Predicting other people shapes the social mind

Diana I. Tamir[1,2*] and Mark A. Thornton[3]

[1] Department of Psychology, Princeton University, Princeton, NJ 08540

[2] Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08540

[3]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755

* Address correspondence to:

Diana I. Tamir

dtamir@princeton.edu

Abstract

People have a rich understanding of the social world within which they are embedded. How do they organize this social knowledge? And to what end? We suggest that these two questions are intimately linked. We review a burgeoning literature which shows how the social mind organizes different layers of social knowledge – including knowledge of actions, mental states, personality traits, situations, and relationships – into parsimonious low-dimensional maps. By distilling much of the complexity of the social world down to coordinates on a few key psychological dimensions, people construct a highly efficient representation of the social world. We go on to review recent research showing that these maps facilitate accurate, automatic prediction of real-world social dynamics. Specifically, the placement of stimuli within these maps implicitly encodes predictions about the social future, both within the same layer of social knowledge, and across different layers. Moreover, the ability of these maps to predict the social future is no coincidence: increasing evidence suggests that the goal of prediction actively shapes the way people organize social knowledge. We conclude by discussing challenges and future directions for studying the predictive social mind.

1. Predicting other people shapes the social mind

A person you've never seen is approaching you. As they walk toward you, your mind rapidly assesses their appearance - their demographic categories, facial expression, and attire (Hester & Hehman, 2023; Willis & Todorov, 2006). Your mind also flies into the future, automatically trying to predict what this person will do. Will they walk right past you? Will they stop and talk to you? Will they treat you like a friend or foe (Cuddy et al., 2008)? How should you respond in return?

People spend at least half of their waking hours in the presence of other people (United States Bureau of Labor Statistics, 2003), meaning that the mind must constantly be prepared to perceive and respond to others' future thoughts, feelings, and actions. Even when one is not actively interacting with others, the social mind spends the bulk of its time spontaneously thinking about other people, anticipating future interactions, and preparing its own actions accordingly (Mar et al., 2012; Song & Wang, 2012).

The mind is tailored to the problem of predicting the world around it. The brain does not passively receive information from the environment. Instead, it actively predicts what will happen, across diverse domains (Hohwy et al., 2008; Rao & Ballard, 1999; Vuust et al., 2009). For example, when watching a ball roll down a hill, people will automatically predict how soon it will reach a street crossing, and how fast they need to run in order to intercept it. When listening to a piece of music, people automatically predict when the chorus will return, or how a chord progression will resolve. When reading, people use the words at the beginning of a sentence to automatically predict the end of that sandwich. Sentences like the preceding one demonstrate how predictions constantly shape our experience of the world: because of the way that sentence was written, you may have expected it to end in the word "sentence" rather than the phonetically similar but semantically disparate “sandwich.” Any surprise or confusion you felt might be the phenomenology attributable to a prediction error. Learning to predict the dynamics of the external world is one of the key goals of early brain development (Emberson, 2017; Richardson & Saxe, 2020).

Learning to predict the *social* components of the external world is a particularly consequential ability for human beings. Imagine trying to please a boss whose mood was unpredictable, trying to court a partner with completely unpredictable preferences, or trying to be friends with someone whose location you could never anticipate. Lacking these insights would make it nearly impossible to build meaningful, satisfying, or productive relationships. However, accessing these insights comes with unique challenges. People are *not* merely physical objects, moved solely by deterministic external forces like gravity. People are self-determined, moved by invisible forces of thought, feeling, and personality that spring from within their own minds. Fortunately, the social mind is well-prepared to predict other people (Tamir & Thornton, 2018a; Thornton, Weaverdyck, & Tamir, 2019b; Thornton & Tamir, 2021b). How does it accomplish these feats of foretelling?

There are two main challenges that the social mind must address in order to successfully navigate the social world. First, people must extract the social information most relevant for making accurate social predictions. Not every detail of a person’s identity or appearance will be equally relevant to their future thoughts, feelings, or behaviors. How do people distill the richness of information in the social world, and represent it in a computationally efficient manner? Second, people must use this knowledge to make predictions. How do people transform the relevant information into a social prediction? In this chapter, we present a multilayered framework of social cognition that explains how people solve the challenges of representing and predicting other people, respectively – the Multilayer Atlas of Person Prediction (MAPP).

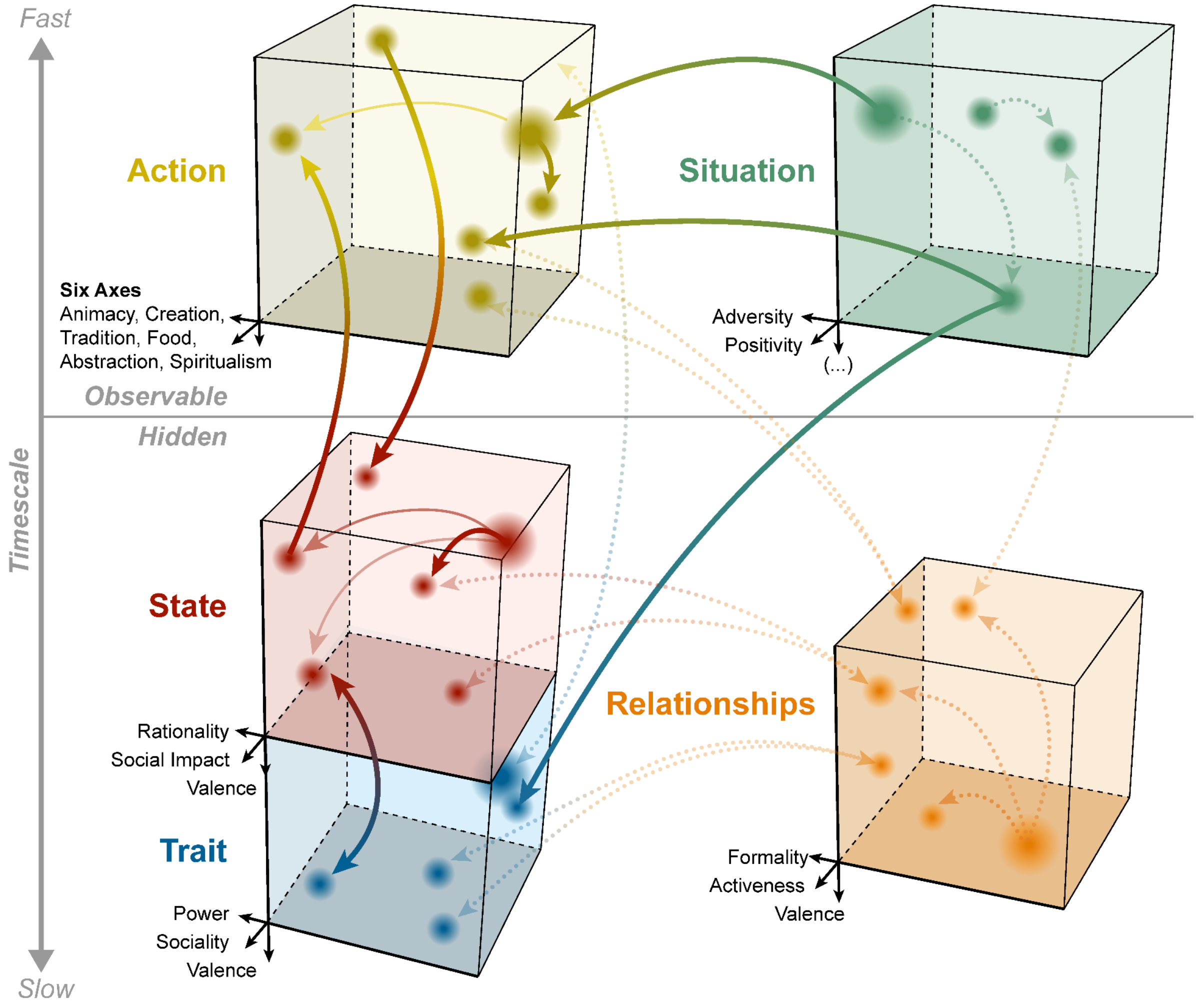

Figure 1. The Multilayered Atlas of Person Prediction (MAPP). This framework for social cognition posits that people make social predictions by representing multiple domains of social knowledge using efficient low-dimensional representational spaces. These locations of stimuli within these maps implicitly encode real-world social dynamics, allowing for accurate, automatic prediction of the social world.

This framework solves the first challenge – representing the social world – by carefully characterizing the simple structure of multiple domains of social knowledge. Specifically, it suggests that the mind represents social information using a series of conceptual "maps." These maps distill the richness and complexity of social information to just the few, key components that will afford the most effective social prediction. The MAPP framework has so far comprised three conceptual maps, representing actions, mental states, and traits (Figure 1). The action map captures observable social information about others' behaviors. The mental state map captures the hidden movers of behavior within a person's mind - feelings, thoughts, and moods – such as excitement, skepticism, or gloom. Finally, the trait map captures aspects of a person's identity that remain stable over time, such as their personality tendencies towards extraversion, competence, or optimism. Maps of additional layers of social knowledge, including knowledge of relationship types and situations, are also starting to be incorporated into the MAPP framework (Figure 1). These latter domains have been taxonomized, as we discuss in the next section, but work on whether and how they predict their own futures has not yet been published. Each map is defined by a small number of psychological dimensions - the latitude and longitude that allow each action, state, or personality trait to be described and understood via a parsimonious set of coordinates.

The MAPP framework solves the second challenge – predicting the social future – by characterizing the dynamics of how actions, states, and traits tend to move within and across these maps. A person first localizes another person's current mood to a set of coordinates within the mental state map or maps their identity to a set of coordinates within the trait map. Each location on a map, in turn, implies a set of likely trajectories. That is, because people move efficiently and predictably through these spaces, each map entails *what* information people need to use to make social predictions, and *how* they leverage it to predict the social future. In this way, the map shows social navigators the usual paths people take through it – the roads, bridges, and tunnels most well-trod. Here we outline the principles people use to predict where others will move within and across maps based on their current location.

In this review, we present the low-dimensional maps for actions, states, traits, relationships, and situations that people use to navigate the social world. We show how people use the structure of these maps to make accurate, automatic social predictions, both within and across maps. We then discuss how these knowledge structures develop, in both human and machine learners. Next, we discuss how social maps and resulting predictions vary from person to person, and the consequences this variation can have. Finally, we consider the MAPP framework as a whole, discussing meta-dimensions that span the layers of social knowledge, how social prediction and causal reasoning complement each other, the methodological advances that facilitated the framework's development, and how it can advance future research and real-world applications.

2. Charting maps of social knowledge

The world population just exceeded 8 billion people. Each of those people knows hundreds or even thousands of unique individuals, on average. On a given day, a person may pass countless strangers on the street, scroll through endless thoughts shared on social media, or enjoy deep conversations with close friends. Each of these strangers, contacts, and friends buzzes with thoughts and feelings, and performs myriad actions. Engaging in this vibrant social world is essential for well-being. Yet, this vibrancy poses a real challenge: how do people keep track of so many people? How does the mind bring order to the complexity of the social world?

We propose that people reign in the richness of the social world by organizing social knowledge using parsimonious conceptual "maps" of social stimuli (Figure 1). Ordinary maps allow people to denote locations in physical space using just two numbers: coordinates on the continuous dimensions of longitude and latitude. Locating a city using these simple dimensions immediately reveals a wealth of information about it. For example, you may learn that a city is near the Pacific Ocean and far from the pollution of Los Angeles, meaning that it would be a fantastic place to visit for a beach vacation. By using social maps, the social mind can likewise reduce the complexity of the social world in a useful and computationally efficient manner. These maps also allow people to generalize what they already know about the social world to new actions, states, traits, relationships, or situations. For example, a person may assume that a nearby beach town would offer a similar experience to another one they had previously visited.

A successful map is parsimonious. The more efficiently the mind can summarize social knowledge, the better, all else equal. The brain is a metabolically greedy organ, demanding far more than its fair share of resources, such as oxygen and glucose, than other body tissues (Magistretti & Allaman, 1999). The parts of the brain typically implicated in social cognition are particularly metabolically demanding, because they are generally more active at "rest" than when engaged in an active task (Raichle, 2015). The more efficiently these regions can process and represent social information; the better-positioned someone will be to survive and thrive in a resource-limited environment. Thus, the mind must choose only the most useful features of the social world to depict in its maps. Below we describe the key dimensions that scaffold maps of social knowledge.

Just as a city planner uses multiple maps to represent the different layers of municipal infrastructure - roads, rails, power, water, and so forth - so too do social perceivers use multiple maps to comprehend the social world, one for each type of social knowledge (Figure 1). So far, work by social psychologists and neuroscientists has charted maps of several key layers of social knowledge, including actions, mental states, traits, relationships, and situations. Each map layers on top of the next. The action map offers quick, observable information about *what* is happening. The state map offers information about the hidden reasons *why* a person acts. The trait map tells us *who* is behaving. The relationship map tells us what types of *interactions* to expect between any pair of people. Finally, the situation map characterizes *external* influences on behavior, and the *affordances* offered by the environment. Together, these maps offer a rich, multilayered view of one's current social world. So, what do these maps look like?

