Phenomenology of Social Explanation

## Shannon Spaulding

## Shannon.spaulding@okstate.edu

#### Penultimate version. Please cite only final published version.

Abstract: The orthodox view of social cognition maintains that mentalizing is an important and pervasive element of our ordinary social interactions. The orthodoxy has come under scrutiny from various sources recently. Critics from the phenomenological tradition argue that phenomenological reflection on our social interactions tells against the orthodox view. Proponents of pluralistic folk psychology argue that our ordinary social interactions extend far beyond mentalizing. Both sorts of critics argue that emphasis in social cognition research ought to be on other elements of our social practices. In this paper, I consider social explanations specifically and argue that social explanations are implicated in many of the social practices highlighted by critics of the orthodox view.

Keywords: theory of mind; mindreading; explanation; mentalizing; phenomenology

# Introduction

The orthodox view of social cognition maintains that mentalizing is an important and pervasive element of our ordinary social interactions. Mentalizing consists in attributing mental states in order to understand and interact with other people.[[1]](#footnote-1) Typically, this understanding is glossed simply as explanation and prediction of behavior. The orthodox view, then, is that in all sorts of everyday circumstances, we attribute beliefs, desires, and emotions to others and, on that basis, explain and predict their behavior. In navigating traffic, ordering coffee from a barista, meeting with students, we figure out what the others’ mental states are so that we can understand what they are doing, saying, and feeling and predict what they will do, say, or feel next.

This orthodox view has come under scrutiny from various sources recently. Critics from the phenomenological tradition argue that reflection on our experience of social interactions tells against the orthodox view. They argue that careful phenomenological reflection reveals that we rarely go through these psychological steps in our social interactions. Proponents of pluralistic folk psychology argue that our ordinary social interactions extend far beyond mentalizing. Indeed, they argue, that things like social norms and scripts are much more prevalent and fundamental to social interactions, and thus the study of social cognition ought to emphasize these elements over mentalizing.

In this paper, I consider one aspect of mentalizing: social explanation. I examine what phenomenology can reveal about social explanation. I argue that phenomenological reflection can shed light on some elements of social explanation, however, I argue that it cannot answer questions at the heart of the critiques from phenomenology and pluralistic folk psychology. Namely, it cannot tell us when social explanation occurs or how prevalent it is. I suggest we turn to behavioral studies for this question. I argue careful examination of the so-called alternatives to mentalizing indicates that mentalizing – in particular, social explanation – is implicated in many of the social practices highlighted by critics of the orthodox view.

# Types of Social Explanation

In our social interactions with others, we generate and receive different types of explanation. These explanations are answers to “why” questions about others’ behaviors and thoughts. In this section, I will explore the nature of these social explanations.[[2]](#footnote-2) Though there are many ways to categorize explanations, I will discuss teleological, mechanistic, and formal explanations. This taxonomy has been used fruitfully in the philosophy and psychology of explanations (Lombrozo 2012, Lombrozo and Carey 2006, Dennett 1987). Teleological explanations answer “why” questions about a target’s behavior or thoughts by citing the target’s goals or intentions. For example, one might explain a mother nursing her newborn by saying that the mother *wanted* to nourish or soothe the baby. The goal, which is not satisfied until the action is complete, is a distal cause of the behavior. Mechanistic explanations answer “why” questions about a target’s thoughts or behavior by citing proximal causes. In explaining the mother feeding her newborn, one might say that a baby’s hunger cries stimulate cortisol, a stress hormone, in the mother, which causes anxiety in the mother and an urge to find ways to relieve the baby’s hunger. Moreover, in anticipation of nursing, the mother’s dopaminergic neurons activate and stimulate a reward response. Both of these mechanistic explanations specify an immediate cause of the behavior in question. Formal explanations answer these “why” questions by citing kind membership norms. A formal explanation of the mother’s behavior would assert that the mother who is nursing her baby is taking care of her baby and taking care of babies is just what mothers do. Formal social explanations identify social norms and stereotypes of social groups. The three types of explanations are not mutually exclusive. Depending on the context, one or more of these explanations may be satisfactory.

|  |  |  |  |
| --- | --- | --- | --- |
| Type of Explanation | Definition | Example | Contextual Use |
| Teleological | Cite goals or intentions | Nursing mother wants to nourish or soothe her baby | Cited goal helps us generalize, predict, or intervene on behavior |
| Mechanistic | Cite proximal causes  | Baby’s cries stimulate stress response, and anticipating nursing generates a reward response  | Causal origin helps us predict and intervene on behavior |
| Formal | Cite kind membership norms | Nursing is one way for a mother to care for her baby, which is what mothers do | Kind membership helps us categorize and generalize behavior |

How exactly do these types of explanation *explain?* Most psychologists and philosophers working in the this literature regard teleological and mechanistic explanations as *causal* explanations (Lombrozo 2012, Lombrozo and Carey 2006). That is, in answering a “why” question, these types of explanations specify a causal path that leads to the behavior in question. For instance, explaining the mother’s behavior by citing the hormones released in her brain specifies (part of) the neurological cause of the mother’s behavior. Citing her desire or intention to nourish and soothe her baby specifies the psychological cause of her behavior. These mechanistic and teleological explanations of a mother’s behavior pick out causes at different levels of description.

The predominant view about formal explanation is that it is non-causal (Lombrozo 2012, Prasada and Dillingham 2006, 2009). On this view, you explain something by categorizing it as part of a larger class of phenomena. To use a common example, you explain why an animal has four legs by citing the fact that it is a dog and dogs are the kind of creatures that typically have four legs. Citing kind membership does not, on this view, trace a causal pathway to state of affairs or behavior in question. However, I think we should regard formal explanations, at least those that are satisfying, as implicitly causal (Spaulding 2018a, ch. 4). Here is why. If formal explanations were devoid of causal information, like it seems to be in the dog/4-legs explanation, then it is hard to see why we would ever find formal explanations compelling at all. They would appear to us as circular restatements of the phenomena in question. We would never produce or accept them as genuine explanations.[[3]](#footnote-3) However, we can distinguish circular explanations from formal explanations. I will argue that we *do* generate and accept non-circular formal explanations of others thoughts and behavior.

