Citation in APA format: Barrett, F. L., & Theriault, J. (2025). What's real? A philosophy of science for social psychology. In D. T. Gilbert, S. T. Fiske, E. J. Finkel, & W. B. Mendes (Eds.), *The handbook of social psychology* (6th ed.). Situational Press. https://doi.org/10.70400/BPQW3358

What's Real? A Philosophy of Science for Social Psychology

Lisa Feldman Barrett, Northeastern University

Jordan Theriault, Northeastern University

All scientists are philosophers. That is, they hold fundamental assumptions about the nature of reality and how best to develop "true" (i.e., justified) knowledge about it. The way a scientist goes about building justified knowledge reveals what they take reality to be. Their assumptions guide their scientific practices and products — their choices of topics to study, their formulation and testing of hypotheses, and ultimately, the inferences they offer as *claims* of knowledge. Other scientists, guided by their own philosophical assumptions, turn those claims into bona-fide knowledge by using them and citing them. Or not.

Consider, for example, a meta-analysis that summarizes about 100 studies on the facial movements that express emotions. In these studies, most of which used the same experimental method, 22,000 participants from 20 countries viewed photos of people posing scowling faces and labeled them as expressions of anger at greater than chance levels. What can be concluded from this meta-analysis? That humans universally recognize scowls as expressions of anger? That humans universally express anger by scowling? A scientist's answers depend on their philosophy of science.

Based on such evidence, some scientists are then inspired to build algorithms to detect scowls and refer to this as "detecting anger." Currently, this so-called "emotion AI" is a multibillion-dollar industry. Scientists and industry professionals worldwide debate the pros and cons of using such "emotion reading" algorithms to determine who is guilty of a crime or well-suited for a job. Would any scientist consent to their own outcomes being influenced with this technology? Their answers, again, likely depend on their philosophy of science.

Now, suppose a scientist conducts their own experiments about facial movements during episodes of emotion using a variety of methods, and their observations contradict those of the preceding meta-analysis. One philosophical practice might lead scientists to conclude that this failure to replicate the meta-analytic results suggests sloppy research practices or even that the newer experiments were flawed in some fundamental way. A different philosophical commitment, however, leads to a different hypothesis. Perhaps the expression of anger fundamentally varies with the context. If so, perhaps the uniformity observed in the original meta-analysis may be evidence that scowling is perceived as an expression of anger only under certain conditions within

a particular experimental setup. Some scientists might wonder about the nature of those conditions and what they suggest about the contextual nature of anger and anger expressions. These scientists might ultimately favor experiments that are designed to cultivate and observe structured variation in emotional expressions, rather than experiments that control variation to observe simpler, stable patterns of expression.

This real-life example comes from a century-long debate in the science of emotion (for review and discussion see Barrett et al., 2019). Scientific evidence has failed to resolve this debate, in part, because the different points of view are rooted in incompatible research traditions (Laudan, 1977) that differ in their philosophical practices, concepts, methodologies, and observations. In the "typological" tradition, e.g., "basic emotion" or "discrete emotion" approaches, English words like "angry," "sad," and "fearful" are each assumed to refer to a distinct grouping of instances that are strongly similar across people and contexts (i.e., a biological and psychological "type"). Each emotion is assumed to be real and inherent in nature, and that nature is assumed to exist independently of any perceiver. By contrast, in the "relational" tradition, e.g., constructionist approaches, the same emotion words are assumed to name folk psychology or commonsense categories that are populations of diverse instances that vary across situations and individuals. Each emotion is assumed to be real in nature, and nature is conceived of as dynamically changing events that emerge from relations that always involve a perceiver. Reality is perspective-independent, full of sameness and stability, as opposed to perspective-dependent, full of variation and complexity. Each side uses similar words , but they are speaking past each other, an important concept from the philosophy of science called *incommensurability* (Feyerabend, 1962; Kuhn, 1962/1970; for a more nuanced history of incommensurability, see Oberheim & Hoyningen-Huene, 2018). Some of the toughest and most intractable scientific debates, such as this one over the nature of emotion, are actually philosophical disagreements in disguise.

Welcome to the first chapter on philosophy of science in the history of *The Handbook of Social Psychology*. Philosophy explores the fundamental nature of reality and humans' place in it, including: How is "real" defined and what meets that definition (i.e., what constitutes reality)? What does justified (i.e., "true" or "approximately true") knowledge look like, and how is knowledge justified? What are the most reliable sources of knowledge, and how does knowledge change? Philosophy of science zeros in on these beliefs as they pertain to the practices and products of science as a source of reliable, useful knowledge about the world.