2.1 The action map

Actions are the touchstone of human social life. Actions are the only dynamic component of a person that can be directly observed. We cannot see thinking or feeling; a person can think and feel within their mind endlessly, but these secret contemplations will only have an impact when that person translates them into action. Thus, action knowledge is a crucial foothold for understanding the social world. However, the action domain is immense. Humans have a vast behavioral repertoire. People can engage in physical actions, such as jumping and running; people can engage in complex actions, such as cooperating and leading. If we use verbs as a proxy for actions, humans can engage in over 10,000 actions. How do people take in information about others' actions and use it to interact with them successfully?

The mind meets this challenge by mapping actions into a 6-dimensional space. Instead of individually representing each action in an entirely atomistic way, the mind represents actions conceptually by placing them on each of these six key dimensions. Specifically, people tune into the following dimensions when thinking about others' actions: the Abstraction, Creation, Tradition, Food, Animacy, and Spiritualism Taxonomy, or ACT-FAST (Thornton & Tamir, 2022).

Each dimension in this map captures one feature that is particularly important for our social minds to grasp when understanding actions. **Abstraction** captures whether an action is abstract and social or concrete and physical. Actions such as protesting and governing are highly abstract; actions such as bicycling or gardening are highly concrete. **Creation** captures whether an action is related to creation or crime. Actions such as filming and singing are high in creation; destroying and stealing are low in creation. **Tradition** captures whether an action has been practiced for generations or is a relatively recent innovation. Cooking and decorating are highly traditional actions; encrypting and welding are highly innovative. **Food** captures whether an action is related to food or not. Actions such as baking and eating are highly food-related; sleeping and jogging are unrelated to food. **Animacy** captures whether a human or animal agent acts or an inanimate force such as weather or machinery does. Actions such as meowing and flossing are highly animate; malfunctioning and extracting are more mechanical. **Spiritualism** captures whether an action is related more to work or worship. Actions such as repenting and preaching are highly spiritual; faxing and haggling are low in spiritualism.

As soon as one places an action within this map, one immediately gains access to a world of knowledge about that action. This map offers rich conceptual knowledge about any location within it. Moreover, this taxonomy predicts practical features of actions (Thornton & Tamir, 2022). For example, the information available across the dimensions in ACT-FAST can predict whether an action is likely to be performed in the morning, afternoon, or night; inside or outside; publicly or privately; because of approach or avoidance motives; with physical or mental effort; with the head arm or leg; and by a person who is (or is not) powerful, positive, or extroverted. ACT-FAST simplifies the space of actions by distilling them down to their most practical, socially-relevant features. In doing so, this map can tell socializers when, where, why, how, and by whom an action is performed.

In addition, the ACT-FAST map allows people to easily apply their existing action knowledge when learning about completely novel actions. By placing a new action within this map, a person learns how far it is from other actions. The closer one action is to another, the more similar they are; the further one action is to another, the more dissimilar they are. For example, freeze-drying and sun-drying are nearby on the food dimension and, thus, could both occur in food preparation situations. However, they are far apart on the tradition dimension and, thus, are likely to be enacted by different people at different times. Even with very little knowledge about sun-drying, knowing its location on these two dimensions, and its proximity to other actions, would be highly informative.

Indeed, information about similarity was critical to initially deriving the structure of people's action knowledge (Thornton & Tamir, 2022). Researchers started by determining the dimensions that could reliably capture the similarities between action words (i.e., verbs) in natural language using a data-driven analysis over large-scale text corpora. These dimensions generalized across multiple large text corpora, predicted human judgments of how similar actions are to each other, and predicted patterns of brain activity elicited by observing actions in a naturalistic functional MRI experiment. Together, these findings suggest that ACT-FAST provides a valid, valuable, and biologically plausible map of how people organize knowledge about other people's actions.

2.2 The mental state map

Mental states are the hidden movers of people's actions. Behind each person's eyes exists their hidden inner world, a buzzing internal mental life filled with thoughts, feelings, and perceptions that define their unique perspectives. One of the great gifts of human sociality is that people can see into these internal worlds, inferring, like magic, the invisible goings on of another person's mind. For example, we can intuit that a person who isn't cooperating is likely feeling grumpy; or that a glassy-eyed person is not paying attention. What map do people use to navigate mental states?

Psychologists and philosophers have long theorized about which dimensions the human mind uses to represent others' mental states. This work offers at least seven dimensional theories about the psychological content that defines mental state representations. These theories include valence and arousal (Posner et al., 2005; Russell, 1980), warmth and competence (Cuddy et al., 2008; S. Fiske et al., 2002), agency and experience (Gray et al., 2007), emotion and reason, mind and body (Forstmann & Burgmer, 2014), social and nonsocial (Britton et al., 2006), and uniquely human and shared with animals (Waytz et al., 2010). These dimensions, though derived independently, share many common insights. Indeed, as a collection, they can be well-described by a lower dimensional space. Using dimension reduction techniques, researchers have found that these psychologically derived dimensions can be captured fully by three orthogonal dimensions: rationality, social impact, and valence (Tamir et al., 2016), which together comprise the 3d Mind Model (Thornton & Tamir, 2020a). **Rationality** captures whether others feel thoughtful or emotional. States such as planning and deciding are high in rationality; embarrassment and ecstasy are low. **Social impact** captures whether mental states arise during intense social interactions or low-energy solo experiences. States such as dominance, friendliness, and lust are high in social impact; sleepiness and pensiveness are not. Finally, **valence** captures whether a person is feeling good or bad. Positive states include affection and satisfaction; negative states include disgust and disarray.

People spontaneously use the 3d Mind Model dimensions to think about others' mental states. For example, these dimensions capture the patterns of activity evoked by thinking about others' mental states. Specifically, the dimensions of rationality, social impact, and valence explain over 80% of the shared variance in neural representations of mental states (Thornton & Tamir, 2020a). That is, they provide a near-complete map of mental states, capturing the inherent complexity of other people's minds in a computationally efficient manner. Researchers demonstrated this by using representational similarity analysis (RSA), an analysis that how psychological dimensions structure neural representations (Kriegeskorte et al., 2008), on using functional neuroimaging data. RSA measures the extent to which a psychological dimension can predict neural patterns of activity based on their similarity. To illustrate: the dimension "emotion" predicts that "ecstasy" and "rage" are represented very similarly in the brain because both are similarly emotional. In contrast, the dimension "valence" predicts that "ecstasy" and "rage" are represented very differently because one state is positive, whereas the other is negative. RSA tests these predictions concurrently by measuring the extent to which distributed patterns of neural activity elicited by thinking about a person in ecstasy are similar or dissimilar to those elicited by thinking about a person in a fit of rage. Each dimension makes predictions about the similarity of each mental state compared with every other. RSA assesses the accuracy of all of these predictions simultaneously. This way, researchers tested *which* dimensions people's mind uses to map mental states, and found that the 3d Mind Model captures the dimensions people are attuned to when thinking about others' mental states (Tamir et al., 2016; Thornton & Tamir, 2020a).

People wield the 3d Mind Model when thinking about a wide range of people's minds. For example, people use the same map when thinking about their own mental states as they do when thinking about the states of a close other *and* a more distant other (Weaverdyck et al., 2021). People also wield this map similarly when prompted to think about mental states using a variety of modalities, from visual to verbal. This map predicts not only neural activity, but also both explicit and implicit behavioral measures of the similarity between mental states, as well as the similarity of mental state word meaning derived from computational text analyses (Thornton & Tamir, 2020a). The 3d Mind Model thus offers people a powerful, validated map for thinking about others' minds. But to which 'people' does it apply?

The 3d Mind Model was initially developed and validated using data from a thin slice of society: the "WEIRD" (Henrich et al., 2010) contemporary population in the United States. To what extent does this map of mental states apply more broadly to other cultures? Researchers tested the cross-cultural generalizability of this model by analyzing large text corpora from a wide range of modern and historical societies (Thornton et al., 2022). In particular, they examined 57 modern countries around the world as they tweeted in English in over 163 million tweets; 17 unique languages, using billions of words from across the Internet; and texts from four unique historical societies spread over more than 2000 years of history, including ancient Chinese and Jewish texts. The 3d Mind Model predicts how similar and different mental state words are, given their distance on each dimension. In this cross-cultural study, researchers used word embeddings - or the locations of words in a high-dimensional linguistic space - to test if a language treats the similarities and differences between two mental state words in the way the 3d Mind Model would predict. The 3d Mind Model predicted mental state word use in 100% of the cultures examined, across countries, languages, and history. Indeed, it could explain over half of the reliable variance in mental state language use on average, suggesting that the 3d Mind Model offers a powerful, generalizable map of how a wide variety of cultures conceptualize thoughts and feelings.

The findings of other cross-cultural studies generally cohere with these results. Although individual mental state concepts are cross-culturally variable, the broad dimensions that shape the map are largely shared across cultures (Jackson et al., 2019; Russell & Lewicka, 1989). This may reflect convergent cultural-evolutionary pressures which favor the emergence of certain dimensions, such as valence, in many different human societies (Heyes & Frith, 2014; Lindquist et al., 2022). Importantly, cultural universality and diversity are not incompatible. Even when the basic dimensions of the map are the same, their relative importance, or the locations of particular states on each dimension, may differ from one culture's version of the map to another. In this way, the 3d Mind Model provides a common framework within which to quantitatively compare the ways in which cultures may differ in their understanding of mental states.

2.3 The trait map

Each individual is a unique landscape of thoughts, feelings, and behaviors. People can gain traction with new individuals by deploying a single map of mental states and a single map of actions described above. However, as people get to know an individual, they learn more about that person's personality traits and emotional or behavioral tendencies. Learning *who* a person is entails gaining insight into these more stable personality features (Roberts & Mroczek, 2008; Srivastava et al., 2003). What map do people use when thinking about others' personality traits?

Psychologists have long pursued the 'correct' map of personality traits. This work offers several prominent theories about the dimensions that matter when people think about others' traits. These dimensions include Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, which together comprise the Five Factor model of personality (Goldberg, 1990;

McCrae & Costa, 1987), as well as the dimensions of warmth and competence from the stereotype content model (Cuddy et al., 2008; S. Fiske et al., 2002), trustworthiness and social dominance from face perception (Oosterhof & Todorov, 2008), and agency and experience from mind perception (Gray et al., 2007). These 11 dimensions derive independently from different psychological traditions. However, as with mental states, these diverse lines of work may have converged on a smaller, shared set of key features that people tune into when considering who another person is. Again using dimension reduction techniques, researchers found that these psychologically derived dimensions can be captured fully by three orthogonal dimensions: power, valence, and sociality (Thornton & Mitchell, 2018). **Power** captures whether a person is dominant and agentic, thus likely capable of exerting their will on others. **Valence** captures whether a person is warm, trustworthy, and agreeable, and is thus likely to help others. This dimension resembles the morality dimension in other taxonomies (Brambilla et al., 2021). Finally, **sociality** captures whether a person is extraverted, and thus likely to engage with the social world.

These three dimensions – power, valence, and sociality – capture the features that people are attuned to when thinking about others' traits (Thornton & Mitchell, 2018). Researchers first demonstrated this in a study where participants thought about a wide range of public figures while undergoing functional neuroimaging. Each of the public figures varied on a wide range of personality features. Researchers tested how well these trait dimensions captured distributed patterns of brain activity when thinking about others' identities. These three dimensions alone explained nearly two-thirds of reliable variance in neural patterns elicited by thinking about these targets. This trait map outperformed alternative maps of trait space, including the Big 5 personality traits. People tune into similar dimensions when thinking about novel people as about famous figures albeit to a different extent (Kim et al., 2022). This map thus offers the most comprehensive view of the dimensions people spontaneously encode when thinking about others' traits.

The dimensions of the trait map may help characterize groups and collectives, as well as individuals. Trait dimensions such as those of the stereotype content model (Cuddy et al., 2008; S. Fiske et al., 2002) have been applied to characterize both individuals and groups. People humanize groups, reasoning about their collective mental states using much of the same mental and neural machinery they apply to individual minds (Alt & Phillips, 2022; Contreras et al., 2013). However, there are also distinct neural correlates of mentalizing about groups. This suggests that trait and group knowledge maps may overlap partially but not entirely. What unique dimensions might groups have that individuals lack? Here we offer some potential candidates. For example, cohesion is a key characteristic of groups: it predicts much about how they will behave, and it lacks a homolog at the individual level (Greer, 2012). Recent research characterized social network knowledge in terms of two dimensions – valence and social distance – and six types of ties between people – admiration, closeness, socializing, active conflict, passive conflict, and contempt (Genkin et al., 2022). We hope to see future research clarify the relationship between individual and group traits, and identify the remaining key dimensions of group knowledge.

2.4 The relationship map

Human behavior depends not only on the characteristics of individuals, but also on the network of social relationships that connect those individuals. Actions that are awkward or impossible alone can become effortless and elegant with the right person. The laugh of a loved one can instantly brighten a sour mood. A generally reserved person can overflow with gregariousness when in the company of close friends. Knowing whether two people are connected, and more importantly, *how* they are connected, is vital to social life. Imagine the headaches of a dinner party planner who inadvertently seats bitterly divided exes next to one another, or the satisfaction of a manager who gets to assign close collaborators to work together. People require a map of relationship knowledge that can characterize the full diversity of dyadic relations in a parsimonious way. The fields of social psychology, sociology, and anthropology have each offered relationship taxonomies, though they have yet to reach consensus on a single validated map for this domain.