Stereotypical explanations – i.e., explanations that cite stereotypes as an answer to a “why” question about a target’s thoughts and behavior – are formal explanations, and we trade in such explanations frequently in social interactions. Consider the following example. Suppose that in conversation with a friend, I wonder why a mutual friend signed up for a long bike race. My friend offers up the following explanation: “Josh always does things that. He’s a fitness junkie.” The explanation asserts that Josh is member of a certain social kind – fitness junkie – and in invoking the stereotype it implies a causal relationship between social kind and the behavior in question. Specifically, it implies that motivations and habits that are typical of the stereotypical fitness junkie – a desire to challenge oneself physically, a desire to stay physically fit, etc. – motivate Josh to do things like sign up for a long bike race. What the explanation does, if it is successful, is focus my attention on a causal pattern of behavior that is typical of a social group.[[4]](#footnote-4)

The question posed above is how these types of explanations explain, and the answer I put forward is that all three types of social explanation – teleological, mechanistic, and formal – explain the phenomenon in question by highlighting a causal path from the *explanans* to the *explanandum*. That is, they specify psychological states, neurological states, or social category norms that lead to the thoughts or behavior in question.

The next question is when we generate and accept each of these kinds of social explanation. Tania Lombrozo and Susan Carey argue that the psychological function of explanation is to provide information that can be exported to novel cases (Lombrozo and Carey 2006). Their hypothesis, called *Explanation for Export*, holds that we prefer explanations that provide information that we can use in generalizations, predictions, and future interventions. Thus, not just any mechanism, goal, or norm is relevant in explaining a target’s thoughts or behavior. Only explanations that cite features that play a causal role in a causal process that conforms to a predictable pattern are acceptable because that information will be useful for generalizing, predicting, and intervening.

The Explanation for Export hypothesis specifies what makes each type of explanation appropriate, but what factors influence our preference for one *type* of explanation over another? We prefer explanations that highlight information that will be useful for inductive inferences we expect to make (Vasilyeva, Wilkenfeld, and Lombrozo 2017). For example, when the task at hand involves identifying the goal of some thing or behavior, teleological explanations are preferred over other kinds of explanations. Teleological explanations are the default type of explanation, and they seem to come most naturally to children and adults (Lombrozo 2012). Thinking about what an object is supposed to do or an agent is trying to do is intuitively useful for generalizing, predicting, and intervening. Sometimes, however, teleological explanations are not as helpful because something behaves counter to presumed goals or we lack information on what the possible goals could be. In such cases, information about the proximal cause of the behavior is extremely useful for predictions and interventions. Finally, in cases where we are interested in categorizing things, behaviors, or events – i.e., making sense of them in terms of familiar frameworks – formal explanations are most appropriate. Such explanations allow us to make predictions about the singular thing we are categorizing and generalizations about members of the category.

With respect to social interactions, teleological explanations are the default (as they are in general). When we want to predict or expect to intervene on someone’s thoughts or behavior, understanding what that person is thinking is extremely useful. However, we sometimes offer stereotypical explanations rather than teleological explanations. Stereotypical explanations, I argue above, are a kind of formal explanation. Stereotypical formal explanations interpret a target’s behavior in terms of a more general social category, which allows us to generalize and make predictions about future thoughts and behavior. Least frequently, we produce or accept mechanistic social explanations. Mechanistic explanations that specify proximal causes of thoughts or behavior may be preferred when explaining unusual or bizarre behavior, especially when teleological and formal explanations fail to explain the behavior in question. For instance, when my child gets progressively crankier and less rational, I stop trying to figure out what it is that she wants, and I start to look for a different kind of explanation: When was the last time she ate? Is she overly tired? Is she overstimulated in this environment? I look for the proximal cause of the crankiness and irrationality… and then give her a snack to raise her blood sugar levels.

The *explanandum* may be a target’s behavior – as many of my examples show – but the *explanandum* also could be the target’s mental states. In such a case, one might be interested in explaining why a target has particular mental states. Teleological and formal explanations could explain either behavior or mental states. For teleological explanations, we may want to figure out the reasons why a target behaves a certain way, but we may also want to know why a target has certain beliefs, desires, emotions, intentions, etc. For formal explanations, we may want to categorize behavioral patterns as typical of certain social groups, but we may also want to categorize a target’s mental states as typical of certain social groups. Mechanistic social explanations aim to explain mental states by reference to proximal causes of those mental states, which may include particular neurological causes, other mental states, sensations, the perceptual environment, etc.[[5]](#footnote-5)

This section provides a taxonomy of social explanation, an account of how these types of social explanations *explain* a target’s thoughts and behavior, and a framework for understanding when we would generate or accept each type of social explanation. In the next section, I discuss the phenomenology of these three types of social explanation.

# Phenomenology of Social Explanation

Phenomenology is the study of the nature and structure of our conscious experiences, from a first-person perspective.[[6]](#footnote-6) A phenomenological analysis may examine conscious experiences in terms of temporal, spatial, attentional, kinesthetic, social, or self-awareness. Phenomenological analyses of social experiences aim to describe the various attentional experiences that constitute or underlie our social interactions. The phenomenology of social explanations consists in a first-person analysis of our conscious experience of social explanations.

In this section, I will discuss numerous dimensions of social explanation: the process vs. product of explanation, producing vs. receiving explanations, explicit vs. implicit explanations, conscious vs. unconscious explanations, and the feeling of knowing or satisfaction that a social explanation can generate. I will examine what phenomenological reflection can reveal about each of these dimensions.