Most psychological scientists have no formal training in philosophy of science. Maybe they've read a bit about the philosopher Karl Popper and his idea that hypotheses can be falsified but never proven true. Maybe they hold the empiricist notion that replication is the gold standard for creating justified knowledge about the world. Maybe they learned about the historian of science Thomas Kuhn and his world-shaking claim that scientific knowledge changes, not by slow and steady progress towards an objective reality that is "out there," separate from humans, but in cataclysmic revolutions that shift everything, including the very concepts of the reality that scientists are trying to know. Because few psychological scientists talk about philosophy, it's easy to assume that it has little to do with boots-on-the-ground science. But ample evidence suggests otherwise. Some of this evidence is covered in this chapter.

Scientists enact a philosophy of some sort every day as they engage in scientific work, whether running experiments, analyzing data, talking to colleagues, or just reading papers. Even a scientist who believes they are purely driven by data is doing philosophy. Every person has potent beliefs about the nature of reality, evidence, inference, and knowledge. In scientists, these beliefs stealthily

impact their scientific practices and products. Correspondingly, their day-to-day scientific practices and products give evidence of their philosophy, regardless of their stated beliefs. A scientist's success in adding to (what counts as) accepted psychological knowledge largely depends on whether other scientists share their philosophy of science and agree with the knowledge claims being advanced in their papers, lectures, and reviews.

This chapter is an introduction to philosophy of science as it is practiced rather than a general survey of important topics. Useful surveys can be found in informative works by the philosophers Peter Godfrey-Smith (2021) and Ian Hacking (1983), Stanford's online Encyclopedia of Philosophy (https://plato.stanford.edu/), and introductory YouTube lectures by Dr. Kane Baker via his channel, Kane B. This chapter relies on all these sources plus original philosophical and historical discussions of how science *should* work (normative claims), as well as descriptions of scientists' *actual* behavior as documented by historians, anthropologists, and sociologists. Several major philosophies of science are surveyed (see Table 1) to explain how certain scientific decisions and actions manifest as the practice of one philosophy versus another (regardless of what a scientist may say or believe). This chapter also includes selected scientific results from social psychology and other fields to weigh in on the suitability of certain philosophical commitments that scaffold scientific practice. (The inherent circularity of this endeavor is impossible to escape.)

***Table 1*: Philosophies of Science**

	Ontology	Epistemology
Traditional Scientific Realism	Reality is structured as fixed, universal categories called natural kind categories. Natural kinds exist independently of any perspective and include both observable and unobservable (i.e., inferred) parts of reality.	The best and most mature theories (as systems of concepts) steadily approach true (or approximately true) descriptions of natural kinds. True knowledge is an objective, perceiver-independent representation of reality, corresponds to reality, or coheres with other objective knowledge of reality. True knowledge predicts and explains reality.
Empiricism	Reality is structured as fixed, universal, empirical kind categories that reject unobservables. Empirical kinds depend on human observation and so are perceiver-dependent but are concept-independent.	The best and most mature theories steadily become empirically adequate descriptions of the observable world (i.e., empirical kinds). Reality may be perceiver-independent, but knowledge is always perceiver-dependent.
Relational Realism	Reality is relational. Everything that is real — a category, its instances and features, its causal impacts — exists as relations among signals within specific contexts. This includes both the observable and unobservable parts of reality. The parts of reality that are relevant for humans emerge from relations that include the categories and concepts that human brains and bodies create.	The best and most mature theories are situation- and perceiver-dependent (i.e., they describe and explain relational kinds). Knowledge is objective when it arises from critically achieved consensus in a diverse scientific community (Appendix, Section 7). Knowledge is justified when predictions materialize with a high probability given the constraints of the situation (i.e., a high posterior probability).
Perspectivism	Perspectivists can be traditional realists who assume reality is organized as natural kinds, or relational realists who assume reality is structured as relational kinds.	Knowledge that predicts and explains reality is only true from within a particular scientific perspective. All knowledge is perceiver dependent.
Pragmatism	Pragmatists can be traditional realists (like Pierce) or relational realists (like William James). See Appendix, Section 6.	Pragmatism is utility with a specific goal in mind. Whatever brings us closer to that goal counts as knowledge.
Functionalism	Functionalists can be traditional realists (aiming at natural kinds) or relational realists (aiming at relational kinds).	The best and most mature theories are true (or approximately true) to the extent that they regularly predict and explain certain outcomes based on certain inputs. In another form of functionalism, knowledge is in service of a particular activity or purpose.

