation of the we-perspective. For example, Tuomela (2013, 122) says that “to make the basic account more realistic, private reasons may play a role at the expense of group reasons, although strictly speaking this goes against acting as a group member.” However, Tuomela does not provide a full account of how I-mode reasons can be weighed against we-mode reasons, or vice versa. Thus, the formal contrast between the two modes does seem to be categorical in character. Given this categorical basis, it seems best to treat the possibility of a continuum between the I-mode and the we-mode as part of Tuomela’s *extrinsic commentary* on his categorical account of the we-perspective, rather than an intrinsic feature of his philosophical account as such. Such commentary is commonly associated with theoretical models in science. Given that any model provides at best a partial representation of reality (Teller 2001), almost all scientific models involve a commentary, by which the modeler makes explicit the representational goals that the model is intended to achieve, as well as the possible epistemic limitations that it may have (Mäki 2009).

To summarize, I have argued above that Tuomela’s account of the I-mode and the we-mode is best understood as a theoretical model of the social cognition and behavior of actual agents in the social world, rather than as “a philosophical theory of the social world” (as Tuomela himself describes it). Nevertheless, a model-based interpretation of the we-perspective also leaves open one avenue for Tuomela to pursue some of his grander theoretical ambitions. This is because a theoretical model of the we-perspective might yet serve as one important *constituent* of an empirical theory of the social world, if it is supplemented by appropriate theoretical hypotheses, and suitable additional conditions, such as systematicity and evidential support have been satisfied (Giere 1988; 2004). However, such theoretical hypotheses are arguably best formulated and later validated by empirical researchers in the cognitive and social sciences, possibly in collaboration with philosophers of social ontology, rather than by philosophers in isolation (see, e.g., Becchio & Bertone 2003; Gallotti & Frith 2013; Rakoczy 2017; Tomasello 2019). In this sense, a Gierean interpretation of the we-perspective can accommodate many of Tuomela’s grander theoretical aspirations (i.e., to construct a “philosophical

theory of the social world”), while settling for a somewhat more modest proposal of what it actually achieves (i.e., a theoretical model that may someday serve as one important constituent of a theory of the social world). Moreover, it provides an adequate response to Tuomela’s naturalist critics about how his highly stylized philosophical framework can be used by cognitive and social scientists for understanding and explaining particular phenomena in the social world that we live in.

A model-based solution to the Naturalist’s conundrum

This paper began with the introduction of the naturalist’s conundrum, or the difficulty of striking a reasonable balance between the relevance and relative autonomy of philosophical research on social ontology relative to the broader range of disciplines studying social reality. The challenge of relevance was to show how social ontology can contribute to our overall, interdisciplinary understanding of the social world that we live in. The challenge of relative autonomy was to show how this goal can be achieved by the analytical tools that philosophers have at their disposal, without stepping over the boundaries of what philosophical methodology can reasonably be expected to deliver. If one fails in the first task, one is threatened with irrelevance outside the domain of strictly philosophical investigation. If one fails in the second task, one is threatened with a loss of philosophical identity. To abandon one challenge in order to focus exclusively on the other would go against the spirit of methodological naturalism, according to which “philosophy of social science and social science itself are continuous, philosophy has no special knowledge or tools that only it can provide, philosophy of social science has to be intimately connected with real social research, philosophy of science can and should try to be of use to social scientists themselves, and social phenomena are susceptible to the broad methods of science in general” (Kincaid 2012a, 391).