In anthropology, relational models theory characterizes dyadic interactions in terms of their resource exchange logic (Haslam, 2004). This theory consists of a four-category typology: **communal sharing**, in which resources are exchanged freely without the expectation of repayment, such as between a parent and child; **authority ranking**, in which one person holds a position of authority above another, and dictates the use of resources; **equity matching**, in which the quid pro quo reciprocity is the dominant logic of resource exchange; and **market pricing**, characterized by the complex allocation of resources via market pricing facilitated by money. Relational models theory is a useful characterization of dyadic resource exchange, but does not necessarily reflect the cognitive model people use to think about relationships.

Social psychology has charted a four-dimensional map of relationship space that may offer a closer approximation of how people conceptualize relationships (Wish et al., 1976). The dimensions of this map span from **cooperative** to **competitive, equal** to **unequal, intense** to **superficial,** and **formal** to **informal**. This research has had an enduring impact, but it could benefit from an update as it approaches the half-century mark. For example, new attempts to map relationship space could use larger and more culturally diverse samples of participants and stimuli to ensure insight into every corner of relationship space. Moreover, techniques such as functional neuroimaging or computational text analysis could be used to generate convergent evidence for the results obtained solely from explicit judgments in the past. Together, these approaches may give us a new and improved understanding of how people organize relationship knowledge.

2.5 The situation map

Relative to other animals, to plants, and certainly to inanimate objects, human behavior seems uniquely endogenous. People often act in ways that simple stimulus-response relationships cannot explain. However, despite being so internally-driven, people are still controlled by external situational constraints. Indeed, the power of the situation to shape behavior was central to the emergence of social psychology as a subfield (Ross & Nisbett, 2011). The ability to understand situations – to read the room, draw upon social norms and scripts, and behave accordingly – are all critical parts of everyday social life.

Situations are simultaneously more concrete and more elusive than people. On one hand, many features of a situation – such as place, time, and objects – are directly observable while the states and traits of the human mind are not. On the other hand, situations are intangible; one can reach out and touch a person, and know that they are a discrete entity, in a way that one cannot do to a situation. What are the elements or features that define a situation?

There are currently two prominent models of situations: the DIAMONDS model (Rauthmann et al., 2014) and the CAPTION model (Parrigon et al., 2017). Both models highlight the dimensions of **positivity, negativity,** and **adversity** as key to understanding situations. In addition, DIAMONDS suggests that people consider factors including **duty, deception, mating, and sociality** when thinking about situations, and CAPTION suggests that people consider

complexity, typicality, importance, and **humor**. These differences may stem from the fact that DIAMONDS was derived from people's judgments about episodes from their daily lives; CAPTION was derived from lexical analysis. Future work using methodologies such as neuroimaging may help to further hone and validate a single, biologically plausible map for these theories of situation knowledge.

2.6 Completing the MAPP framework

This section has described recent research mapping the organization of different layers of social knowledge. How close are the maps described here to their ideal forms? That is, do these maps really capture *everything* that's happening in each layer of social knowledge? The answer varies by layer. The 3d Mind Model, for instance, has been extensively validated across multiple studies, using multiple methodologies. It captures the vast majority of variance shared across individual brains, and more than half of the variance across cultures (Thornton et al., 2022; Thornton & Tamir, 2020a). In this case, we anticipate the future will consist more of minor refinements than wholesale overthrow. In other domains, the situation differs. For example, dimensional taxonomies are relatively new in the action domain. ACT-FAST was one of the first, and thoroughly grounded action taxonomy, and we expect it to facilitate the emergence of newer and better theories of this layer of social knowledge (Thornton & Tamir, 2022). Other domains have languished for some time. For example, there has been little work taxonomizing relationship knowledge – at least within social psychology since Wish (1976). This may reflect the extent to which that work was ahead of its time, but we suspect that recent methodological innovations may help to update our understanding of this domain.

The models we discuss here have limitations. First, they cannot represent propositional content (Saxe, 2018). The 3d Mind Model captures the general concept of happiness, but it cannot represent what a person might be happy *about* at any given moment. Likewise, trait models capture impressions of a person, but cannot encode arbitrary propositional content like a birthdate. Thus, content such as specific beliefs about, or attitudes towards, certain people are not well captured by these models. Second, these models assume that the psychological domain in question is a smooth, continuous space. In such spaces, the properties of any social stimulus can be interpolated from the properties of nearby neighbors. This may well be true of many social stimuli, but it is not guaranteed to be true of all of them. For instance, logical operators behave create spaces that are not susceptible to interpolation. Thus, there may be limits to what these maps can encode. Third, these maps – as conceived of here – are metric. Thus, they must satisfy the requirements of metric spaces. This places limits on their capacity, in that they cannot capture phenomena such as asymmetric judgements of similarity, which violate metric assumptions, but occur in human cognition (Tversky, 1977). Finally, the current dimensions are those with strong evidence *across* people. These are the dimensions that most minds agree are useful when thinking about others' states, traits, and actions. However, individuals could also use maps that incorporate additional dimensions. An individual may find a particular dimension useful when thinking about certain targets; or a community or culture might use dimensions specifically relevant to their group. We have yet to identify the idiosyncratic dimensions of these maps. Ameliorating these limitations is an important direction for future research.

3. Social maps predict the social future

In the last section, we described the maps that characterize multiple domains of social knowledge. These low-dimensional structures provide parsimonious descriptions of how people represent others' actions, mental states, personality traits, relationships, and situations. Why has the social mind settled into using these particular maps of social knowledge? Human social knowledge is not a random collection of files. The *usefulness* of this information for achieving social goals should shape the knowledge people retain, and the form in which they retain it.

In the non-social domain of vision science, research has demonstrated that two simple goals – efficiency and accuracy – shape visual representations in the brain (Olshausen & Field, 1996). Efficiency meant that the simulated neurons should expend as little energy as necessary via their activity. Accuracy meant that the representation encoded by these neurons should resemble the original image as closely as possible. These two goals - in combination with the visual statistics of natural scenes - are sufficient to produce simulated receptive fields that mirror those in primary visual cortex. The same logic applies further along the visual processing pathway, as well as to other domains such as audition and language (Doshi et al., 2022; Goldstein et al., 2022; Schrimpf et al., 2020; Yamins & DiCarlo, 2016).

So, what are the primary computational goals the social brain is optimized to accomplish? Prior work in social psychology has offered explanations for the computational goals achieved by individual social dimensions. For example, the stereotype content model's dimensions of warmth and competence may serve as a basic threat detection module (Cuddy et al., 2008). Assessing another's warmth tells you whether they are more likely to help you or harm you; assessing their competence tells you how effective they are likely to be in either endeavor. Others have argued that these dimensions encode a fundamental division in social roles: gender with warmth representing stereotypical femininity and competence representing stereotypical masculinity (Martin & Slepian, 2021). By placing others' along these dimensions, one could thus determine the types of stereotypical behaviors one might predict they will perform. The 3d Mind Model may indicate state-level threat, just as the stereotype content model may indicate trait-level threat. Other characterizations of mental states, such as the Circumplex Model, have likewise been assigned other functional purposes.

Each of these accounts provides a valuable, domain-specific explanation for the emergence of certain socially important psychological dimensions. We propose an overarching goal unites them: the goal of accurately predicting the social future. That is, accurate social prediction is the key goal that constrains the structure of social knowledge, much like it constrains the representations present in the visual system. This perspective aligns with the broader theoretical paradigm of predictive coding (Friston & Kiebel, 2009; Huang & Rao, 2011). This paradigm rests on the hypothesis that the brain does not passively perceive the outside world, but rather actively predicts its own inputs. The brain then compares these predictions to incoming sense data. The discrepancies between them – prediction errors – are then passed back up the processing hierarchy to keep the brain's model of the world in touch with reality. The process is highly efficient because passing the small prediction errors along can be far more efficient than transmitting the entire incoming signal. This paradigm also emphasizes accuracy with respect to a clear target: the future.

Inspired by work on predictive processing in more perceptual domains, we and others have suggested that it might also provide a high-level description of the social mind (Koster-Hale & Saxe, 2013; Tamir & Thornton, 2018b). That is, the social mind may organize social knowledge to afford accurate, efficient predictions of the social future. In the remainder of this section, we discuss research demonstrating that the structure of social knowledge indeed supports within-domain prediction. For example, the structure of action knowledge supports predicting future actions from present actions, and likewise with mental states and traits. As of yet, we know of no working investigation how situations predict situations, or relationships predict relationships – or linking such predictions to their respective taxonomies – so we omit these domains from this section. In the following section, we discuss how the structure of social knowledge facilitates prediction across domains, such as predicting mental states from personality traits. Together, the results we describe corroborate the hypothesis that social knowledge is organized in the service of the predictive social mind.

3.1 Action prediction

The ability to anticipate others' actions plays a key role in social interactions (Bach & Schenke, 2017). In cooperative activities, for example, one must predict what one's interaction partners will do in order to plan one's own actions accordingly. This could entail synchronizing with them or taking complementary actions as the situation demands. Likewise, competitive social activities require people to foresee their opponents' moves in order to forestall them. The more deeply one can peer into this future, the more effectively one can compete (Holding, 1992). Difficulties in making accurate social predictions can lead to functional impairments in everyday life (Von Der Lühe et al., 2016). Humans possess a uniquely vast behavioral repertoire (Changizi, 2003), and as a result, opportunities for misprediction are prevalent. Fortunately, people can draw upon many different sources of information to guide their action predictions. Here, we focus on how people use a current action to predict future actions.

Actions do not occur in random order. People routinely plan and execute carefully structured sequences of actions to achieve their goals (Dezfouli & Balleine, 2012; Haggard, 1998). For example, one must design a study, collect data, analyze it, and then write it up – in that order – to produce an empirical publication. Sometimes these sequences are forced upon us by logical necessity: one cannot collect data for a study one has yet to design. Some sequences unfold to maximize the probability of achieving one's goal. Some sequences unfold as a matter of social convention, as in the case of social scripts, or normative action sequences (Abelson, 1981). Other cultures may follow different norms around sequences of gastronomic actions. For example, in the USA, salads are typically eaten at the start of a meal, before the main course, whereas in Europe, salads are traditionally eaten after the main course. Finally, some actions are more likely to occur in close temporal proximity to each other, regardless of the order in which they occur. For example, if you see someone kicking a soccer ball, you could accurately guess that you will also see them tackle a soccer player. Kicking and tackling tend to co-occur within the same context (i.e., a soccer game), though there is no prescribed order in which these activities must happen. All of these factors mean that actions occur in regular, predictable sequences and clusters. Indeed, the transition probabilities between both human and nonhuman actions are reliable, albeit remarkably complex in structure (Berman et al., 2016; Desmurget & Grafton, 2000; Hudson et al., 2016). Knowledge of these transition probabilities can facilitate action prediction in a wide range of contexts.

People hold accurate, explicit knowledge of the transition probabilities between actions (Thornton & Tamir, 2021a). Infants and toddlers make accurate predictions within highly restricted sets of actions, and adults are proficient in learning a restricted set of artificial transition probabilities (Baldwin et al., 2001; Monroy et al., 2017, 2019; Saylor et al., 2007). Adults are also highly accurate at predicting real-world action dynamics. For example, the American Time Use Survey is a large, nationally representative survey where people report activities they engaged in over a day (United States Bureau of Labor Statistics, 2003). These data provide insight into the *actual* transition probabilities between actions at coarse and fine-grained levels. People are highly knowledgeable of these transition probabilities. When asked to rate the likelihood that someone currently “working” will next be “socializing” or someone currently "eating" will next be "walking," people's estimates match the ground truth data remarkably well. People are likewise remarkably accurate at predicting action sequences cataloged from other diverse sources, such as actions depicted in movie scripts, semi-naturalistic sequences of actions in how-to instructions from Wikihow.com, and visually decoded action sequences depicted in Google's Atomic Video Actions dataset. People are highly attuned to sequences of actions, from moment-to-moment changes to changes over hours. People's knowledge of action sequences is also highly nuanced, understanding both which actions are likely to occur together and which actions are likely to proceed vs. follow each other.

Notably, the structure of action knowledge – operationalized in the form of the ACT-FASTaxonomy discussed above (Thornton & Tamir, 2022) – facilitates these accurate predictions. How do social navigators translate knowledge structures into predictions about dynamics? Distance within the ACT-FAST space directly translates into transition probabilities between actions. Actions nearby on the ACT-FAST dimensions have high transition probabilities. For example, eating and drinking are nearby on the Food dimension and are highly likely to proceed from one to the next. In contrast, actions far away on the ACT-FAST dimensions have low transition probabilities. For example, eating and faxing are far away on the Food dimension and are highly unlikely to proceed from one to the next. Arranging actions this way eases the memory burden of encoding every possible pair of action transitions. Instead, people can simply encode each action's location on just six dimensions. Then, when called upon to predict actions, they can compute the distance between them on the fly to determine how likely any one action is to precede or follow another.

The relation between distance and transition probabilities shows up in several ways. First, ACT-FAST proximity predicts people's estimates of transition probabilities. This indicates that people use this map as a mental model of action dynamics. Second, ACT-FAST proximity predicts ground truth transition probabilities. This indicates that ACT-FAST captures real-world action dynamics. Finally, ACT-FAST proximity accounts for the shared variance between rated and real transition probabilities. Together, these data suggest that the structure of action knowledge is uniquely suited to facilitate action prediction by encoding action dynamics as proximity within a conceptual space.