In structure, the chapter considers three related questions:

1. Which scientific method do psychological scientists use to curate knowledge claims, and crucially, which assumptions are required to use this method to build justified scientific knowledge?

2. What assumptions do psychological scientists make about the nature of reality when using their scientific method of choice?

3. What sort of philosophy offers the best chance to build justified knowledge about psychological phenomena?

The chapter opens with question 1 and provides a brief overview of an icon of scientific methodology: the *hypothetico-deductive method*. Most psychological scientists have encountered this method during their scientific training and use it every day. However, it requires some surprising preconditions to deliver the robust and replicable knowledge that most scientists seek, and it's unclear whether psychological science currently meets — or can ever meet — these preconditions. Moreover, the hypothetico-deductive method is not the *only* scientific method available, and it's not equally suited for every philosophy of science practiced in psychological science.

This overview of methodology is the foundation for question 2: What do psychologists assume about the structure of the reality that they are trying to know scientifically? Science is fundamentally a framework of concepts for understanding reality, and accordingly, it depends on other concepts about the nature of reality. In philosophy, such questions belong to the domain of metaphysics or *ontology*, which asks questions such as: What things exists in space and time? (What does it mean to be a "thing" vs. a "process" or an "event"? How should "space" and "time" be understood?) What properties describe a thing? (And what counts as a "property," anyway?) What causal relations exist between this thing and other things? (What is "cause"? Is it deterministic or probabilistic? Does a "cause" differ from a "regularity," a "contingency," or an "explanation"?)

While exploring psychology's ontological foundations, the chapter touches on beliefs about the nature of scientific knowledge and its relation to reality, known as *epistemology*. Examples of epistemic questions are: What is a scientific observation? How is "observation" different from "inference," and when does the former permit the latter? What does it mean to claim that something is true, and what is "truth," anyway? Is "truth" the ultimate point of knowledge, or are we after something else, like "useful tidbits" for acting on the world (or for getting and keeping a job)? Are there limits on what can be known about reality? What makes scientific knowledge more trustworthy than other methods for gaining knowledge?

Question 3 builds from the discussions from questions 1 and 2. Are current scientific practices, and their attendant beliefs about the nature of psychological reality, optimal for accumulating justified knowledge about psychological phenomena? For example, concerns about methodological rigor when designing, running, and analyzing controlled laboratory experiments can obscure deeper, thorny philosophical issues about whether those experiments are well-suited for learning about psychological phenomena. Debating the former while ignoring the latter can wreak havoc in insidious ways. This tends to happen when scientists simply redouble their efforts to improve rigor in using the hypothetico-deductive method rather than question its suitability for building justified knowledge in the first place. Greater attention to the philosophy of doing psychological science may therefore improve the validity and trustworthiness of psychological knowledge beyond what greater methodological rigor can achieve alone.

And because science is performed by a community of humans, the chapter briefly considers the social nature of scientific knowledge, and the ways in which scientific knowledge becomes more objective when the entire enterprise of science is viewed as a community activity. Psychological science tends to assume an individualistic picture of scientific work (if the incentive structure is any guide), but the practices of every individual scientist are always embedded in social transactions that encourage and support certain ideas and behaviors while discouraging or outright punishing others. These *non-epistemic* influences challenge an individualistic conception of scientific knowledge as built by lone scientists who rationally employ a scientific method in the pursuit of truth.

Philosophy can be infuriatingly abstract, so to make things concrete throughout, the chapter uses examples of building and using knowledge about everyday things like cats and water; examples from physics, chemistry, and biology; and closer to home, examples from the science of emotion, such as the debate described earlier.

Ultimately, science is more than making precise measurements that replicate across experiments. It's about more than being correct. It's also more than racing to publish a mountain of papers. Science, at its best, crafts a usable system of concepts for thinking about what exists in the world, a.k.a., reality, a system of concepts that works well enough for humans to act in the world, and even add to the world, in ways that interest them. Until it doesn't. And then it's time to entertain a different conceptual framework, or even more than one, thereby unlearning the most fundamental ideas about reality and replacing them with others. This is what Kuhn meant when he penned one the most controversial ideas in *The Structure of Scientific Revolutions* (Kuhn, 1962/1970), the notion that scientific concepts *construct* the world that scientists are trying to know.