This paper has addressed the naturalist’s conundrum through a detailed case study of Raimo Tuomela’s account of the (we-mode) we-perspective. While Tuomela (2007, viii) originally described his account of the we-perspective as a “philosophical theory of the social world,” I have argued that Tuomela’s account is best understood as a theoretical model of social cognition and behavior in a Gierean sense that is compatible with, but less demanding than Tuomela’s own description of it as a theory of the social world. Given a model-based interpretation of the we-perspective, I have argued that Tuomela can proceed in a relatively autonomous manner in articulating the main conceptual features of the we-perspective, while remaining largely non-committal (at the stage of model-construction) how realistic or truthful his highly stylized philosophical account may be. Moreover, he can do so with confidence that his abstract and idealized philosophical framework can

at least in principle be regarded as relevant for cognitive and social scientists, who may use it to formulate theoretical hypotheses about whether individuals do indeed have mental states in two distinct modes or whether they behave as predicted by Tuomela's abstract philosophical framework (see, e.g., [Becchio & Bertone 2003](#); [Gallotti & Frith 2013](#); [Tomasello 2019](#)).

Tuomela's naturalistic aspirations are symptomatic of a wider shift of sentiment in contemporary social ontology, even if we brush aside his talk of providing a "theory of everything" in the social world" as the unintentional slip of an over-zealous academic tongue.¹⁵ This sounds like an old-fashioned aspiration at a time when the trend in many social sciences has been towards more local and mechanistic explanations, which stand in contrast to the grand unifying theories of the past ([Hedström & Ylikoski 2010](#)). However, regardless of the precise scope of their accounts, many contemporary philosophers of social ontology would clearly like to make some type of concrete and tangible contribution to our overall understanding of the social world. For example, in the introduction to her book "Joint Commitment: How we Make the Social World," [Gilbert \(2013, 3\)](#) says that while the proximate goal of her Weberian-interpretivist account of *joint commitment* is to understand "the ideas in terms of which individuals would couch their goals and their reasons," the ultimate goal of her account is not to analyze everyday language but to "perspicuously to describe the phenomenon to which the relevant everyday statements refer." In his equally ambitiously named "Making the Social World: the Structure of Human Civilization," [Searle \(2010, 12–13\)](#) draws on his earlier philosophical work on speech act theory to defend the claim that "all of institutional reality, and therefore... all of human civilization, is created by speech acts that have the same logical form as Declarations."

These philosophers (and many others) *practice* conceptual analysis, but ultimately, they see their accounts as contributing to much more than elucidation of the *meanings* of colloquial or technically defined concepts. However, for all that they have said, the precise contribution that they make to our overall, interdisciplinary understanding of the social world remains difficult to pin down. The goal of paper has been to fill this lacuna in the methodological foundations of contemporary social ontology (as practiced by analytic philosophers) by framing it as theoretical modeling and model-construction in a sense that is familiar from contemporary naturalistic philosophy of science. To the extent that my approach differs from how philosophers have previously understood the nature of their theoretical enterprise, my paper can be understood in a revisionary spirit—I have argued that despite possible appearances to the contrary, philosophical research on social

¹⁵Tuomela (in correspondence) qualifies this assertion by indicating that the passage is intended "to be understood in the sense that the theory to be created concerns (the part of) the social world that relies on the we-perspective."

ontology can (at least in some circumstances) be understood in terms of the Gierean approach that I have advocated. At the same time, I do not want to claim all-encompassing scope to my approach, since there are also alternative approaches to social ontology, which do not agree with the basic tenets of my approach or the methodologically naturalist spirit that motivates it. For example, philosophers in the critical realist tradition (e.g., Archer 2017; Bhaskar 1979; Elder-Vass 2012; Lawson 2019; cf. Kaidesoja 2013) have often drawn on some form of transcendental reasoning to argue that social ontology uses *different research methods* and addresses *different topics of inquiry* than other parts of social science. While I am inclined to view such projects with skepticism, my goal in this paper has not been to criticize these approaches, but to offer an alternative to them by identifying important continuities between social scientific methodology and some central approaches to social ontology.