The prototypical conception of a social explanation seems to be an internally generated, explicit, conscious explanation of a target’s behavior. For instance, when I wonder, or someone asks me, why David is irritable, I may try to generate an explanation of his behavior – e.g., I may say, to myself or to the other person, that David is anxious about an upcoming event. In this kind of case, the process that yields an explanation is deliberative and conscious, and the explanation that results is conscious and explicit. Any of the three types of social explanation discussed above can be of this sort. In other words, I can deliberately and consciously produce an explicit, conscious teleological, mechanistic, or formal explanation of a target’s thoughts or behavior.

Careful phenomenological reflection could reveal quite a lot about consciously, deliberatively produced explicit social explanations. It could illuminate the attentional focus of the observer, i.e., which elements of the situation, behavior, or person the observer is attending to in generating an explanation. To use the example from above, careful reflection on my conscious experience could reveal what aspects of David’s demeanor and behavior I am focused on (e.g., his body language, facial expressions, tone of voice) and what background information I am considering (current stresses and upcoming events in his life). It could reveal the content (expressed in inner speech) and valence (positive or negative) of the explanation produced. Phenomenological reflection can also probe the feeling of knowing produced by the explanation, for example, how satisfied I am with the explanation and how well I feel I understand the behavior in question.

However, even for the prototypical conscious, explicit social explanation, phenomenological reflection has limits. There are general concerns about the reliability of introspective reports that are relevant in this debate. These are worries about the consistency, accuracy, and potential bias of such reports (Spaulding 2015). In terms of consistency, generally there seems to be quite a lot of interpersonal and intrapersonal variability in phenomenological reports of first-person conscious experiences. Data for this claim comes from experiments on visualization, peripheral perception, introspection of the features of an emotional experience, among other areas (Schwitzgebel 2008). Different individuals will report different first-person conscious experiences of the same event, and a single individual over time will report different first-person conscious experiences of the same event over time.[[7]](#footnote-7) Thus, in experiments on visualization, peripheral perception, and emotional experience, there is a notable lack of consistency in reporting on one’s conscious experiences. There is no particular reason to think that phenomenology of *social explanation* would an exception to this pattern.

Variability is not a problem if we have a standard for assessing the accuracy of phenomenological reports, i.e., a way to tell if what the subject reports as her phenomenology accurately reflects what is happening in her mind. However, we lack such a standard in many cases, including in the case of phenomenology of social explanation. We do not have a simple way to assess whether what the subject reports in social interactions is all and only what the subject is in fact experiencing.[[8]](#footnote-8)

Finally, there is a worry about a reporting bias in phenomenological reports. Subjects do not simply read off their thoughts, as if reading from a page. They interpret their conscious experiences, their unfinished, elliptical thoughts in inner speech, their visualizations, and gut feelings, and they make interpretive decisions about what they are experiencing. In asking them to report their experiences, subjects must interpret and articulate their conscious experiences, and in doing so they may lose some of the granularity of the experience, ignore some relevant experiences, or shift how they interpret or recall the experience. This is true in all cases of introspection, including the phenomenological introspection of our social explanations. Thus, we should not rest *too* much theoretical or empirical weight on our phenomenological reports on the existence, content, or nature of social explanations, even in the case of conscious, explicit social explanations.

What about the other forms that social explanations may take? These include non-conscious and implicit social explanations.[[9]](#footnote-9) Why think that explanations can be implicit and non-conscious? At first blush, the idea of non-consciously, implicitly explaining a target’s behavior may sound strange. The term *explanation* connotes a deliberative, conscious cognitive process. Indeed, some proponents of phenomenological accounts of social cognition argue that unconscious, implicit explanation and prediction simply make no sense. Shaun Gallagher argues, “Explanation (or theory) seems to mean (even in our everyday psychology) a process that involves reflective consciousness. The term ‘prediction’ also seems to me to describe a reflective conscious act… ‘Explanation’ and ‘prediction’ are personal-level terms” (Gallagher 2005, 215). Let’s set aside the slide between conscious/non-conscious and persona/sub-personal here (Drayson 2012).[[10]](#footnote-10) Gallagher’s point seems to be that explanation is exclusively a conscious-level phenomenon.

I admit that the notion of unconscious explanation seems counterintuitive. (The notion of implicit explanation, to my mind, is less counterintuitive. Unarticulated or inchoate answers to “why” questions seem like conceptually coherent theoretical posits.) The terms explanation and prediction descended from philosophy of science discussions of what a theory does (Spaulding 2015). Logical Positivism held that scientific theories deduce explanations and predictions from a description of the world and general principles (or laws) of the theory. This Deductive-Nomological conception of scientific theories fell by the wayside as many scholars pointed out various technical difficulties. However, later conceptions of scientific theories kept this emphasis on explanation and prediction. The contemporary study of social cognition started around the same time that alternatives to the Deductive-Nomological model were being developed, and philosophers and psychologists explicitly used language from philosophy of science to articulate theories of how we understand others. These scholars regarded our ability to understand and interact with other people in terms of a *folk psychological theory*. The idea was that just as scientists can explain and predict the position of the stars in the sky with their astronomical theories, we ordinary folk can explain and predict the behavior of people with our folk psychological theories. The difference is that folk psychological theories operate implicitly and non-consciously, they clarified. Later, an alternative to this so-called Theory Theory emerged in the form of Simulation Theory of folk psychology. This theory offered a different account of how we understand and interact with other people – via mental simulation, not application of a theory – but the language of explanation and prediction remained. Even now, with nuanced theories of pluralistic folk psychology on offer, explanation and prediction still feature prominently in theories of how we understand and interact with other people (Andrews 2008, Andrews, Spaulding, and Westra 2020, Spaulding 2018b).