A CLOSER LOOK AT THE SCIENTIFIC METHOD

Science is widely considered the best approach to build justified knowledge about the world, at least by scientists. The standard claim is that science is a rational method of observation and logical inference — something like the *hypothetico-deductive method* — that systematically exposes beliefs to the reality of experience to adjudicate which ones are sufficiently justified to count as knowledge. The hypothetico-deductive method is drilled into the brain of every budding psychological scientist and is the official mantra of proper scientific technique (e.g., L. D. Nelson et al., 2018): Ideas are fashioned into hypotheses and tested against rigorous, public observations of the world. When hypotheses and observations are consistent, scientists should repeat the process with increasing boldness and ever more risky and rigorous tests. If the two are inconsistent, then the hypothesis is likely false and should be rejected. This is the standard process of building justified knowledge about the world.

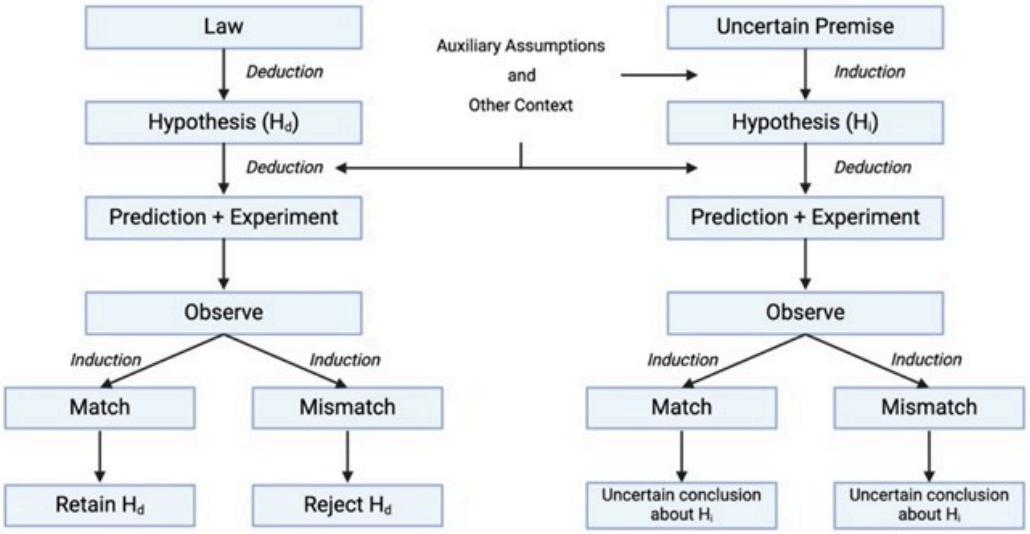
Consider the following two phrases from the preceding paragraph: to "expose beliefs to the reality of experience" and to "test hypotheses against public observations of the world." In the hypothetico-deductive method, the two phrases mean the same thing. The equivalence of "belief" and "hypothesis" seems straightforward. A hypothesis is a belief that is lightly held but sufficiently justified to be worth testing. The equivalence of "experience" and "observation," however, might seem confusing to scientists who were trained to disavow experiences in the methods of science.

Observations are experiences that scientists cultivate for themselves using some set of procedures. To observe is to carefully quantify changes in something during an experiment. A physicist might cultivate experiences of electrons by observing fluctuations of kinetic energy when a light shines on metal (the emission of electrons from a material caused by electromagnetic radiation is called the photoelectric effect). A chemist might cultivate experiences of a chemical reaction by observing how long it takes for a volume of liquid to heat up or cool down (as an indication that a chemical reaction is occurring). A neuroscientist might cultivate experiences of neural activity by observing fluctuations in a magnetic field due to changes in blood flow and blood oxygenation levels (called blood-oxygen-level-dependent or BOLD signal). A psychological scientist might experience other people as emotional by observing the time it takes for participants to press a computer key or to speak a word (called reaction time), the speed of each participant's heart beats (called heart rate), or the velocity, location, and duration of eye movements (called saccades). To say that observations are

public simply means that that the procedures needed to cultivate similar experiences are available for anyone to test the same hypothesis by attempting to reproduce the experimental findings with the described method.