To generalize beyond the confines of my case study, a model-based approach to social ontology solves the naturalist's conundrum through its reliance on a distinctively two-stage picture of scientific theorizing (Godfrey-Smith 2006; Weisberg 2013). In the first stage of model-based science, a suitable model is constructed for studying some phenomenon of interest, possibly with only a vague or imprecise conception of how the model may help in understanding the phenomenon in question. In the second stage of model-based science, the model is applied to some real phenomenon by formulating and testing theoretical hypotheses about how the model may be used to represent the phenomenon in question (Giere 1988; 2004). Such theoretical hypotheses may be only partial, relating to the predictions that the model yields, or some other subset of its features—or they may be more complete, involving a mapping from each element in the model to some part or aspect of the world. Of course, there may in practice be considerable back-and-forth motion between these two stages, as successive generations of a theoretical model are developed and calibrated in order to capture the details of a particular phenomenon of interest (Wimsatt 2007). However, even if the construction of theoretical models does not necessarily precede empirical investigation in the way in which philosophers have traditionally thought of social ontology as being (conceptually or logically) prior to social science, the analytic distinction between the two stages of the model-based enterprise can still guarantee a degree of relative autonomy to social ontology. Moreover, although theoretical models are also constructed in other disciplines than philosophy, the interest that some philosophical accounts of social ontology have sparked among empirical researchers during recent decades indicates that such models may be viewed as relevant not only in principle, but also in practice to current empirical science (see, e.g., Gallotti & Frith 2013; Rakoczy 2017; Tomasello 2019).

To draw together the results of the preceding discussion, the two-stage picture of model-based science can be used to ground a feasible division of labor between philosophers of social ontology and empirical researchers in the cognitive and social sciences, which does not depend on transcendental reasoning or any other dubious form of a priori investigation of the kind that many anti-naturalist philosophers of social ontology have laid claim to (cf. [Bhaskar 1979](#); [Elder-Vass 2012](#); [Lawson 2019](#); [Searle 2010](#)). This is because the two stages of model-based science can in principle be performed relatively independently of one another—at different times, by different individuals, and sometimes even by entirely different scientific disciplines ([Godfrey-Smith 2006](#); [Mäki 2009](#); [Weisberg 2013](#); [Wimsatt 2007](#)). Some parts of science, such as certain branches of theoretical economics, seem to be almost exclusively concerned with the construction of theoretical models ([Hausman 1992](#); [Mäki 2009](#); [Rodrik 2015](#); [Sugden 2000](#)). Other parts of science are primarily concerned with the formulation of theoretical hypotheses and the description of empirical phenomena ([Bechtel & Richardson 2010](#); [Cummins 2000](#); [Weisberg 2007](#)). Given the analytical and logical tools that they have at their disposal, the construction of theoretical models would seem like a task that philosophers of social ontology should be both eager and well-equipped to perform. However, a natural corollary of this view is that they are not engaged in the direct investigation of empirical truths about the world at all. Thus, the construction of theories would not be an appropriate description of the type of theoretical enterprise that they are engaged in. [Grüne-Yanoff and Mäki 2014](#); [Knoblich et al. 2011](#); [Tuomela 2009](#).

Acknowledgments

This paper has benefited from valuable comments from Prof Raimo Tuomela, Dr Tuukka Kaidesoja, Prof Michiru Nagatsu, Dr Raul Hakli, and an anonymous referee for *Philosophy of the Social Sciences*.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Matti Sarkia  <https://orcid.org/0000-0001-6523-9993>