This historical interlude is meant to show that the language of explanation and prediction became part of the fabric of this field from the very beginning. But it did not have to be that way. We could perfectly coherently substitute *interpretation* for *explanation* and *anticipation* for *prediction*. The onlyreason that we use explanation and prediction in this context is residual terminological influence from the philosophy of science. Interpretation and anticipation encompass all the phenomena we want to describe with the terms *explanation* and *prediction*. Moreover, it seems perfectly sensible to talk about implicit and unconscious interpretation/anticipation. These do not have the hint of conceptual incoherence that some might decry for unconscious explanation/prediction.

Inertia is a powerful thing, however, and I will not attempt to change decades-long use of terminology in a huge interdisciplinary field. The field will likely continue to talk about explanation in this context. I hope to have defused the concern that unconscious explanation is incoherent by highlighting that this term could easily be replaced by unconscious interpretation. With that long aside complete, we can return to the main issue, which is what phenomenology can tell us about non-conscious, implicit social explanations.

Any of the three types of social explanation detailed above – teleological, mechanistic, or formal – can be non-conscious and implicit. More precisely, both the underlying cognitive processes and the end product can be unarticulated and not consciously accessed. For teleological explanations, this would involve inferring a goal implicitly and non-consciously. For mechanistic explanations, inferring a proximal cause would be implicit and non-conscious. And for formal explanations, categorizing a person or behavior in terms of a social category norms would be implicit and non-conscious.

What can phenomenology tell us about these types of explanations? Recall that for conscious, explicit explanations phenomenology could illuminate the attentional focus of the observer, content and valence of the explanation, and the feeling of knowing produced by the explanation. Phenomenology is less helpful in illuminating the attentional focus of an observer (or more broadly, the processes that underlie explanation) and content of the explanation for non-conscious, implicit explanations. Simply put, if the process and product are non-conscious, even a careful phenomenological reflection will not shed light on these features for any of the three types of explanations. However, phenomenological reflection may still be helpful in identifying the valence of implicit, non-conscious explanation. We may have lingering positive or negative feelings that are subject to phenomenological reflection. We may not recognize the source of these feelings, but the feelings themselves may be relatively transparent to reflection. Similarly, we may also be able to probe our feeling of knowing, i.e., how well we feel we understand. The feeling of knowing can be quite strong while the explanation is implicit or unarticulated. In other words, feeling strongly that one understands a social interaction is an indication that social explanation occurred, but it does not imply that the social explanation is conscious or explicitly articulated. Evidence for this claim comes from Frank Kiel’s work on the psychology of the feeling of knowing (Rozenblit and Keil 2002). In one particular experiment, psychologists asked subjects how strongly they felt like they understood how a bicycle operates, and then subjects were instructed to draw a bicycle. The feeling of knowing did not correlate with more accurate drawings of bicycles. Indeed, most people reported a strong feeling of knowing how a bicycle works and still drew bicycles that would have been physically impossible to ride or even stand up. Thus, phenomenological reflection can reveal the feeling of knowing, which is an indication that the subject engaged in some sort of social explanation. However, it does not thereby reveal the content of that social explanation and, in some cases, may mislead us into thinking we understand more than we in fact do.

Let’s sum up the terrain covered in this section. I have argued that phenomenological reflection is a useful tool for analyzing the valence of social explanations in general and the feeling of knowing associated with social explanations of any sort. These are important features of explanation that receive much discussion in the psychological literature. Phenomenology may also shed light on the attentional focus of an observer and content of explanations for explicit, conscious explanations. Though I urged we proceed with caution here, as there are concerns about the consistency, accuracy, and potential for reporting bias with phenomenological reports. Finally, I argued that phenomenology is not useful for analyzing the content or processes that underlie social explanations that are non-conscious and implicit.

# Mentalizing Debate

Social explanations are explanations of others’ thoughts and behavior. Not every social explanation counts as mentalizing, of course. One might adopt an anthropological stance in explaining a person or group’s thoughts and behavior. However, social explanations that invoke mental states to explain thoughts and behavior do count as mentalizing. Any of the three types of explanation described above can invoke mental states to explain behavior and thereby count as mentalizing. Teleological explanations do this by default insofar as they cite goals, beliefs, and desires to explain thoughts and behavior. Formal explanations may count as mentalizing by invoking social norms and stereotypes about the mental states of certain groups or types of people to explain their behavior. Less commonly, mechanistic explanations may cite mental states as proximal causes of thoughts and behavior, i.e., mental states as immediate triggers of thoughts and behavior. For example, when you hear a song that reminds you of a deceased loved one, this may trigger an immediate feeling of sorrow.

Debates about the nature and prevalence of mentalizing in our ordinary interactions pose a challenge to the orthodox view in the literature on social cognition. For decades, the orthodox view in social cognition has been that mentalizing – attributing mental states in order to explain and predict behavior – is fundamental to and a ubiquitous part of normal human social interactions. On this view, from relatively early on in childhood and throughout adulthood, neurotypical humans attribute mental states and explain and predict behavior virtually constantly throughout the day. Teaching a room full of students, chatting with a cashier, negotiating with colleagues, caring for your children, being engrossed in a work of fiction, mentally rehearsing a potential interaction, all allegedly involve mentalizing. The orthodox view holds that, in various ways, each of these examples from ordinary life involve attributing a mental state (to another person, to one’s future or past self, to a fictional character) and explaining and predicting behavior on the basis of that mental state attribution.