Different dimensions help explain different types of predictions. . For example, the food dimension is particularly important in predicting actions at a coarse level, as in predicting sequences over a day. In contrast, animacy is a relatively good predictor of action transitions on a much shorter time scale, on the order of seconds. This variation suggests that different ACT-FAST dimensions may describe transition probabilities at different temporal and conceptual granularities.

ACT-FAST not only captures people's explicit action predictions – it also captures naturalistic perception of and implicit neural predictions of future actions. For example, when people passively viewed a naturalistic stimulus – the first episode of the popular TV series *Sherlock* (Chen et al., 2017) – their brains automatically coded each current action using the ACT-FAST coordinates (Thornton & Tamir, 2020b). That is, a multivoxel decoding model could predict not only the ACT-FAST coordinates of the action a participant was *currently* viewing based on the pattern of brain activity elicited by viewing it, but also *future* actions that participants had not yet seen. These neural predictions leapt up to three actions or about 13 s into the future. The ACT-FAST dimensions are uniquely suited to support action prediction, outperforming nearly all possible other models of action representation and prediction.

Together, these results suggest that people automatically and accurately predict others' future actions whenever they observe their present actions. They also suggest that the structure of action knowledge supports this predictive coding of others' actions. This spatial map likely aids generalization as well. For example, if one learns about a new, unfamiliar action, one would not need to observe it within many sequences of actions to make an educated guess about its transition probabilities to and from other actions. Instead, simply inferring where to place the new action on the ACT-FAST dimensions would license an informed prediction about its transition probabilities.

3.2 Mental state prediction

People make accurate predictions of others' future actions based on their perception of others' current actions. Observable behaviors, however, are only one piece of the social pie that people must predict. People also devote significant time and energy to understanding the hidden movers of action: people's thoughts, beliefs, feelings, and intentions. How do people make such mental state predictions? Just as with action prediction, people also predict future mental states from current states and use the structure of mental state knowledge to do so.

Mental states, like actions, unfold in reliable and, thus, predictable ways. That is, one's current mental state reliably predicts the mental states that one is likely to experience in the future (Eldar et al., 2016; Kuppens et al., 2010; Sudhof et al., 2014). For example, if one is angry now, they will likely feel regret soon, and are unlikely to soon feel joy. Researchers have measured these real-world emotion dynamics using experience sampling methods (Sudhof et al., 2014; Trampe et al., 2015; Wilt et al., 2011), where people report on their momentary emotion experiences in everyday life. These reports provide insight into people's *actual* transition probabilities between mental states. People are highly attuned to these actual transitions. For example, when asked to rate the likelihood that someone currently experiencing “hunger” will next experience “exhaustion” or someone currently feeling “thoughtful” will next feel “calm”, people’s estimates match the ground truth data remarkably well (Thornton & Tamir, 2017). People are even more accurate when estimating mental state transitions than actions, despite the latter being objectively easier to observe.

Prediction is built into the neural hardware people use to think about mental states. In addition to being adept at making accurate, explicit emotion predictions, people also make emotion predictions spontaneously whenever they think about a mental state. Indeed, the predictive coding account suggests that the computational goal of mentalizing is to predict others’ future states – not just to perceive their current states (Koster-Hale & Saxe, 2013; Tamir & Thornton, 2018b). In line with this account, neuroimaging offers several convergent sources of evidence that the brain processes mental states with prediction in mind (Thornton, Weaverdyck, & Tamir, 2019b).

First, the brain responds differently to expected than unexpected sequences of mental states (Thornton, Weaverdyck, & Tamir, 2019b). When a person sees a sequence in an unexpected order, this elicits a prediction error signal; when a person sees a sequence proceed in an expected order, brain activity is reduced (Summerfield et al., 2008). For example, imagine seeing a happy face preceded by either another happy face or an angry one. If a brain region encodes facial expressions, it will be less active when the preceding face is happy since this would prime participants to expect another happy face (Ishai et al., 2004). This is exactly what occurs when the brain sees expected vs. unexpected sequences of mental states. Brain regions that process mental states, including the precuneus, become more active as they reconcile a mismatch between predictions and perceptions (Thornton, Weaverdyck, & Tamir, 2019b). That is, when the precuneus sees states occur in an order unlikely to occur in the real world, it registers a mental state prediction error.

Second, the neural representation of each state systematically resembles the states that are likely to follow it (Thornton, Weaverdyck, & Tamir, 2019b). That is, representations of each state include traces of all the states that it portends. This effect occurs throughout the social brain network, and focally in the dorsal medial prefrontal cortex, a region implicated in social cognition and predictive coding more generally (Brunec et al., 2018). The brain does more than just track the overall similarity between states and their likely future states. Using *asymmetric* representational similarity analysis, researchers have shown that neural pattern similarity tracks with asymmetrical transition probabilities. That is, representations of mental states were specifically similar to representations of other states that tended to *follow* them (i.e., predicted future states) rather than those that precede them.

Together, these findings suggest that people not only predict future mental states based on current mental states, but they also do so automatically, baking their predictions of the social future directly into their neural representations of the social present. By automatic, in this instance we mean that these predictions are (i) fast, happening in a few seconds, as most, (ii) goal independent, made even when unhelpful for the task at hand, and (iii) efficient, in that merely encoding a stimulus also makes predictions “for free”. It is plausible, *prima facie*, that these predictions are also automatic in other senses, such as being unintentional, below conscious awareness, or uncontrollable, but these features have not yet been directly tested.

As with actions, the structure of mental state knowledge – as embodied by the 3d Mind Model (Tamir et al., 2016; Thornton & Tamir, 2020a) – supports accurate emotion predictions. Distance in mental state space likewise translates directly into transition likelihood. All three dimensions of this model – rationality, social impact, and valence – correlate with both actual emotion transitions and with people's estimates of others' emotion transitions. Distance also mediates the shared variance between the two, suggesting that people use the 3d Mind Model space to encode mental state transition dynamics. Emotion dynamics and emotion knowledge seem highly entwined.

3.3 Trait prediction

Actions and mental states both vary over relatively short timescales. These rapid dynamics make it clear why it might be useful to predict them. In contrast, people's personality traits are relatively fixed, though they do change at a gradual and often imperceptible pace throughout the lifespan (Roberts & Mroczek, 2008). Given this slow rate of change, it hardly seems pressing to predict someone's future traits from their current traits. However, trait-to-trait prediction is nonetheless a useful ability, as we will explain.

Like trait change, the formation of trait trait impressions occur at a slower timescale than impressions of other domains of social knowledge like mental states or actions. Traits represent long-term response tendencies rather than deterministic behavioral rules. So, a single example of behavior is rarely sufficient to infer a trait unambiguously. Indeed, although people readily form first impressions in a split-second (Willis & Todorov, 2006), these thin-sliced judgments have little validity (Todorov, Funk, et al., 2015; Todorov, Olivola, et al., 2015). Instead, accurate trait impressions rely on extended observation periods across multiple settings. This process allows perceivers to disentangle this situational from the dispositional determinants of a person's behavior (Jones & Davis, 1965; Kelley, 1973). This slow impression formation process may be the surest route to accurately understanding another's personality. However, people cannot always wait months or years to get to learn another person's traits – nor do they need to.

Traits are not mutually independent. There are highly reliable correlation structures in the traits that people self-report and perceive in others (Cuddy et al., 2008; Goldberg, 1990; McCrae & Costa, 1987). These trait-trait associations make it possible to infer a given trait based on knowledge of a person's other traits, even if one has no direct diagnostic information about the focal trait in question. For example, if you think a person looks cheerful based on a facial first impression, and you think that cheerful people are typically cooperative, then you are likely to ascribe cooperativity to this target person, even without direct evidence. People capitalize on this approach to make trait-to-trait predictions: bootstrapping their way from relatively narrow insights to a broader view of others' personalities by way of the correlation structure between traits (Stolier et al., 2020). Thus, trait-to-trait prediction can indeed be useful – it's just that one tends to predict which unobserved traits are likely to co-occur with observed traits, rather than which traits transition to other traits over time.

Importantly, the structure of trait correlations is not universal, either in perception or reality. For instance, a single perceiver may hold different stereotypes about different groups of people, and these stereotypes may dictate different trait correlational structures within each of these groups (Xie et al., 2021). For example, in a group of American participants, intelligence and attractiveness were viewed as more correlated traits among women, but less correlated traits among men. Thus, the trait-to-trait predictions that people make within those groups are liable to be modulated by this stereotype, and thus differ from trait-to-trait predictions made in other cultures.

The true associations between traits vary across world cultures (Oh et al., 2022). For example, in Argentina, caring people are likely to be unhappy, whereas these traits are less correlated in the United States. These differences in actual personality structure are reflected in the trait-to-trait predictions that people make on the basis of facial first impressions. Cultures with more similar personality trait correlation structures also make more similar trait-to-trait predictions. These results show how trait prediction is influenced by both the actual statistics of personality within the perceiver's environment, as well as the stereotypes that those perceives may hold about different social groups.

4. Prediction across different layers of social knowledge

The last section details how the structure of social knowledge supports predictions within a given "layer" or domain of knowledge. Actions predict actions; states predict states; and traits predict traits. The structures of action, mental state, and trait knowledge carry information about their own respective domains. However, each layer of social knowledge also carries information about the future of *other* layers of social knowledge. Actions, thoughts, emotions, traits, and situations may be different type of social construct, but they are not independent. The way someone feels may motivate them to take action; situations afford behaviors; personality traits manifest in patterns of thought. By attending to this intricate web of cross-layer connections, people could capitalize on multiple sources of information to hone their predictions of the social future. In this section, we describe research into how people make these connections across different layers of social knowledge.

4.1 Mental state and trait cross-layer predictions

People's momentary mental states and enduring personality traits are deeply intertwined. People use many of the same words to describe both emotions and dispositions. For instance, one might say that a person is happy, meaning they are *currently* happy or *habitually* happy. This practice represents more than polysemous word use – it reflects the strong conceptual overlap between the psychological dimensions defining state space and the psychological dimensions defining trait space. Specifically, each of the three state dimensions of the 3d Mind Model matches with a corresponding trait dimension. State and trait valence not only share the same name, they also load onto similar sets of psychological dimensions. State rationality and trait power also share a similar loading structure (e.g., both correlate with competence from the stereotype content model). The same is true of state social impact and trait sociality.

We see this blending of trait and state space in two additional ways: First, mental states and traits are represented in largely overlapping sets of brain regions, spanning much of the social brain network (Thornton & Mitchell, 2018). This neural association goes beyond mere co-location of neural information. Specifically, a model trained to decode the valence of different mental states can be applied out-of-sample to decode the valence of different target people (Thornton & Mitchell, 2018). In other words, thinking about a dispositionally happy person elicits a similar multivariate pattern of brain activity to thinking about a person who is momentarily experiencing happiness. These findings suggest that the social knowledge for mental states and traits overlap, even if not entirely. These partially overlapping psychological spaces may represent similar content, but at different temporal scales.

How might people's minds link mental states and traits? One clue comes from personality psychology in the form of the Whole Trait Theory (Fleeson & Jayawickreme, 2021; Jayawickreme et al., 2019). This theory suggests that repeatedly measuring an individual's states gives a better measurement of their enduring traits than a one-off global self-report. The premise of this theory is that personality traits manifest in people's momentary states. Although Whole Trait Theory considers this premise primarily from a methodological perspective, one could also consider it from a social cognitive perspective. Specifically, if people's habitual states indicate their dispositions, social perceivers might leverage states to infer them.

Convergent evidence for this hypothesis comes from research on first impressions from faces. Much research in this domain has focused on why people draw such strong snap judgments from faces, given the generally low diagnosticity of facial appearance (Todorov, Funk, et al., 2015; Todorov, Olivola, et al., 2015). One of the leading explanations for this phenomenon is the facial emotion overgeneralization hypothesis (Zebrowitz & Montepare, 2008). This hypothesis derives from the observation that people use the same cues to judge the momentary valence of facial expressions as they use to judge the traits of faces (Oosterhof & Todorov, 2009). For example, the more closely a person's resting "neutral" facial appearance resembles a positive expression – like a smile – the more likely perceivers will attribute positive traits to them, such as trustworthiness. A possible explanation for this effect is that people assume that a person's resting expressions indicate their habitual affect (e.g., the formation of "smile lines" from persistent smiling). If people believed this theory, then it would give them a reasonable basis for drawing first impressions from faces in the ways they do.

Recent research drew inspiration from Whole Trait Theory and the facial overgeneralization hypothesis to test whether habitual mental states provide a good explanation for how the brain represents other people (Thornton, Weaverdyck, & Tamir, 2019b). This study compared patterns of brain activity elicited by thinking about specific people with patterns of brain activity elicited by thinking about the mental states those people experience. The results showed that the person-specific patterns could be reconstructed by summing up patterns of the mental states each person habitually experiences. For example, say a person is known to experience high levels of happiness and excitement, medium levels of thoughtfulness and calmness, and low levels of guilt and self-consciousness. If researchers then summed together the patterns associated with these states, weighted by their relative frequencies, the resulting pattern would resemble the pattern associated with that person. In other words, the brain represents others as the sum of their states. A person's landscape of state experiences was even better at predicting their neural patterns than their landscape of traits. This suggests that traditional trait terms – such as extraversion – do not fully capture our impressions of others' dispositions. Instead, they might be an efficient shorthand for communicating about frequency distributions over mental states, which would be tedious to describe verbally.