References

- Abend, G. 2008. "The Meaning of 'Theory'." *Sociological Theory* 26 (2): 173-199.
- Archer, M. 2017. *Structure, Culture and Agency: Selected Papers of Margaret Archer*. London: Routledge.
- Bacharach, M. 1999. "Interactive Team Reasoning: A Contribution to the Theory of Co-operation." *Research in Economics* 53 (2): 117-147.
- Bardsley, N. 2007. "On Collective Intentions: Collective Action in Economics and Philosophy". *Synthese* 157 (2): 141-159.
- Becchio, C, and Bertone, C. 2003. "Wittgenstein Running: Neural Mechanisms of Collective Intentionality and the We-Mode." *Consciousness and Cognition* 13 (1): 123-133.
- Bechtel, R., and R. Richardson. 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA: MIT Press.
- Bhaskar, R. 1979 (1998). *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. 3rd ed. London: Routledge.
- Boyd, R. 1991. "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds." *Philosophical Studies* 61 (1/2): 127-148.
- Brante, T. 2010. "Review Essay: Perspectival Realism, Representational Models, and the Social Sciences." *Philosophy of the Social Sciences* 40 (1): 107-117.
- Bratman, M. 2014. *Shared Agency: A Planning Theory of Acting Together*. New York: Oxford University Press.
- Brewer, M. 2003. *Intergroup Relations*. 2nd ed. Philadelphia: Open University Press.
- Coleman, J. 1990. *The Foundations of Social Theory*. Cambridge: Harvard University Press.
- Craver, C. 2002. "Structures of Scientific Theories." In *Blackwell Guide to the Philosophy of Science*, edited by P. Machamer, and M. Silberstein, 55-79. Oxford: Blackwell.
- Cummins, R. 2000. "“How Does it Work” Versus “What Are the Laws?": Two Conceptions of Psychological Explanation." In *Explanation and Cognition*, edited by F. Keil, and R. Wilson, 117-145. Cambridge: MIT Press.
- Dolk, T., and W. Prinz. 2016. "What it Takes to Share a Task: Sharing Versus Shaping Task Representations." In *Shared Representations: Sensorimotor Foundations of Social Life*, edited by S. Obhi, and E. Cross, 3-21. Cambridge: Cambridge University Press.
- Downes, S. M. 2011. "Scientific Models." *Philosophy Compass* 6 (11): 757-764.
- Elder-Vass, D. 2012. *The Reality of Social Construction*. Cambridge: Cambridge University Press.
- Elster, J. 2007. *More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.
- Epstein, B. 2015. *The Ant Trap. Rebuilding the Foundations of the Social Sciences*. New York: Oxford University Press.

- Epstein, B. 2018 "Social Ontology." In *The Stanford Encyclopedia of Philosophy (Summer 2018 Edition)*, edited by E. Zalta. <https://plato.stanford.edu/archives/sum2018/entries/social-ontology/>.
- Fumagalli, R. 2016. "Why We Cannot Learn from Minimal Models." *Erkenntnis* 81 (3): 433-455.
- Gallotti, M., and C. D. Frith. 2013. "Social Cognition in the We-Mode." *Trends in Cognitive Sciences* 17 (4): 160-165.
- Giere, R. 1999. *Science Without Laws*. Chicago, IL: University of Chicago Press.
- Giere, R. N. 1985. "Philosophy of Science Naturalized." *Philosophy of Science* 52 (3): 331-356.
- Giere, R. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.
- Giere, R. N. 2004. "How Models are Used to Represent Reality." *Philosophy of Science* 71 (5): 742-752.
- Gilbert, M. 2013. *Joint Commitment. How We Make the Social World*. Oxford: Oxford University Press.
- Godfrey-Smith, P. 2003. *Theory and Reality*. Chicago: Chicago University Press.
- Godfrey-Smith, P. 2005. "Folk Psychology as a Model." *Philosopher's Imprint* 5 (6): 1-16.
- Godfrey-Smith, P. 2006. "The Strategy of Model-Based Science." *Biology & Philosophy* 21 (5): 725-740.
- Goertz, G. 2006. *Social Science Concepts: A User's Guide*. Princeton: Princeton University Press.
- Gold, N., and R. Sugden. 2007. "Collective Intentions and Team Agency." *Journal of Philosophy* 104 (3): 109-137.
- Grüne-Yanoff, T. 2009. "Learning from Minimal Economic Models." *Erkenntnis* 70 (1): 81-99.
- Grüne-Yanoff, T., and Mäki, U. 2014. "Introduction: Interdisciplinary Model-Exchanges." *Studies in History and Philosophy of Science Part A* 48: 52-59.
- Guala, F. 2007. "The Philosophy of Social Science: Metaphysical and Empirical." *Philosophy Compass* 2 (6): 954-980.
- Guala, F. 2016. "Naturalism and Anti-Naturalism in the Philosophy of Social Science." In *The Oxford Handbook of Philosophy of Science*, edited by F. Guala, 43-61. Oxford: Oxford University Press.
- Hacking, I. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.
- Hacking, I. 1995. "The Looping Effect of Human Kinds." In *Causal cognition: A Multidisciplinary Debate*, edited by A. Premack, 351-383. Oxford: Clarendon Press.
- Hakli, R., Miller, K., and Tuomela, R. 2010. "Two Kinds of We-Reasoning." *Economics and Philosophy* 26 (3): 291-320.
- Hari, R., Henriksson, L., Malinen, S., and Parkkonen, L. 2015. "Centrality of Social Interaction in Human Brain Function." *Neuron* 88: 181-193.