The idea that mentalizing is both fundamental to and ubiquitous in our ordinary social interactions has come under scrutiny from multiple sources. Some phenomenologists and others in the embodied cognition research programme argue that our ordinary social interactions do not involve mentalizing. Instead, they argue that more basic capacities that underlie mentalizing are (i) developmentally prior to mentalizing, (ii) present in adults, (iii) and sufficient for sustaining social interactions in the absence of mentalizing. This line of argumentation is meant to establish that mentalizing is not developmentally fundamental or essential to our interactions as adults. Furthermore, they argue that we have good reason to think that mentalizing is a rarely used specialized skill. They argue that careful phenomenological reflection shows that we are rarely attributing mental states and explaining and predicting behavior (Spaulding 2010). Finally, some proponents of pluralistic folk psychology argue that while we do on occasion mentalize, we also do many other things besides this, and an adequate theory of social cognition ought to emphasize the diversity of folk psychology (Spaulding 2018a, ch. 2).

There is a lot to consider in these objections, so let’s extract the threads that are relevant to this project. One thread concerns the prevalence of mentalizing. Everyone accepts that there are developmental precursors to a mature, fully developed capacity for mentalizing. Even dyed-in-the-wool nativists accept this claim. Thus, it is not problematic to highlight the role of primary intersubjectivity (the capacity to perceive actions as intentional, read others’ eye movements, facial gestures, and feelings) and secondary intersubjectivity (the capacity for joint attention).[[11]](#footnote-11) The relevant question is whether having these capacities makes mentalizing unnecessary in most social interactions. Can the ordinary social interactions I described above proceed with just an ability to read body language, joint attention, and knowledge of social scripts? And, for our purpose, what kind of evidence is relevant for answering this question?

Some scholars have argued that phenomenology can answer this question. Shaun Gallagher argues that careful phenomenological reflection shows that our ordinary interactions do not seem to involve mental state inferences or explanation and prediction. (Gallagher 2001, pp. 89-92, 2004, pp. 201-204, 2005, pp. 208-216). He claims that “a careful and methodical phenomenology should be able to tell us whether, when we hear the exasperated voice, our usual response involves formulating an explanation or predicting what the person will do next. Our encounters with others are in fact not normally occasions for theorizing or simulating if such nonconscious procedures are cashed out phenomenologically as explaining or predicting on the basis of postulated mental states” (Gallagher, 2001, 89).

At this point, we can revisit the conclusions from the previous section about what phenomenology can reveal about social explanations. I argued that for explicit, conscious social explanations, phenomenology may illuminate the content and valence of an explanation, the attentional focus of the explainer, and his or her feeling of knowing. I expressed some qualms about how reliable phenomenology can be with respect to the content of explanations but allowed that it can play this role. Thus, Gallagher is partially right insofar as phenomenology can shed some light on conscious, explicit social explanations.[[12]](#footnote-12)

However, phenomenology cannot answer the question of how *prevalent* mentalizing is because it tells us very little about implicit, non-conscious explanations. It cannot tell us whether such explanations occur or the content of them. Though it may be relevant when assessing valence and feeling of knowing that arises in a social interaction, this is less helpful when the question is whether one has unconsciously, implicitly explained (or interpreted, if you prefer) a target’s thoughts or behavior. Having phenomenological access only to the valence and feeling of knowing and not the cognitive process that generates those effects may lead one to mistakenly conclude that the underlying cognitive process did not – and does not need to – occur.

Using phenomenology to assess the existence or prevalence of social explanations is an instance of the streetlight effect, wherein we look for something only where it is easy to look. On a dark night, we might look for our lost keys in the light of a streetlamp. The streetlamp is helpful only if our keys are in the lit area. If they are not, we might mistakenly conclude our keys are gone. Phenomenological reflection reveals the presence and nature of social explanation in the same way that the streetlamp sheds light on a dark night. It is a helpful tool when what we are looking for just so happens to be where we are looking for it, but otherwise it is a fairly limited tool.

Thus, to summarize, phenomenological reflection cannot answer questions about the prevalence of social explanations because it is not the right tool for examining the presence, prevalence, or nature of implicit, unconscious explanations. That entails that phenomenology is not the right tool for examining the orthodox view’s claim that mentalizing is essential and pervasive.

How then do we answer those questions? Behavioral studies. Earlier, I discussed research from the psychology of explanation that shows when we generate and accept various types of explanation. This research finds that we generate/accept teleological explanations when we expect to intervene on behavior. When we want to generalize, categorize, or moralize about a social group, we generate/accept formal social explanations that identify a pattern of typical thoughts and behavior. When we lack the relevant information for inferring goals, or when the situation is unusual, we generate/accept mechanistic social explanations that infer the proximal cause of a target’s thoughts and behavior. With this framework in hand, we can simply ask how often people have these motivations. If we frequently have the motivation to intervene on others’ thoughts and behavior,[[13]](#footnote-13) or if we are frequently motivated to generalize or moralize about others’ behavior, other things being equal, this will lead us to expect that we frequently generate/accept these types of social explanations of others’ thoughts and behavior.

There are other behavioral studies that can answer the question of how prevalent mentalizing is in social interactions. Psychologists have crafted clever behavioral studies aimed at testing whether subjects implicitly, unconsciously attribute mental states to others in their social interactions. These studies document various forms of level-2 perspective taking that are spontaneous, rapid, implicit, and non-conscious (Samson et al. 2010, Elekes, Varga, and Király 2016, Surtees, Apperly, and Samson 2016). They find that subjects will spontaneously and rapidly adopt a target’s level-2 perspective. That is, they will process not just *what* a target can see but also *how* it looks from the target’s perspective. The spontaneously do this even when not instructed to do so and even when doing so interferes with completing their own tasks. These studies find that we spontaneously, unconsciously, implicitly represent others’ psychological perspective.[[14]](#footnote-14)

Some of the findings advanced in favor of implicit mentalizing have a complex replication record, especially anticipatory looking and violation of expectation (Kulke, Johannsen, and Rakoczy 2019, Dörrenberg, Rakoczy, and Liszkowski 2018, Powell et al. 2018). It is unclear what to make of the pattern of successful and failed replications of this work. I do not want to rest too much on claims of specific studies because the question we are addressing here is methodological: How can we assess this claim about the scope of mentalizing? I have offered a negative answer and a positive answer. The negative answer is: Not through introspective phenomenology. The positive answer is: Through behavioral studies that test whether and when we implicitly, unconsciously mentalize.