People learn about the mapping between mental state and trait representations via bi-directional causal links (Lin & Thornton, 2021). State frequencies inform trait impressions, and trait impressions inform predictions about mental states. For example, in one experiment, participants learned about a target person's disposition using a photograph and short biography, which provided no explicit information about their mental states. Manipulating dispositional information shaped how people attributed various thoughts and feelings to these targets. In another experiment, participants learned to predict a new person's mental states across different situations. Manipulating the frequencies of their states shaped how people attributed traits to the targets. These bidirectional causal effects between mental states and traits accrued only for conceptually meaningful relations. For example, manipulating the trait warmth of a target person had the largest effect on the corresponding state dimension – valence.

The results from these investigations are highly consistent with a predictive coding explanation of person perception. The brain represents other people as bundles of predictions about their likely states. Mapping personality to the base rates of another's emotions should allow people to infer others' global traits based on observations of their local states (and then use those traits to make behavioral predictions), and to predict others' states from their traits. Ongoing research aims to replicate the effects of (facial) trait impressions on situated state predictions and extend this effect to a broad range of other cultures beyond the US (Lin et al., accepted in principle).

In sum, traits may represent a sort of frequency distribution over states. That is, people collapse across time all of the individual instances of mental states one has observed in a person to generate a distribution over states that represents their disposition. To what extent are people's representations of other domains of knowledge, such as mental states, actions, or situations likewise composed similarly?

4.2 Situation and action cross-layer predictions

People's actions depend on the situations they find themselves in (Chemero, 2003, 2003; Darley & Batson, 1973; Ross & Nisbett, 2011). People reliably use their knowledge of situations to make accurate predictions about how others are likely to behave (Bach et al., 2014; Bach & Schenke, 2017; McDonough et al., 2020). And conversely, people use their observations of others' behavior to interpret situations. People might be able to effectively translate across these layers of knowledge if representations at each level are defined by frequency distributions over the other levels. For example, a situation could be thought of as a different prior probability distribution over the actions people might take within it. Do people construct situations by observing the frequencies of actions within them?

The subfield of ecological psychology has long held that situations are defined by the action affordances they offer (Chemero, 2003; Gibson, 1977). Each situation contains objects that enable specific actions. A room with a TV affords "watching," a room with a chair affords "sitting," a room with a ball affords "throwing," and so forth. By attending to the menu of objects in a space, a person may glean insight into the actions they are likely to observe (Bach et al., 2014; McDonough et al., 2020). However, affordances are not determined purely by the physical constitution of a situation. For example, a park might host a wedding one day and a memorial service the next. Despite offering similar physical affordances, the social affordances of these situations are markedly different. People hold mental schemas that inform them about which actions are appropriate in different situations (S. T. Fiske & Linville, 1980). These schemas are culturally constructed – some societies have funeral dances, but if you dance at most US funerals, you'd likely be shown the door. Many situations come with a mental script – a sequence of actions that tend to be performed in a particular order (Abelson, 1981). For example, when one arrives at a formal restaurant, one waits to be seated, orders, eats, pays, and then leaves, in that order. Strong situations can even cause people to behave in ways that radically depart from their typical behavior (Darley & Batson, 1973; Milgram, 1963). The combination of physical and social constraints embodied by a situation makes it a powerful predictor of actions. By applying this knowledge, people can navigate new instances of familiar situations with relative ease.

The predictive relationships between actions and situations suggest that the knowledge structures of these domains may be closely linked. A recent neuroimaging study demonstrated this by examining the entwined neural representations of actions and situations (Thornton & Tamir, 2023). That is, much in the same way that neural representations of individuals are comprised of the mental state they experience (Thornton, Weaverdyck, & Tamir, 2019b), this investigation showed that neural representations of situations were comprised of the actions they afforded. Representations of actions summed up to reconstruct representations of the situations that typically afforded those actions. Situations are both conceptually and neurally defined by the actions that occur within them.

4.3 State and action cross-layer predictions

People's actions depend not only on the situation, but also on their current mental states and dispositions (Carver et al., 2000; Frijda, 2004; Ross & Nisbett, 2011). For example, observing that someone is angry may lead you to anticipate that they will be more likely to engage in aggressive actions than if they were calm. Similarly, you could anticipate that a happy person is more likely to dance than someone feeling sad. Different states expand or curtail people's range of actions. Imagination may help us see new courses of action that were not previously obvious (Tsai, 2012); intense emotions like panic may limit one to a highly constrained set of stereotyped responses: fight or flight (McCarty, 2016). By re-evaluating our own or others' past actions through cognitive states such as reappraisal, we can also change the likelihood of those actions recurring (Aldao et al., 2010). Expressive actions, such as making facial expressions, gesturing, or engaging in other forms of body language, are also critical to inferring what mental state someone is currently occupying in the first place (Atkinson et al., 2004; Zaki et al., 2009).

Indeed, the neural representations of actions and states are entwined, in much the same way as actions and situations (Thornton & Tamir, 2023). As was the case with situations, mental state representations can be reconstructed from sums of the actions caused by those states. In other words, adding up the neural representations of actions that people often perform when angry – yelling, throwing objects, slamming doors, etc. – can reproduce a pattern of brain activity that looks like the pattern elicited by thinking about someone feeling angry (even if they are not engaged in any of these behaviors).

Together, these findings suggest that the layers of social knowledge are connected hierarchically via aggregation. For example, action representations sum to mental state representations, and mental state representations sum up to person representations. Other dimensions of aggregation have yet to be explored (Figure 1). For example, situations shape the expression of people's traits: people express greater extraversion at a party than at a library (Matz & Harari, 2021; Mischel & Shoda, 1995). Likewise, personality traits influence relationships (Asendorpf & Wilpers, 1998) and relationships can reshape people's personalities (Neyer et al., 2014). The literature is full of individual examples of such cross-layer interactions, but many such examples await a unifying explanatory framework such as MAPP. Although this section has examined multiple layers at once in the context of information in one layer predict the future of another layer, there are also other types of interactions which require investigation. In particular, we await research investigating how information at two or more levels are combined to make predictions.

Hierarchical aggregation may suggest why different layers of social knowledge are distinct: they may capture different sources of variance in behavior, operating at different time scales. These findings also tie together disparate ends of psychology: the internal versus external influences on behavior, or the person versus situation (Ross & Nisbett, 2011). Although commonly viewed as opposites in attribution theory, both person and situation may ultimately be constructed from the same building blocks: actions. The process of construction, rather than the component parts, may be the key to understanding these distinctions. In the next section, we dive more deeply into this idea. Specifically, we explore how social knowledge is not only useful for predicting social dynamics, social dynamics may be the very source of social knowledge itself.

5. Social dynamics shape social knowledge

So far, we have reviewed research that describes the structure of social knowledge and shows that this structure facilitates social prediction. Is this a happy coincidence, or something deeper? That is, does the structure of social knowledge come into existence independent of dynamics, and simply happen to be useful for prediction? We suggest not. Rather, we propose that the organization of social knowledge resembles the contours of real-world social dynamics because that knowledge is shaped specifically in order to support social predictions.

Most existing accounts in the literature either implicitly assume or explicitly state that the structure of social knowledge derives from static (i.e., not dynamic) features of social stimuli. This is true of some constructivist theories, as well as theories that posited more “hard-wired” evolutionary understandings of the social world. For example, one constructivist account of the origins of emotions attributes them to clustering experiences on static dimensions (Lindquist et al., 2022). That is, for each event one experiences, one can place it on different observable feature dimensions, such as where it occurred, with whom, or what actions or objects were present. Over time, some parts of this feature space would accumulate a greater concentration of experiences than others. These high-density regions could then be clustered together to give rise to discrete emotion concepts such as “happy” or “sad”. This is a powerful theory of the formation of emotion concepts. However, because it lacks a dynamic component, we suggest it cannot be a complete picture of this process.

In this section, we present evidence that the structure of mental state knowledge derives in large part from mental state dynamics (Figure 2). A wide range of learners – adults, babies, and machines – translate dynamics into structure, suggesting that the goal of social prediction is so important that it defines how we organize our understanding of other minds.

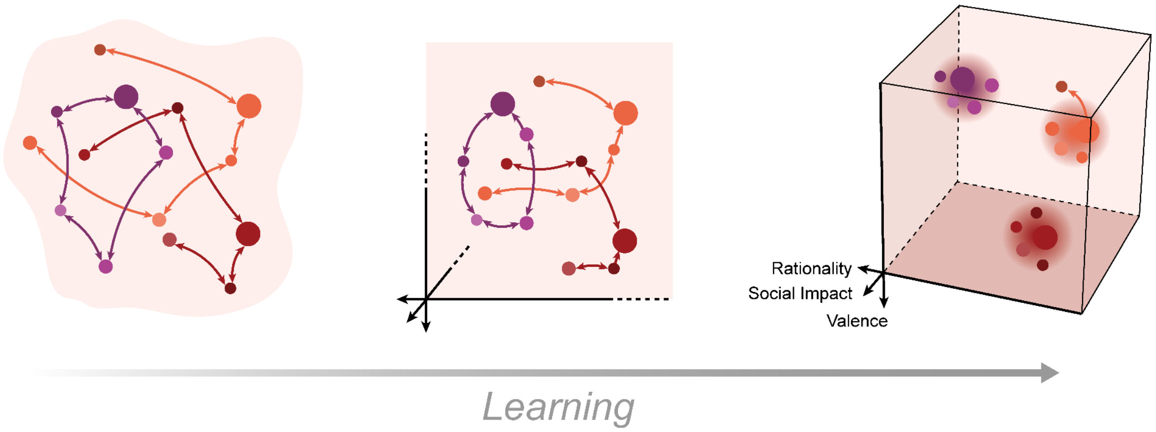

Figure 2. Learning the structure of mental state knowledge from dynamics. As adults, children, and machines observe state transition dynamics, they translate this information into structured representations. States with high transition probabilities between them move closer together in the space, while those with infrequent transitions move apart. Over time, this leads to the emergence of a knowledge structure that implicitly encodes state dynamics, and interpretable dimensions emerge to describe this structure.

5.1 Adults translate dynamics into structure

Imagine you are on a spaceship bound for another world. When you arrive, you will join an expedition to understand alien life. As a xenopsychologist, your job is to make sense of the mental states of a seemingly intelligent alien creature. So, you watch as their mental states change over time. Eventually, you report what you have learned to the rest of the crew.

This was the scenario participants encountered in a series of experiments testing if state dynamics shape state structure (Thornton et al., in press). As participants watched the alien transition from state to state, they may have noticed that some states were more likely to proceed or follow each other than others. If state dynamics shape state concepts, the higher the transition likelihood between two states, the more people should see them as conceptually similar. This is exactly what participants learned across a wide variety of transition structures. Participants reliably translated dynamics into conceptual similarity, even when they had to infer the alien's states through a noisy, naturalistic inference process.

Participants did more than just directly translate transition dynamics into pairwise conceptual similarity. Instead, they engaged in a more complex account of mental state concept formation. Specifically, participants embedded states in a low-dimensional space, such that states with higher transition probabilities between them landed closer together in the space. This geometric account explained systematic distortions in people's representations. For example, people reported relatively symmetric conceptual similarities between two states, even when the transition probabilities were highly asymmetric. If state A always precedes state B but not vice versa, then this would predict symmetric similarity *only* when A and B are embedded in a low-dimensional geometric space since distance in such a space is symmetrical. Moreover, the alien's state dynamics also influenced the mental state concepts people attributed to them, such as the valence of states and judgments about which alien states corresponded to which human mental states. The results of these experiments explain how people could form conceptually meaningful low-dimensional representational spaces – such as the 3d Mind Model (Thornton & Tamir, 2020a) – by observing mental state dynamics (Thornton et al., in press).

5.2 Children translate dynamics into structure

By adulthood, people typically have a strong understanding of mental states. Moving around adults' mental state concepts – making happiness seem negative or sadness seem positive – would likely be difficult, hence the need for artificial targets like the aliens described above. However, young children are still in the process of acquiring their understanding of the mental world. By observing the natural formation of this understanding, we can learn more about how people acquire and organize their knowledge of the social world.

Infants aren't born understanding mental states as adults do (Hoemann et al., 2019; Nook et al., 2017, 2018; Russell & Ridgeway, 1983; Saarni & Harris, 1991). Instead, extensive evidence suggests that children learn about the structure and meaning of their world through observation of transitional probabilities in sequences of speech, objects, and actions (Lew-Williams & Saffran, 2012; Monroy et al., 2017, 2019; Saffran et al., 1996; Saffran & Kirkham, 2018). The same may be true for how children learn about emotions. Infants can observe predictable transitions in adults' emotions, for example, seeing that positive emotions more often follow other positive rather than negative emotions (Thornton & Tamir, 2017). Young learners (i.e., infants and young children) accumulate these kinds of observations of emotion transitions and, in doing so, develop a representation of the structure of mental states defined by the canonical dimensions of the 3d Mind Model.

Indeed, as children age from infancy through early childhood, their own emotion transitions become increasingly structured by the dimensions that define adult transitions. This structure is scaffolded most strongly by the dimension of valence (Nencheva et al., 2023). Over time, children's emotion dynamics converge on the canonical pattern of transitions present in adulthood: they become more likely to transition between emotions that are similar in valence (e.g., sad to angry) and less likely to transition between emotions that are different in valence (e.g., sad to happy). The changes in emotional experience arise alongside the development of more accurate predictions about how states are supposed to transition from one to the next. For example, even babies as young as ten months of age respond differently to 'expected' transitions between states similar in valence than unexpected ones between states different in valence (Nencheva, Tamir, Lew-Williams, in prep). Infants track the statistical regularities of emotion transitions and converge in how they process familiar transitions. Together, these findings show from infancy through childhood, young learners may develop social knowledge by observing their own mental state dynamics and the dynamics of other people.