- Harsanyi, J., and Selten, R. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press.
- Hausman, D. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Hedström, P., and Ylikoski, P. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36: 49-67.
- Heinonen, M. 2016. Minimalism and Maximalism in the Study of Shared Intentional Action. *Philosophy of the Social Sciences* 46 (2): 168-188.
- Hindriks, F., and Guala, F. 2015. "Institutions, Rules, and Equilibria: A Unified Theory." *Journal of Institutional Economics* 11 (3): 459-480.
- Joas, H., and Knöbl, W. 2009. *Social Theory: Twenty Introductory Lectures*. Cambridge: Cambridge University Press.
- Kaidesoja, T. 2013. *Naturalizing Critical Realist Social Ontology*. London: Routledge.
- Kim, J. 1984. "Concepts of Supervenience." *Philosophy and Phenomenological Research* 45 (2): 153-176.
- Kincaid, H. 1986. "Reduction, Explanation, and Individualism." *Philosophy of Science* 53 (4): 492-513.
- Kincaid, H. 2012a. "How Should Philosophy of Social Science Proceed?." *Metascience* 21 (2): 391-394.
- Kincaid, H. 2012b. "Introduction: Doing Philosophy of Social Science." In *The Oxford Handbook of Philosophy of Social Science*, edited by H. Kincaid, 3-12. Oxford: Oxford University Press.
- Kincaid, H. 2021. "Concrete Ontology: Comments on Lauer, Little, and Lohse." *Philosophy of the Social Sciences* 51 (1): 40-47.
- Knoblich, G., Butterfill, S., and Sebanz, N. 2011. "Psychological Research on Joint Action: Theory and Data." In *Psychology of Learning and Motivation*, edited by B. Ross, vol. 51, 121-54. Burlington: Academic Press.
- Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lawson, T. 2019. *The Nature of Social Reality: Issues in Social Ontology*. Cambridge: Cambridge University Press.
- List, C., and Pettit, P. 2011. *Group Agency. The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.
- Mäki, U. 2009. "MISSing the World. Models as Isolations and Credible Surrogate Systems." *Erkenntnis* 70 (1): 29-43.
- Michael, J. A., and Szgeti, A. 2019. "The Group Knobe Effect": Evidence that People Intuitively Attribute Agency and Responsibility to Groups." *Philosophical Explorations* 22 (1): 44-61.
- Niiniluoto, I. 1999. *Critical Scientific Realism*. Oxford: Oxford University Press.
- Pacherie, E. 2011. "Framing Joint Action." *Review of Philosophy and Psychology* 2 (2): 173-192.
- Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.