There is a second thread in the objections above that is relevant here, and this thread regards the prevalence of mentalizing social explanations vs. alternative social practices. Phenomenologists like Gallagher and some proponents of pluralistic folk psychology, like Kristin Andrews (2008) and Tadeusz Zawidzki (2013), regard mentalizing as just one small, relatively unimportant component of our ordinary social interactions. These critics emphasize the role of trait attributions, stereotypes, behavioral schema, and social scripts. These other elements of our social interactions are, on their views, much more significant than mentalizing. Indeed, they argue, we can get by without mentalizing when we have these other components in place. We mentalize only when we encounter truly strange behavior that we cannot make sense of with our stereotypes, scripts, and schema. Relatedly, Zawidzki and Victoria McGeer argue social interactions often are aimed at shaping and regulating others rather than explaining and predicting their behavior (McGeer 2007, Zawidzki 2013). The primary function of our social practices is to make others conform to certain ways of thinking and behaving to facilitate cooperation. Thus, our social interactions often do not involve explanation and prediction.

I think that these claims about our ordinary social interactions are, by and large, right. We have many tools in our social cognition toolkits, and we do so much more than explain and predict behavior. However, I do not think these claims are incompatible with mentalizing being fundamental and ubiquitous. To the contrary, I think that mentalizing – in particular, teleological and formal social explanations of thoughts and behavior – is implicated in many of these social practices.

Mentalizing is a necessary element in almost all the social practices highlighted by the critics above (Spaulding 2018b, Spaulding 2019). For example, mindshapingand regulative folk psychology presuppose mentalizing insofar as one needs to know what a target thinks and feels in order to shape or regulate their minds and behavior. That is, we need a teleological social explanation in order to intervene on the thoughts and behavior. Suppose I have the goal of mindshaping my students to believe that with a bit more effort and organization they can perform better in the class. First, I need to understand what my students think about their performance. Consider a student who simply does not care about maximizing her grade. She is content with getting a merely passing grade. In order to mindshape her, I need to emphasize the importance of the course content or the significance of GPA and relate this to values and goals she does care about. A student who wants to do well but does not know how to study will require a different intervention. A student who thinks she is not smart enough to understand the material and therefore does not even try is different still. Without some idea of what a target’s goals are, we could not engage in mindshaping effectively. Thus, while I agree with the claim that we are frequently engaged in mindshaping, I contend that mindshaping involves mentalizing, in particular, teleological social explanations.

Mentalizing is also involved in the attribution of many social stereotypes, which I argue is a kind of formal explanation. Stereotypes play a role in our social interactions in two ways. First, we sometimes directly attribute a stereotype to an individual or group of people to explain some thoughts or behavior. Second, a salient stereotype can influence how we interpret an individual’s traits, mental states, and behaviors. In both cases, the inferred thoughts are essential to the stereotype use (Westra 2017b). They help us make sense of a wide range of behaviors. Consider, for example, the stereotype that women are more empathetic. Directly attributing this stereotype to a person or group of people implies that the target has certain beliefs, values, and desires. In this case, it implies that the target cares about others’ suffering, she likes listening to others, and that she desires to help others. When this stereotype is salient to us, it makes us certain inferences about personality traits, behavior, and mental states more likely. In this case, when the female empathy stereotype is salient, we are more likely to interpret behavior as prosocial, attribute caring motives, expect that helping will make the target happy, etc. Social stereotype usage, a kind of formal social explanation, is rife with mentalizing.

Similar remarks apply to personality trait attribution, as well (Westra 2017a). When we infer a personality trait, we infer a host of mental states, such as desires, cares, and values. Like stereotype attribution, personality trait attribution is effective because the personality trait presupposes mental states that both explain a pattern of thoughts behavior and predict future thoughts and behavior.

Summing up, I agree with the claim from phenomenologists and pluralistic folk psychology that our social interactions involve a broad array of social practices, such as mindshaping, stereotyping, attributing personality traits, employing social scripts and norms. Furthermore, I agree that we should incorporate these practices into our theories of social cognition. I have argued here that mentalizing – in particular, teleological, and formal social explanation – is at the root of many of the social practices emphasized by phenomenologists and pluralistic folk psychology. Thus, the critiques of mentalizing research successfully highlight social practices that are important and prevalent in our social interactions. However, these practices involve mentalizing, and thus highlighting them does undermine the claim that mentalizing is fundamental and ubiquitous.

# Conclusion

The orthodox view of social cognition maintains that mentalizing is an important and pervasive element of our social interactions. The claim is that in ordinary social interactions, we attribute beliefs, desires, emotions to others and, on that basis, explain and predict their behavior. Phenomenologists and proponents of pluralistic folk psychology argue that mentalizing is a rarely used, specialized skill that we employ only when our more basic cognitive tools fail. Instead of mentalizing, we attribute personality traits, stereotypes, social scripts, read body language, use joint attention, etc.

In this paper, I considered social explanations specifically and examined what phenomenological reflection can reveal about this aspect of mentalizing. I argued that phenomenological reflection can shed light on some elements of social explanation, it cannot answer questions at the heart of the critiques from phenomenology and pluralistic folk psychology. Namely, it cannot tell us when social explanation occurs or how prevalent it is. I suggested we turn to behavioral studies for this question. I argued for a broad conception of mentalizing in which social explanation and other elements of mentalizing are implicated in many of the social practices highlighted by critics of the orthodox view. Though the broad conception mentalizing I articulate is in some ways a deviation from the orthodox picture of mentalizing, it supports the orthodox view’s claim that mentalizing is fundamental and ubiquitous.