5.3 Machines translate dynamics into structure

Researchers interested in studying the origins of social knowledge cannot always experimentally manipulate the structure of this knowledge. Adults already have solidified views of the social world that are difficult to push around, requiring artificial stimuli instead of actual human emotions. Children's views on the social world are more malleable, but it would be impractical and unethical to distort their understanding of mental states by exposing them to a lifetime of manipulated emotion dynamics. What if there was a different experimental subject that suffered from the limitations of neither adult nor infant humans?

Artificial neural networks (ANNs) provide one such possibility. ANNs have three significant advantages over human subjects. First, they start off completely ignorant – without a human infant's instincts or biological preparedness. This makes them an ideal blank slate: anything they learn must come from the inputs you feed them. Second, ANNs are patient: they can "observe" the equivalent of a lifetime's worth of mental state dynamics in a few minutes. Third, they are focused. Humans have a wide range of goals – those programmed into them by evolution, like eating to mitigate hunger, and those they have consciously adopted. When a human subject steps into an experiment, the experimenter may add a new goal, but that does not mean that people will forget all their other goals. In contrast, ANNs seek only to optimize the specific goal – or loss function – set for them by their programmer. Therefore, any behavior the network manifests can be cleanly attributed to the combination of their inputs and prespecified goal. Together, these virtues make ANNs powerful computational tools for modeling the human mind and brain (Richards et al., 2019).

Recent work using ANNs to shows how mental state dynamics shape mental state concepts (Thornton et al., in press). Simple ANNs with a single hidden layer were trained to use one's current mental state to predict their next mental state. The sequences of mental states that they were trained on were derived directly from human mental state transition dynamics. These models spontaneously learned close facsimiles of human mental state concepts. When the size of its hidden layer was set to 3 units, it rediscovered the 3d Mind Model (Thornton & Tamir, 2020a); when the hidden layer was set to 2 units, the model rediscovered the Circumplex Model (Russell, 1980); and when the model had only one hidden unit, it rediscovered unidimensional valence. These results suggest that a system with the sole goal of predicting mental states can learn the structure of mental state knowledge simply by observing mental state dynamics.

Together with the results from adult and infant humans described above, these findings suggest that the structure of social knowledge may derive from the structure of social dynamics. As yet, this hypothesis has been thoroughly tested only in the domain of mental states, but future work may expand its explanatory breadth across all the layers of the social mind.

6. Individual differences in social knowledge and prediction

So far, we have primarily focused on the social knowledge and predictions that are shared across individuals and cultures. However, this commonality coexists with individual variation, both within and between cultures. In this section, we explore these individual differences in social knowledge and prediction, their causes, and their impacts. We focus primarily on the domain of mental states here, because this is the domain in which this variability has been best explored so far. That said, we expect many of the principles identified in this domain may generalize to other layers of social knowledge.

While the 3d Mind model generally reflects how people think about mental states, there is considerable variance in the structure of each person's idiosyncratic version of this map. By analogy, each student in a high school may have a shared understanding of the layout of their high school building. But each student has a different class schedule, meaning they walk through the building using different paths, arriving at different classrooms at different times of the day. Thus, each student's mental map will highlight different locations and routes through the space. Mental state maps may work similarly. Each person's experience of their own and others' emotions leads to an idiosyncratic map of the salient dimensions at play . The canonical 3d Mind Model can explain what is common across these maps – including the cardinal directions and the typical locations of states within them. It can also capture some forms of variability – for example, some individuals might place more emphasis on valence, whereas others might place more on social impact. However, the 3d Mind Model cannot fully describe the variability in people's trajectories through mental state space, nor their predictions about how others will traverse it.

We can track idiosyncrasies in the structure of people's mental state maps using two metrics: (i) The *size* of the map. This measure captures the granularity with which one represents states. For example, a person who clearly distinguishes between all states will have a large map; a person who has difficulty distinguishing states will have a small map. (ii) The *shape* of the map. A person attuned to all 3 dimensions will show a canonical map shape; a person who fails to tune into one dimension or is overly attuned to another will deviate from the canonical shape such that one dimension shrinks or grows in importance, respectively. With a canonical map, people can accurately predict others' future states. Despite the shared structure of this map across individuals and cultures, there remains *enormous* variation in how people navigate the social world (Thornton et al., 2022; Thornton & Tamir, 2017; Zhao et al., 2022). Each person may wield an idiosyncratic version of this map. *If* people derive their predictions from the map's structure, *then* people with different maps will make different predictions. Indeed, individual differences in map structure translate into individual differences in prediction accuracy. People with larger maps make more accurate inferences; people particularly attuned to the valence and social impact dimensions also make more accurate inferences (Barrick, Tamir, & Lincoln, in prep). Thus, individual differences in the map's integrity shape one's ability to make accurate inferences.

One factor that impacts the integrity of the map is one's own emotion experiences. Individuals with *typical* experiences – experiences that mirror the canonical mental state map – make more accurate predictions than individuals with atypical experiences (Barrick et al., 2023; Zhao et al., 2022). For example, in one study, researchers asked newly acquainted first-year college roommates to make predictions about their own emotion transitions and their new roommate's transitions. People with more canonical transitions made more accurate predictions for their roommate. Thus, individual differences in personal emotion experiences shape one's model of mental states, in turn shaping one's ability to accurately model others' emotion transitions. People whose map mirrors canonical state dynamics make accurate predictions; people whose map deviates from canonical state dynamics make less accurate predictions.

Deviations from the canonical map reduce accuracy, which, in turn, can impact people's real-world social relationships. The more accurately a person can predict others' emotion transitions, the more adeptly they can navigate the social world. This effect occurs at the relationship level, such that the better a person can predict the typical emotion transitions in their community, the less lonely they feel, and the larger their social networks are (Barrick et al., 2023). This effect also occurs within individual relationships. The better a person can predict the emotion transitions of a specific friend, the better their relationship with that friend (Zhao et al., 2022). People who accurately predict others' emotions enjoy myriad social benefits: less loneliness, better friendships, and richer social networks.

A second factor that impacts the integrity of the map is one's observations of others' emotion experiences. Children, in particular, learn about the structure of mental states from their observations of state dynamics. And indeed, there is a clear link between the *idiosyncratic* input children receive from their primary caregivers and their idiosyncratic emotion processing. Infants who observe more "typical" emotion transitions at home process emotion transitions in a more typical way (Nencheva, Tamir, Lew-Williams, in prep). As infants age into young childhood, the linguistic input they receive from their primary caregiver further shapes their conceptual understanding of emotion words. For example, caregivers who use emotion labels in sentences rich in emotional context are likelier to have kids with richer emotion vocabularies (Nencheva et al., in press). That is, the temporal proximity of positive emotion labels amongst other positively valenced words – much like the temporal proximity of positive state experiences to other positive states – helps children identify valence's key role in organizing state experiences. The development of solid emotion labels, or categories, in turn, helps kids to identify emotions and extract statistical regularities in emotion transitions (Cole et al., 2010; Hoemann et al., 2019; Nook & Somerville, 2019; Ruba et al., 2022; Thiessen et al., 2013). By adulthood, learners have accumulated plenty of observations of emotion dynamics. As people observe more and more transitions in the world, their 3d Mind Model should converge upon a *canonical* form that reflects the dynamics of the general population. Although most adults share this canonical representation of mental states and use it successfully to predict others' states, there remains significant variation around this canonical model in the adult population.

Importantly, variation in social maps can occur not only across individuals, but also within individuals. Each person carries with them multiple, interrelated copies of their social maps. For example, research suggests that people apply a relatively detailed, “high-resolution” version of the map to understanding their own mental states, but use comparatively lower granularity maps to chart others’ mental states (Thornton, Weaverdyck, Mildner, et al., 2019). One can think of this as an intra-mind version of the outgroup homogeneity effect (Quattrone & Jones, 1980). That is, just as people represent outgroups of individuals as being more homogenous than ingroups, they also seem to represent the mental states in others’ minds as being relatively homogeneous compared to the mental states in their own mind. People can also tailor multiple copies of their social maps to best suit different target people or populations. Research suggests that people quickly acquire the ability to make target-specific predictions of mental state dynamics for others with whom they closely interact (Zhao et al., 2022).

6.1 Clinical implications

Some clinical disorders, such as autism and schizophrenia, are characterized by severe disruption in social cognition and social relationships (Couture et al., 2006). These social deficits significantly impact a patient's daily functioning and overall quality of life (Fett et al., 2011). To what extent are disrupted social knowledge and predictions related to disrupted social relationships in these clinical disorders? Populations with known deficits in social interaction make markedly inaccurate predictions, including individuals with Autism Spectrum Disorder (Barrick et al., 2023) and Schizotypy (Barrick, Tamir, & Lincoln, in prep). The degree of inaccuracy in these individuals’ predictions reflects distortions in the structure of their mental state maps – both decreased size and atypical shape – *and* predicts their levels of social dysfunction. Together, this work suggests that developing reliable, accurate maps of social knowledge may be key to successfully navigating the social world.

How do social maps become miscalibrated in these disorders? Both internal and external sources of emotion observation may be at fault. That is, any disorder that disturbs one's own affective experience or one's ability to perceive others' affective experiences will lead to distorted maps. If people construct their mental state knowledge, in part, by observing the emotion dynamics that they themselves experience, then individuals with atypical emotion dynamics may form a knowledge structure that is miscalibrated for other people. Similarly, people who struggle to infer others' emotions may not be able to use observations of these others' state transitions to construct their mental model. Indeed, people with atypical emotion transitions, alexithymia, or lower face-state inference accuracy are all less accurate at predicting others' emotion transitions (Barrick et al., 2023). If these social deficits *cause* distortions in the map, then providing clinical populations with concrete feedback about typical mental state dynamics may help them construct social maps that are better calibrated for navigating the social world. That said, it is also possible that distorted maps exacerbate social dysfunction. A person with a distorted map will make less accurate predictions. Because accurate predictions facilitate social connection, distorted maps can limit the quality and quantity of a person's social interactions. This can set up a negative feedback loop in which people with distorted maps have fewer opportunities to take in new data about others' mental state transitions, leading to an even less accurate map.

7. Conclusions and future directions

People chart the social world using a set of mental maps. Each of these maps organizes social stimuli, such as actions, situations, mental states, or personality traits, in an efficient way by locating their coordinates on a small set of interpretable psychological dimensions. These maps allow people to understand the regularities in their social worlds and generalize to new stimuli. Critically, they also allow people to accurately and efficiently predict the future of their social world. Indeed, each map appears specifically tailored to capture the dynamics within its corresponding domain. This organization helps the map-bearer anticipate how others will think, feel, and behave in the future.

This line of research illustrates how tracing the influence of a relatively simple set of basic goals can reveal deep insights into the human mind. The goal of accurately representing and predicting the world and the oft-competing goal of parsimony come together here to explain phenomena across a wide swath of social psychology. These goals are not unique to social cognition: they apply to disparate domains such as vision as well. However, their application within the social sphere has led us to a compelling new understanding of the social mind and brain. By applying these simple goals to the complex, messy structure of the real social world, we can see how social understanding emerges in all of its regularity and diversity.

Importantly, the evidence presented here to show that social prediction is a default goal of the brain does not mean that the brain makes all *possible* social predictions by default. Making detailed predictions about specific events, or rolling out predictions deep into the future, are likely to be computationally costly exercises. Careful arbitration is required to determine whether and when such predictions are worthwhile (S. W. Lee et al., 2014). Chess offers a good analogy. Novice chess players often cannot foresee beyond the immediately consequences of their own moves. As chess players improve their skills, their ability to search through the space of possible future moves grows in breadth and depth (Holding, 1992). The types of predictions we describe here are analogous to a convenient notation system, whereby the encoding of a player's immediate future moves is implicitly encoded in representations of their current moves. This notation could serve to facilitate deeper predictions, but such predictions are not necessarily made as a matter of course. Thus, work presented here remains consistent with prior research suggesting that being asked to make predictions explicitly changes subsequent social judgments (Wyer et al., 1984), an observation which might otherwise seem inconsistent with the assertion that social prediction is automatic. Understanding when and why social predictions are made at different temporal depths and levels of granularity remains an interesting avenue for future research.

In this final section of the chapter, we aim to synthesize across the topics discussed so far. First, we ask whether there might be a meaningful map of maps: a way to organize across the layers of social knowledge within the MAPP framework. In this vein, we consider potential "meta-dimensions" that may span multiple layers of social knowledge. Second, we discuss key methodological innovations that have facilitated the research reviewed here. New data collection methods, and new analytic and modeling techniques, have played a crucial role in advancing our understanding of social knowledge and prediction. Third, we compare and contrast the story of predictive social cognition with an alternative view rooted in causal modeling. We suggest that these two perspectives are complementary, and capture different aspects of the social mind. Finally, we look ahead to the future of this research. We believe this future will feature research with increasing integration between the different layers of social knowledge, and will hew ever closer to naturalistic social interactions as they occur in the real-world.