- Paternotte, C. 2011. "Comment on Raimo Tuomela. Joint action: How rational? How irreducible?." *Analyse & Kritik* 33 (1): 87-92.
- Pietroski, P., and Rey, G. 1995. "When Other Things Aren't Equal: Saving Ceteris Paribus Laws from Vacuity." *The British Journal for the Philosophy of Science* 46 (1): 81-110.
- Rakoczy, H. 2017. "The Development of Individual and Shared Intentionality." In *The Routledge Handbook of the Philosophy of the Social Mind*, edited by J. Kiverstein, 139-151. London: Routledge.
- Rodrik, D. 2015. *Economics Rules: Why Economics Works, When it Fails, and How to Tell the Difference*. Oxford: Oxford University Press.
- Salmela, M., and Nagatsu, M. 2016. "Collective Emotions and Joint Action." *Journal of Social Ontology* 2 (1): 33-57.
- Sarkia, M. 2021. "Modeling Intentional Agency: A Neo-Gricean Framework." *Synthese*. DOI: [10.1007/s11229-021-03103-w](https://doi.org/10.1007/s11229-021-03103-w).
- Schaffer, J. 2016. "Grounding in the Image of Causation." *Philosophical Studies* 173 (1): 49-100.
- Searle, J. 2009. "Language and Social Ontology." In *Philosophy of the Social Sciences. Philosophical Theory and Scientific Practice*, edited by C. Mantzavinos, 9-27. Cambridge: Cambridge University Press.
- Searle, J. 2010. *Making the Social World. The Structure of Human Civilization*. Oxford: Oxford University Press.
- Soros, G. 2013. "Fallibility, Reflexivity and the Human Uncertainty Principle." *Journal of Economic Methodology* 20 (4): 309-329.
- Suárez, M. 2004. "An Inferential Conception of Scientific Representation." *Philosophy of Science* 71 (5): 767-779.
- Sugden, R. 2000. "Credible Worlds: The Status of Theoretical Models in Economics." *Journal of Economic Methodology* 7 (1): 1-31.
- Suppes, P. 1960. "A Comparison of the Meaning and Use of Models in Mathematics and the Empirical Sciences." *Synthese* 12 (2-3): 287-301.
- Suppe, F. 1989. *The Semantic Conception of Theories and Scientific Realism*. Chicago: University of Chicago Press.
- Swedberg, R. 2014. *The Art of Social Theory*. Princeton: Princeton University Press.
- Teller, P. 2001. "Twilight of the Perfect Model." *Erkenntnis* 55 (3): 393-415.
- Thomasson, A. L. 2003. "Realism and Human Kinds." *Philosophy and Phenomenological Research* 67 (3): 580-609.
- Thomson-Jones, M. 2005. "Idealization and Abstraction: A Framework." In *Idealization XII: Correcting the Model*, edited by M. Thomson-Jones, and N. Cartwright, 173-217. Amsterdam: Rodopi.
- Tomasello, M. 2019. *Becoming Human: A Theory of Ontogeny*. Cambridge: Harvard University Press.
- Tuomela, R. 2007. *The Philosophy of Sociality*. Oxford: Oxford University Press.
- Tuomela, R. 2009. "Collective Intentions and Game Theory." *Journal of Philosophy* 106 (5): 292-300.

- Tuomela, R. 2013. *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press.
- Van Fraassen, B. C. 1987. "The Semantic Approach to Scientific Theories. In *The Process of Science: Contemporary Philosophical Approaches to Understanding Scientific Practice*, edited by N. Nersessian, 105-124. Dordrecht: Kluwer.
- Weisberg, M. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
- Wettersten, J. 2009. "The Philosophy of Sociality." *Philosophy of the Social Sciences* 39 (3): 531-534.
- Wettersten, J. 2010. "Reply to Tuomela." *Philosophy of the Social Sciences* 40 (3): 518-522.
- Weisberg, M. 2007. Who is a Modeler? *British Journal for the Philosophy of Science*. 58 (2): 207-233.
- Wimsatt, W. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge: Harvard University Press.
- Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Author Biography

Matti Sarkia is a researcher at the Department of Political and Economic Studies of the University of Helsinki, with experience of research at the City University of New York (CUNY) Graduate Center, the University of Munich, and the University of Melbourne, in addition to his alma mater. In recent work, he has defended a model-based approach to the philosophy of mind and action in [Sarkia \(2021\)](#). The present article extends his approach to the domain of social ontology.