# Compliance with Ethical Standards

The author has no relevant financial or non-financial interests to disclose.

**References**

Andrews, Kristin. 2008. "It’s in your nature: A pluralistic folk psychology." *Synthese* 165 (1):13-29.

Andrews, Kristin, Shannon Spaulding, and Evan Westra. 2020. "Introduction to Folk Psychology: Pluralistic Approaches." *Synthese*. doi: 10.1007/s11229-020-02837-3.

Brenner, Andrew, Anna-Sofia Maurin, Alexander Skiles, Robin Stenwall, and Naomi Thompson. Winter 2021 Edition. Metaphysical Explanation. In *The Stanford Encyclopedia of Philosophy*, edited by Edward. N. Zalta.

Carruthers, Peter. 2016. "Two systems for mindreading?" *Review of Philosophy and Psychology* 7 (1):141-162.

Dennett, D. C. 1987. *The Intentional Stance*. Cambridge, MA: MIT press.

Dörrenberg, Sebastian, Hannes Rakoczy, and Ulf Liszkowski. 2018. "How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures." *Cognitive Development* 46:12-30.

Drayson, Zoe. 2012. "The uses and abuses of the personal/subpersonal distinction." *Philosophical perspectives* 26:1-18.

Elekes, Fruzsina, Máté Varga, and Ildikó Király. 2016. "Evidence for spontaneous level-2 perspective taking in adults." *Consciousness and Cognition* 41:93-103.

Gallagher, S. 2001. "The practice of mind. Theory, simulation or primary interaction?" *Journal of Consciousness Studies* 8 (5-7):83-108.

Gallagher, S. 2004. "Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind." *Philosophy, Psychiatry, & Psychology* 11 (3):199-217.

Gallagher, S. 2005. *How the Body Shapes the Mind*: Oxford University Press, USA.

Gelman, Susan A. 2003. *The essential child: Origins of essentialism in everyday thought*: Oxford Series in Cognitive Dev.

Gelman, Susan A. 2004. "Psychological essentialism in children." *Trends in cognitive sciences* 8 (9):404-409.

Kulke, Louisa, Josefin Johannsen, and Hannes Rakoczy. 2019. "Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts." *PloS one* 14 (3):e0213772.

Lombrozo, T. 2012. "Explanation and abductive inference." In *Oxford Handbook of Thinking and Reasoning*, edited by Keith J. Holyoak and Robert G. Morrison, 260-276. Oxford University Press.

Lombrozo, T., and S. Carey. 2006. "Functional explanation and the function of explanation." *Cognition* 99 (2):167-204.

Malle, B. F. 2004. *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction*. Cambridge, MA: MIT Press.

Mancosu, Paolo. Summer 2018 Edition. Explanation in Mathematics. In *The Stanford Encyclopedia of Philosophy*, edited by Edward. N. Zalta.

McGeer, V. 2007. "The regulative dimension of folk psychology." In *Folk psychology re-assessed*, 137-156. Springer.

Piaget, J. 1969. *The Child's Conception of Physical Causality*. Totawa, NJ: Littlefield, Adams & Co.

Powell, Lindsey J., Kathryn Hobbs, Alexandros Bardis, Susan Carey, and Rebecca Saxe. 2018. "Replications of implicit theory of mind tasks with varying representational demands." *Cognitive Development* 46:40-50. doi: <https://doi.org/10.1016/j.cogdev.2017.10.004>.

Prasada, S., and E. M. Dillingham. 2006. "Principled and statistical connections in common sense conception." *Cognition* 99 (1):73-112.

Prasada, S., and E. M. Dillingham. 2009. "Representation of principled connections: A window onto the formal aspect of common sense conception." *Cognitive Science* 33 (3):401-448. doi: 10.1111/j.1551-6709.2009.01018.x.

Rozenblit, Leonid, and Frank Keil. 2002. "The misunderstood limits of folk science: An illusion of explanatory depth." *Cognitive science* 26 (5):521-562.

Samson, D., I.A. Apperly, J.J. Braithwaite, B.J. Andrews, and S.E. Bodley Scott. 2010. "Seeing it their way: Evidence for rapid and involuntary computation of what other people see." *Journal of Experimental Psychology: Human Perception and Performance* 36 (5):1255.

Schwitzgebel, E. 2008. "The unreliability of naive introspection." *Philosophical Review* 117 (2):245-273.

Spaulding, S. 2010. "Embodied cognition and mindreading." *Mind & Language* 25 (1):119-140.

Spaulding, S. 2015. "Phenomenology of social cognition." *Erkenntnis* 80 (5):1069-1089. doi: 10.1007/s10670-014-9698-6.

Spaulding, S. 2018a. *How We Understand Others: Philosophy and Social Cognition*. New York, NY: Routledge.

Spaulding, S. 2019. "What is Mindreading?" *Wiley Interdisciplinary Reviews: Cognitive Science* 11 (3).

Spaulding, Shannon. 2018b. "Mindreading beyond belief: A more comprehensive conception of how we understand others." *Philosophy Compass* 13 (11):e12526. doi: 10.1111/phc3.12526.

Surtees, Andrew, Ian Apperly, and Dana Samson. 2016. "I’ve got your number: Spontaneous perspective-taking in an interactive task." *Cognition* 150:43-52.

Vasilyeva, N., D. Wilkenfeld, and T. Lombrozo. 2017. "Contextual utility affects the perceived quality of explanations." *Psychon Bull Rev*. doi: 10.3758/s13423-017-1275-y.

Westra, E. 2017a. "Character and theory of mind: an integrative approach." *Philosophical Studies* 175 (5):1217-1241.

Westra, E. 2017b. "Stereotypes, theory of mind, and the action-prediction hierarchy." *Synthese* 196 (7):2821-2846.