7.1 Meta-dimensions of social knowledge

A primary focus of this chapter has been the discovery of the psychological dimensions that scaffold social knowledge. This includes the dimensions of models such as the 3d Mind Model of mental states and the ACT-FASTaxonomy of actions (Tamir et al., 2016; Thornton & Tamir, 2020a). These dimensions facilitate prediction within a given layer of social knowledge (Thornton et al., in press; Thornton & Tamir, 2017, 2020b, 2021a) and may even derive from the goal of prediction (Thornton et al., in press). However, they do not describe how these layers of social knowledge differ. Is there a set of meta-level dimensions that describe why people distinguish one layer of social knowledge from another?

The most salient potential meta-dimension is observability. Social stimuli at different levels differ in their observability. Actions are often directly observable: if you can see someone, there is typically little doubt whether they are sprinting versus lying down. Mental states are notoriously more difficult to infer, requiring integration over noisy cues such as facial expressions, tone of voice, body language, and context (Barrett et al., 2011; Zaki et al., 2009). Traits take this a step further, in that one must often observe a target over a long period of time, and across multiple situations to accurately infer their personality (Jones & Harris, 1967; Kelley, 1973). This spectrum of observability thus distinguishes actions, states, and traits.

Another potential meta-dimension that distinguishes each layer is timescale. Different social layers change at different timescales. Discrete physical actions – such as reaching for a glass or lifting a box – are often complete in seconds. Abstract activities like writing an email may persist longer, for minutes or hours. Emotions also tend to change over minutes or hours, while moods may persist for days (Thornton & Tamir, 2017); traits change slowly, evolving over years or decades (Roberts & Mroczek, 2008). These differences in timescale may contribute to the explanation of observability: the more quickly something happens, the more data we can accumulate about it, and the better we can infer new instances. Cross-cultural or comparative psychology studies which examine societies or species which operate at different timescales may help us to get a better grasp on how this meta-dimension comes to distinguish between different layers of social knowledge.

7.2 Methodological innovations

This review describes research that capitalizes on methodological innovations that were largely unavailable to psychologists working on these issues in the last century. These innovations include new ways to design studies, new approaches to collect, quantify, and analyze data, and new models and theories to explain the results. In this subsection, we highlight a few of these methodological innovations.

The use of functional magnetic resonance imaging (fMRI) burst into social psychological research in the early 2000s (Mitchell et al., 2002; Saxe & Kanwisher, 2003). The first generation of social fMRI research focused heavily on the localization of function and establishing functional dissociations between different tasks or types of stimuli (Mitchell, 2008). Over the last decade, more sophisticated fMRI designs and analysis techniques have emerged, making it easier for neuroimaging to answer purely psychological questions. For instance, analytic methods such as representational similarity analysis (Kriegeskorte et al., 2008) and encoding and decoding models (Naselaris et al., 2011) allow researchers to test dimensional taxonomies of social knowledge, compare them to one another, and synthesize them into more effective models (Tamir et al., 2016; Thornton & Mitchell, 2017, 2018; Thornton & Tamir, 2020b, 2020a). FMRI has numerous advantages over other data collection methods when applied in this way. For example, the neural activity that fMRI measures are implicit, difficult to control consciously, and richly multidimensional. They can also be collectively entirely passively from the participants point of view, allowing research design to decouple tasks from the need to produce measurable behavioral outcomes.

Researchers have since applied the analytic techniques pioneered in fMRI research to other measures of social cognition. For example, researchers have applied representational similarities analysis to rating data (Oh et al., 2022; Stolier et al., 2020; Thornton & Mitchell, 2017; Thornton & Tamir, 2017), reaction times (Thornton & Mitchell, 2017), text data (Thornton et al., 2022; Thornton & Tamir, 2020a), and computational models (Thornton et al., in press). Indeed, this technique was developed to put different types of data into dialogue with one another by modeling the associations between similarities measured using different methods (Kriegeskorte et al., 2008). By doing so, it has become a powerful tool by which psychologists can acquire convergent evidence across many different approaches.

Techniques such as representational similarity analysis call for new ways of designing studies. In particular, they benefit from “condition-rich” studies: studies in which the number of experimental conditions rises from the 2-4 common in traditional psychological experiments up to dozens or even hundreds. This affords researchers a potent means to address another issue: the generalizability crisis (Yarkoni, 2022). As Egon Brunswik pointed out decades ago, the importance of representative sampling extends beyond participants to the stimuli those participants engage with (Brunswik, 1955). However, representatively sampling stimuli is impossible when you only have 2-4 of them to cover an entire domain. The use of condition-rich designs and optimized stimulus selection techniques has afforded many of the studies discussed here with greater generalizability and enhanced statistical power than they could have attained using more traditional approaches (Thornton, Weaverdyck, Mildner, et al., 2019; Thornton, Weaverdyck, & Tamir, 2019c).

Finally, artificial neural networks (ANNs) are becoming increasingly useful to the study of social cognition. Pre-trained ANNs allow researchers to automatically quantify complex social behavior in naturalistic stimuli and datasets (Thornton et al., 2022; Thornton & Tamir, 2020b). Although the results are typically not as accurate as humans, and are susceptible to all the same social biases that human annotators would be, they are immensely more scalable. Research that would be impractical to conduct using manual annotation, such as labeling the facial expressions of dozens of participants in every frame of hours-long videos, suddenly becomes possible (Chang et al., 2021). ANNs are also more than just labor-saving tools. They are promising as cognitive models (Richards et al., 2019) in the domains of vision and language (H. Lee et al., 2020; O'Toole & Castillo, 2021; Schrimpf et al., 2020), and in more recent applications within social cognition as well (Dehghani et al., 2017; Dima et al., 2022; Goldstein et al., 2022; Noguchi et al., 2022; Thornton et al., in press). We believe that they holds great promise for advancing the future of the field.

7.3 Prediction vs. causation

This review has focused on how people's understanding of the social world is optimized for prediction. However, humans are not passive observers and predictors of their environment. We can, and frequently must intervene in the social world to achieve our goals. For example, if you notice that your friend is feeling down, you may take action to cheer them up. To intervene requires not just an ability to predict that world but to predict how it might counterfactually evolve depending on the different actions you might take. You must weigh how much your friend would enjoy eating ice cream versus going for a walk in the park. Understanding causation is critical to making such counterfactual predictions (Roese & Morrison, 2009).

Prediction and causation are *not* the same. As the familiar refrain goes: correlation does not imply causation. For prediction, all one needs is correlation, but for causation, one must build a structured model of causal relationships. The true causal model is not guaranteed to be the most accurate predictive model. For example, a social system's true causal model might involve variables that are difficult to infer, or it might be so complex as to become computationally intractable. As such, mental models and knowledge structures optimized for prediction are not necessarily optimized for causal reasoning, and vice versa (Ho et al., 2022).

How can the observation that much of social knowledge appears to be structured for prediction be reconciled with the fact that people will often need causal understanding to engage in critical processes like planning? There may be two non-mutually exclusive explanations. First, predictive statistical models and structured casual models may coexist in the brain (Ho et al., 2022). These models could then be called upon when individuals face different types of problems under different constraints. Second, although correlation does not *imply* causation, it is possible that correlation is empirically *correlated* with causation to a sufficient extent that the same knowledge structures can support both prediction and causal reasoning. A purely statistical predictive approach may only partially explain some facets of the social mind. Still, it is worthwhile to examine just how far we can get without evoking more complex mechanisms.

7.4 Putting it all together

Can we predict what the future of research into social knowledge and prediction looks like? We believe that two currently emerging trends point the way. First, as evidenced by several papers discussed in this review (Lin & Thornton, 2021; Thornton, Weaverdyck, & Tamir, 2019a), the next logical step in this line of research is to integrate the different layers of social knowledge. Understanding the map of even a single layer represents an important step forward, but ultimately, we cannot understand actions, situations, mental states, or traits in isolation. Each of these social domains interacts intensely and complexly with the others. We anticipate that studies that span multiple layers will both illuminate these critical interactions and solve some of the lingering puzzles within each individual domain.

Second, as these models of the social mind grow in complexity, so too will the phenomena which they can explain. Already, this line of research is characterized by the use of advanced stimulus selection techniques and the use of naturalistic stimuli, both of which can help ensure the generalizability of their results. As we chart increasingly detailed maps of the social mind, these maps will guide us ever better in our attempts to understand the messy complexity of the real world. We anticipate that this explanatory and predictive power may guide useful applications of this research in domains ranging from clinical psychology to affective computing, and from organizational management to managing interpersonal relationships.

8. References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, *36*(7), 715–729.

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237.

Alt, N. P., & Phillips, L. T. (2022). Person perception, meet people perception: Exploring the social vision of groups. *Perspectives on Psychological Science*, *17*(3), 768–787.

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, *74*(6), 1531–1544.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746.

Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: How objects guide action understanding and prediction. *Frontiers in Human Neuroscience*, *8*, 254.

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, *11*(7), e12312.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708–717.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290.

Barrick, E., Thornton, M. A., Zhao, Z., & Tamir, D. (2023). Individual differences in emotion prediction and implications for social success. *PsyArXiv*.

Berman, G. J., Bialek, W., & Shaevitz, J. W. (2016). Predictability and hierarchy in Drosophila behavior. *Proceedings of the National Academy of Sciences*, *113*(42), 11943–11948.

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. In *Advances in experimental social psychology* (Vol. 64, pp. 187–262). Elsevier.

Britton, J. C., Phan, K. L., Taylor, S. F., Welsh, R. C., Berridge, K. C., & Liberzon, I. (2006). Neural correlates of social and nonsocial emotions: An fMRI study. *NeuroImage*, *31*(1), 397–409.

Brunec, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z.-X., Grady, C., Rosenbaum, R. S., Winocur, G., & Barense, M. D. (2018). Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Current Biology*, *28*(13), 2129–2135.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217.

Carver, C. S., Sutton, S. K., & Scheier, M. F. (2000). Action, emotion, and personality: Emerging conceptual integration. *Personality and Social Psychology Bulletin*, *26*(6), 741–751.

Chang, L. J., Jolly, E., Cheong, J. H., Rapuano, K. M., Greenstein, N., Chen, P.-H. A., & Manning, J. R. (2021). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Science Advances*, *7*(17), eabf7129.

Changizi, M. A. (2003). Relationship between number of muscles, behavioral repertoire size, and encephalization in mammals. *Journal of Theoretical Biology*, *220*(2), 157–168.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, *15*(2), 181–195.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125.

Cole, P. M., Armstrong, L. M., & Pemberton, C. K. (2010). The role of language in the development of emotion regulation. In S. D. Calkins & M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition.* (pp. 59–77). American Psychological Association.

Contreras, J. M., Schirmer, J., Banaji, M. R., & Mitchell, J. P. (2013). Common brain regions with distinct patterns of neural responses during mentalizing about groups and individuals. *Journal of Cognitive Neuroscience*, *25*(9), 1406–1417.

Couture, S. M., Penn, D. L., & Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: A review. *Schizophrenia Bulletin*, *32*(suppl_1), S44–S63.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 61–149.

Darley, J. M., & Batson, C. D. (1973). " From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*(1), 100–108.

Dehghani, M., Boghrati, R., Man, K., Hoover, J., Gimbel, S. I., Vaswani, A., Zevin, J. D., Immordino-Yang, M. H., Gordon, A. S., & Damasio, A. (2017). Decoding the neural representation of story meanings across languages. *Human Brain Mapping*, *38*(12), 6096–6106.

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, *4*(11), 423–431.

Dezfouli, A., & Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *European Journal of Neuroscience*, *35*(7), 1036–1051.

Dima, D. C., Tomita, T. M., Honey, C. J., & Isik, L. (2022). Social-affective features drive human representations of observed actions. *Elife*, *11*, e75027.

Doshi, F., Konkle, T., & Alvarez, G. (2022). Human-like signatures of contour integration in deep neural networks. *Journal of Vision*, *22*(14), 4222–4222.

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016). Mood as representation of momentum. *Trends in Cognitive Sciences*, *20*(1), 15–24.

Emberson, L. L. (2017). How does experience shape early development? Considering the role of top-down mechanisms. *Advances in Child Development and Behavior*, *52*, 1–41.

Fett, A.-K. J., Viechtbauer, W., Penn, D. L., van Os, J., & Krabbendam, L. (2011). The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 573–588.

Fiske, S., Cuddy, A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902.

Fiske, S. T., & Linville, P. W. (1980). What does the schema concept buy us? *Personality and Social Psychology Bulletin*, *6*(4), 543–557.

Fleeson, W., & Jayawickreme, E. (2021). Whole traits: Revealing the social-cognitive mechanisms constituting personality's central variable. In *Advances in experimental social psychology* (Vol. 63, pp. 69–128). Elsevier.

Forstmann, M., & Burgmer, P. (2014). Adults Are Intuitive Mind-Body Dualists. *Journal of Experimental Psychology. General*, *144*(1), 222–235.

Frijda, N. H. (2004). Emotions and action. *Feelings and Emotions: The Amsterdam Symposium*, 158–173.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221.

Genkin, M., Harrigan, N., Kanagavel, R., & Yap, J. (2022). Dimensions of social networks: A taxonomy and operationalization. *Social Networks*, *71*, 12–31.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Erlbaum.

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229.

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., & Cohen, A. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, *25*(3), 369–380.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of Mind Perception. *Science*, *315*(5812), 619.

Greer, L. L. (2012). Group cohesion: Then and now. *Small Group Research*, *43*(6), 655–661.

Haggard, P. (1998). Planning of action sequences. *Acta Psychologica*, *99*(2), 201–215.

Haslam, N. (2004). *Relational models theory: A contemporary overview*. Psychology Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD: to understand human psychology, behavioural scientists must stop doing most of their experiments on Westerners. *Nature*, *466*(7302), 29–30.

Hester, N., & Hehman, E. (2023). Dress is a Fundamental Component of Person Perception. *Personality and Social Psychology Review*, 1–20.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*(6190), 1243091–1243091.

Ho, M. K., Saxe, R., & Cushman, F. (2022). Planning with Theory of Mind. *Trends in Cognitive Sciences*.

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*, *55*(9), 1830.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*(3), 687–701.

Holding, D. H. (1992). Theories of chess skill. *Psychological Research*, *54*(1), 10–16.

Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593.

Hudson, M., Nicholson, T., Ellis, R., & Bach, P. (2016). I see what you say: Prior knowledge of other's goals automatically biases the perception of their actions. *Cognition*, *146*, 245–250.

Ishai, A., Pessoa, L., Bikle, P. C., & Ungerleider, L. G. (2004). Repetition suppression of faces is modulated by emotion. *Proceedings of the National Academy of Sciences*, *101*(26), 9827–9832.

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522.

Jayawickreme, E., Zachry, C. E., & Fleeson, W. (2019). Whole trait theory: An integrative approach to examining personality structure and process. *Personality and Individual Differences*, *136*, 2–11.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. In *Advances in experimental social psychology* (Vol. 2, pp. 219–266). Elsevier.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*(1), 1–24.

Kelley, H. H. (1973). The Processes of Causal Attribution. *American Psychologist*, *28*(2), 107–128.

Kim, M., Young, L., & Anzellotti, S. (2022). Exploring the Representational Structure of Trait Knowledge Using Perceived Similarity Judgments. *Social Cognition*, *40*(6), 549–579.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron*, *79*(5), 836–848.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*(7), 984–991.

Lee, H., Margalit, E., Jozwik, K. M., Cohen, M. A., Kanwisher, N., Yamins, D. L., & DiCarlo, J. J. (2020). Topographic deep artificial neural networks reproduce the hallmarks of the primate inferior temporal cortex face processing network. *BioRxiv*.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, *81*(3), 687–699.

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, *122*(2), 241–246.

Lin, C., Keles, U., Thornton, M. A., & Adolphs, R. (accepted in principle). Trait impressions from faces shape mental state inferences. *Nature Human Behaviour*.

Lin, C., & Thornton, M. A. (2021). Linking attributions of mental states and traits: Evidence for bidirectional causation. *PsyArXiv*.

Lindquist, K. A., Jackson, J. C., Leshin, J., Satpute, A. B., & Gendron, M. (2022). The cultural evolution of emotion. *Nature Reviews Psychology*, 1–13.

Magistretti, P. J., & Allaman, I. (1999). Brain energy metabolism. *Fundamental Neuroscience*, *3*, 271–293.

Mar, R. A., Mason, M. F., & Litvack, A. (2012). How daydreaming relates to life satisfaction, loneliness, and social support: The importance of gender and daydream content. *Consciousness and Cognition*, *21*(1), 401–407.

Martin, A. E., & Slepian, M. L. (2021). The primacy of gender: Gendered cognition underlies the Big Two dimensions of social cognition. *Perspectives on Psychological Science*, *16*(6), 1143–1158.

Matz, S. C., & Harari, G. M. (2021). Personality–place transactions: Mapping the relationships between Big Five personality traits, states, and daily places. *Journal of Personality and Social Psychology*, *120*(5), 1367.

McCarty, R. (2016). The fight-or-flight response: A cornerstone of stress research. In *Stress: Concepts, cognition, emotion, and behavior* (pp. 33–37). Elsevier.

McCrae, R. R., & Costa, J., Paul T. (1987). Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90.

McDonough, K. L., Costantini, M., Hudson, M., Ward, E., & Bach, P. (2020). Affordance matching predictively shapes the perceptual representation of others’ ongoing actions. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(8), 847–859.

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, *67*(4), 371–378.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246–268.

Mitchell, J. P. (2008). Contributions of functional neuroimaging to the study of social cognition. *Current Directions in Psychological Science*, *17*(2), 142–146.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences*, *99*(23), 15238–15243.

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology*, *157*, 14–28. https://doi.org/10.1016/j.jecp.2016.12.004

Monroy, C. D., Meyer, M., Schröer, L., Gerson, S. A., & Hunnius, S. (2019). The infant motor system predicts actions based on visual statistical learning. *NeuroImage*, *185*, 947–954. https://doi.org/10.1016/j.neuroimage.2017.12.016

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage*, *56*(2), 400–410.

Nencheva, M. L., Nook, E. C., Thornton, M. A., Lew-Williams, C., & Tamir, D. I. (2023). The co-emergence of emotion vocabulary and organized emotion dynamics in childhood. *PsyArXiv*.

Nencheva, M. L., Tamir, D. I., & Lew-Williams, C. (in press). Caregiver speech predicts the emergence of children's emotion vocabulary. *Child Development*.

Neyer, F. J., Mund, M., Zimmermann, J., & Wrzus, C. (2014). Personality-relationship transactions revisited. *Journal of Personality*, *82*(6), 539–550.

Noguchi, W., Iizuka, H., Yamamoto, M., & Taguchi, S. (2022). Superposition mechanism as a neural basis for understanding others. *Scientific Reports*, *12*(1), 1–14.

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, *1*(12), 881–889.

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2018). The nonlinear development of emotion differentiation: Granular emotional experience is low in adolescence. *Psychological Science*, *29*(8), 1346–1357.

Nook, E. C., & Somerville, L. H. (2019). Emotion concept development from childhood to adulthood. *Emotion in the Mind and Body*, 11–41.

Oh, D., Martin, J. D., & Freeman, J. B. (2022). Personality across world regions predicts variability in the structure of face impressions. *Psychological Science*, *33*(8), 1240–1256.

Olshausen, B., & Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*(1), 128–133.

O’Toole, A. J., & Castillo, C. D. (2021). Face recognition by humans and machines: Three fundamental advances from deep learning. *Annual Review of Vision Science*, *7*, 543–570.

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, *112*(4), 642–681.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(03), 715–734.

Quattrone, G. A., & Jones, E. E. (1980). The Perception of Variability Within In-Groups and Out-Groups: Implications for the Law of Small Numbers. *Journal of Personality and Social Psychology*, *38*(1), 141–152.

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, *38*, 433–447.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718.

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., & Ganguli, S. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, *22*(11), 1761–1770.

Richardson, H., & Saxe, R. (2020). Development of predictive responses in theory of mind brain regions. *Developmental Science*, *23*(1), e12863.

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science*, *17*(1), 31–35.

Roese, N. J., & Morrison, M. (2009). The psychology of counterfactual thinking. *Historical Social Research/Historische Sozialforschung*, 16–26.

Ross, L., & Nisbett, R. E. (2011). *The person and the situation: Perspectives of social psychology*. Pinter & Martin Publishers.

Ruba, A. L., Pollak, S. D., & Saffran, J. R. (2022). Acquiring complex communicative systems: Statistical learning of language and emotion. *Topics in Cognitive Science*.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russell, J. A., & Lewicka, M. (1989). A Cross-Cultural Study of a Circumplex Model of Affect. *Journal of Personality and Social Psychology*, *57*(5), 848–856.

Russell, J. A., & Ridgeway, D. (1983). Dimensions underlying children's emotion concepts. *Developmental Psychology*, *19*(6), 795–804.

Saarni, C., & Harris, P. L. (1991). *Children's understanding of emotion*. CUP Archive.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203.

Saxe, R. (2018). Seeing Other Minds in 3D. *Trends in Cognitive Sciences*, *22*(3), 193–195.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842.

Saylor, M. M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007). Infants' on-line segmentation of dynamic human action. *Journal of Cognition and Development*, *8*(1), 113–128.

Schrimpf, M., Blank, I., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J., & Fedorenko, E. (2020). Artificial neural networks accurately predict language processing in the brain. *BioRxiv*, *10*(2020.06), 26.174482.

Song, X., & Wang, X. (2012). Mind Wandering in Chinese Daily Lives – An Experience Sampling Study. *PLOS ONE*, *7*(9), e44423. https://doi.org/10.1371/journal.pone.0044423

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of Personality in Early and Middle Adulthood: Set Like Plaster or Persistent Change? *Journal of Personality and Social Psychology*, *84*(5), 1041–1053.

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, *4*(4), 361–371.

Sudhof, M., Goméz Emilsson, A., Maas, A. L., & Potts, C. (2014). *Sentiment expression conditioned by affective transitions and social forces*. 1136–1145.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006.

Tamir, D. I., & Thornton, M. A. (2018a). Modeling the predictive social mind. *Trends in Cognitive Sciences*, *22*(3), 201–212.

Tamir, D. I., & Thornton, M. A. (2018b). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. https://doi.org/10.1016/j.tics.2017.12.005

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences*, *113*(1), 194–199.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792–814.

Thornton, M. A., & Mitchell, J. P. (2017). Consistent neural activity patterns represent personally familiar people. *Journal of Cognitive Neuroscience*, *29*(9), 1583–1594.

Thornton, M. A., & Mitchell, J. P. (2018). Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral Cortex*, *28*(10), 3505–3520.

Thornton, M. A., Reilly, B. J., Slingerland, E., & Tamir, D. (2022). The 3d Mind Model characterizes how people understand mental states across modern and historical cultures. *Affective Science*, *3*(1), 93–104.

Thornton, M. A., & Tamir, D. (2023). The brain represents situations and mental states as sums of their action affordances. *PsyArXiv*.

Thornton, M. A., & Tamir, D. I. (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences*, *114*(23), 5982–5987.

Thornton, M. A., & Tamir, D. I. (2020a). People represent mental states in terms of rationality, social impact, and valence: Validating the 3d Mind Model. *Cortex*, *125*, 44–59.

Thornton, M. A., & Tamir, D. I. (2020b). Perceiving actions before they happen: Psychological dimensions scaffold neural action prediction. *Social Cognitive and Affective Neuroscience*, *16*(8), 807–815.

Thornton, M. A., & Tamir, D. I. (2021a). Perceptions accurately predict the transitional probabilities between actions. *Science Advances*, *7*(9), eabd4995.

Thornton, M. A., & Tamir, D. I. (2021b). The organization of social knowledge is tuned for prediction. In K. N. Ochsner & M. Gilead (Eds.), *The Neural Bases of Mentalizing*. Springer Nature.

Thornton, M. A., & Tamir, D. I. (2022). Six dimensions describe action understanding: The ACT-FASTaxonomy. *Journal of Personality and Social Psychology*, *122*(4), 577–605.

Thornton, M. A., Vyas, A. D., Rmus, M., & Tamir, D. (in press). Transition dynamics shape mental state concepts. *Journal of Experimental Psychology. General*.

Thornton, M. A., Weaverdyck, M. E., Mildner, J. N., & Tamir, D. I. (2019). People represent their own mental states more distinctly than those of others. *Nature Communications*, *10*(2117).

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019a). The brain represents people as the mental states they habitually experience. *Nature Communications*, *10*(1), 1–10.

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019b). The social brain automatically predicts others' future mental states. *Journal of Neuroscience*, *39*(1), 140–148.

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019c). The social brain automatically predicts others' future mental states. *Journal of Neuroscience*, *39*(1), 140–148.

Todorov, A., Funk, F., & Olivola, C. Y. (2015). Response to Bonnefon et al.: Limited 'kernels of truth'in facial inferences. *Trends in Cognitive Sciences*, *8*(19), 422–423.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545.

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in Everyday Life. *PloS One*, *10*(12), e0145450.

Tsai, K. C. (2012). Play, Imagination, and Creativity: A Brief Literature Review. *Journal of Education and Learning*, *1*(2), 15–20.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

United States Bureau of Labor Statistics. (2003). *American Time Use Survey*. https://www.bls.gov/tus/

Von Der Lühe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150373.

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music–brain responses to rhythmic incongruity. *Cortex*, *45*(1), 80–92.

Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science*, *19*(1), 58–62.

Weaverdyck, M. E., Thornton, M. A., & Tamir, D. I. (2021). The representational structure of mental states generalizes across target people and stimulus modalities. *NeuroImage*, *238*, 118258.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598.

Wilt, J., Funkhouser, K., & Revelle, W. (2011). The dynamic relationships of affective synchrony to perceptions of situations. *Journal of Research in Personality*, *45*(3), 309–321.

Wish, M., Deutsch, M., & Kaplan, S. J. (1976). Perceived Dimensions of Interpersonal Relations. *Journal of Personality and Social Psychology*, *33*(4), 409–420.

Wyer, R. S., Srull, T. K., & Gordon, S. (1984). The effects of predicting a person's behavior on subsequent trait judgments. *Journal of Experimental Social Psychology*, *20*(1), 29–46.

Xie, S. Y., Flake, J. K., Stolier, R. M., Freeman, J. B., & Hehman, E. (2021). Facial impressions are predicted by the structure of group stereotypes. *Psychological Science*, *32*(12), 1979–1993.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1.

Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion*, *9*(4), 478.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517.

Zhao, Z., Thornton, M. A., & Tamir, D. I. (2022). Accurate emotion prediction in dyads and groups and its potential social benefits. *Emotion*, *22*(5), 1030–1043.