Westra, Evan. 2017c. "Spontaneous mindreading: A problem for the two-systems account." *Synthese* 194 (11):4559-4581.

Williams, Heath. forthcoming. "The unbearable lightness of the personal, explanatory level." *Phenomenology and the Cognitive Sciences*.

Woodward, James, and Lauren Ross. Summer 2021 Edition. Scientific Explanation. In *The Stanford Encyclopedia of Philosophy*, edited by Edward. N. Zalta.

Zawidzki, T. W. 2013. *Mindshaping: A New Framework for Understanding Human Social Cognition*: MIT Press.

1. This phenomenon goes by many names: mindreading, theory of mind, folk psychology, among others. I use “mentalizing” because a recent interdisciplinary survey found that this term is the most commonly used and recognized term for the phenomenon. [↑](#footnote-ref-1)
2. This section draws on Chapter 4 of Spaulding (2018a). [↑](#footnote-ref-2)
3. I do not mean to imply that *all* genuine explanations are causal. That would require more argumentation than I have given here. We can make room for non-causal explanations, perhaps reasons or justifying explanations, mathematical explanations, or metaphysical explanations. See Malle (2004), Mancosu (Summer 2018 Edition), Woodward and Ross (Summer 2021 Edition), and Brenner et al. (Winter 2021 Edition) for more on the possibility of non-causal explanations. What I object to is the combination of formal explanations (categorizing X as part of kind Y) and the purported lack of causal information. Categorizing an instance in terms of a kind would not be genuinely explanatory – would not answer the “why” question – if the categorization did not imply some relevant causal information. [↑](#footnote-ref-3)
4. Lombrozo and Carey argue that people judge teleological explanations as acceptable explanations only when they reference functions that play a predictable causal role (Lombrozo and Carey 2006). Teleological explanations that do not meet this condition are regarded as not explanatory. We tend to regard those explanations as circular or uninformative. For reasons such as these, many theorists have concluded that teleological explanations often presuppose causal information (Lombrozo 2012, Piaget 1969, Gelman 2003, 2004). In the main text, I employ an analogous argument for formal explanations. [↑](#footnote-ref-4)
5. Though these are social explanations, many theorists would not regard these types of explanation as mentalizing. Mentalizing mechanistic explanations probably are rare. [↑](#footnote-ref-5)
6. Williams (forthcoming) argues that we should distinguish *Phenomenology* (with a big P) from *phenomenology* (with a little P). The former refers to philosophical discussions of the nature of conscious experiences, such as we see in Husserl’s writings. The latter consists in verbalization of conscious mental states. I will be discussing little-p phenomenology in this paper. [↑](#footnote-ref-6)
7. I do not want to overstate this claim. Clearly there is overlap in what people say about their conscious experiences. If individuals are looking at a large red circle on a wall, it is extremely likely that they would all provide a similar phenomenological report of their experience. The variability arises when subjects are asked more nuanced questions about the nature of their conscious experience – e.g., not just what emotion are you feeling, but whether and where in your body you feel this emotion – or they are asked about their experience of phenomena that are less central, like peripheral version or anxiety. The variability comes in the answers to these more difficult questions. [↑](#footnote-ref-7)
8. One could design violation-of-expectation experiments to test *some* elements of these reports. One could have subjects observe/participate in a social interaction and then self-report their conscious experiences, in particular what they are attending to. Then, one could replay the interaction, manipulating a scene to make some elements of the social interaction more obvious, and test whether looking times at the more prominent elements in the replay correlate with self-reports of attention. This method could test the accuracy of only some elements of phenomenological reports of social interactions, and even then, the results would likely be ambiguous and subject to numerous interpretations. [↑](#footnote-ref-8)
9. To clarify, I do not presuppose a hard line between conscious and non-conscious or between implicit and explicit. I take it the difference between conscious and non-conscious is one of degree, and that *conscious* and *non-conscious* represent two ends of a spectrum. On the view I adopt here, fully conscious mental experiences are at the center of our attention and broadly available for use by many cognitive systems. On the other end of the spectrum are fully unconscious mental experiences, which are not currently – or not ever – the target of our attention and thus are inaccessible to some higher-level cognitive systems. Fully explicit mental experiences are articulated or easily articulatable, in inner speech, verbally, visually, etc. Fully implicit mental experiences are difficult or impossible to articulate in these ways. On my view, the process and the product of explanation can be more or less the target of our attention and more or less articulatable. [↑](#footnote-ref-9)
10. See also Williams (forthcoming) for an application of Drayson’s distinction to the social cognition debate. [↑](#footnote-ref-10)
11. In fact, some would maintain that these capacities presuppose rudimentary mentalizing. [↑](#footnote-ref-11)
12. In other works, I argued that phenomenology plays other roles, as well (Spaulding 2015, 2018a). Specifically, phenomenology helps us carve out the phenomenon of social cognition, may be essential for studying the role of emotions in social cognition, and *weakly* constrains theories of social cognition insofar as these theories should not have false predictions about our phenomenological experience of social interactions. I argued that existing theories of mentalizing satisfy the weak constraint. [↑](#footnote-ref-12)
13. Though these are not behavioral studies, Zawidzki (2013) and McGeer (2007) give persuasive reasons to think that we frequently are motivated to intervene, shape, and regulate others’ mental states. Though Zawidzki regards this as an alternative to mentalizing, I argue below that mindshaping presupposes teleological explanations. It is difficult to change someone’s mind if you do not know what they are thinking. Thus, if it is true that we are frequently motivated to intervene on, change, and shape others mental states, this is good reason to think we frequently generate teleological social explanations. [↑](#footnote-ref-13)
14. See Westra (2017c) and Carruthers (2016) for a review of some of these findings. [↑](#footnote-ref-14)