The hypothetico-deductive method, depicted in Figure 1, is not an actual description of how scientists practice their craft. It's more like a prescription for building bona-fide justified knowledge about objects (including people), processes, events, and relations — theories that explain, predict, act on, and modify aspects of the world. The details of how, and even when, to use the hypothetico-deductive method make all the difference to the success of any scientific endeavor.

Figure 1: **The Hypothetico-Deductive Method**



Note: (left) Starting premise is a scientific law. (right) Starting premise is less certain than a law. H_d = Hypothesis deductively inferred from a scientific law. H_i = Hypothesis inductively inferred from a premise that is less certain than a law. Figure created in BioRender.

Categories And Concepts

When scientists employ the hypothetico-deductive method, they aim it at electrons, DNA, emotions, cognitions, and other potential objects of scientific knowledge. Psychological scientists often call their objects "constructs." To philosophers, they are "entities" or "kinds." Whatever they're called, they are ideas about the categories of nature, such as the category ELECTRON, DNA, EMOTION, or COGNITION. When scientists conduct an experiment, they curate *instances* of a category (i.e., scientists "sample" a group of instances from the larger population of all possible category instances), they experience those instances (a process called "observing"), and they expose their hypothesis, which is part of their concept for the category, to their experiences, for better or worse. Whether they reject their hypothesis or not, scientists attempt to infer something about the entire category — all instances that ever have existed or will exist — from the sample of instances they experienced. Over time, if all goes well, the variation and uncertainty of an ever-changing world is reduced and made more predictable (and even more controllable) by appreciating the equivalence among different instances. Scientific knowledge about the category develops as scientists hone and extend the associated scientific concept (that describes the category). The

hypothetico-deductive method is a recipe for this entire process of uncertainty reduction, with the ultimate purpose to discover justified, generalizable, and therefore useful knowledge about the category in question.

Let's review some of what is conventionally considered justified knowledge about categories and concepts in cognitive science and philosophy. A *category* is a grouping of individual occurrences — specific objects, processes, and events — that are equivalent to one another in some way (Murphy, 2004; Sidman, 1994; E. E. Smith, 1989). The table of contents in this *Handbook*, for example, includes all sorts of categories in social psychology, such as SELF-CONTROL, GOALS, ATTITUDES, AFFECT, WELL-BEING, PREJUDICE, MENTALIZING, ROMANTIC RELATIONSHIPS, INTERGROUP RELATIONS, and so on. *Every* phenomenon of interest to a scientist is a grouping of instances that are similar enough to be interchangeable, yet different enough from the instances of other groupings so as not to be confused with them.

A category's instances are similar to one another because they have some features (properties, qualities, behaviors) in common, allowing them to be treated as equivalent or interchangeable. For example, if instances of FEAR in humans around the world, in cage-reared and wild-type rats, in fish and in flies all share some feature that makes them instances of the same category, then whatever scientists learn from experiencing a small sample of instances of that category in one species (say, observing humans whose hearts are racing) should generalize as useful knowledge about the larger category of FEAR instances in humans and in other species. The features of equivalence shared by all members of category are called the *intension* of the category. All instances that possess the intension are the category's *extensions* or exemplars. So, all extensions of FEAR contain the intension for FEAR. (Philosophers typically refer to the intension and extensions of a word or a concept but this chapter also uses "intension" and "extension" to refer to the corresponding category because this philosophical jargon becomes useful later when discussing ideas about the different kinds of categories that might exist and what they allow a scientist to accomplish with a scientific method.) Scientific questions about occurrences of fear, for example, often include questions about intensions of the category FEAR. Do instances of FEAR (the category's extensions) share a pattern of physiological changes in the body, a facial expression, an action like freezing or fleeing, and so on? Do different instances of FEAR issue from the same assembly of neurons? Do extensions of FEAR, such as fear of heights, fear of spiders, fear of social evaluation, and fear of physical harm share some set of features that make them all instances of FEAR? If so, then the instances are sufficiently similar to one another for a person to predict, infer, or explain new instances of FEAR based on what is known about prior instances that have been experienced, observed, or otherwise learned. This inference is a hypothesis. The new extensions of FEAR are an experiment that tests the hypothesis. Whatever is learned about these new extensions of FEAR should generalize to all instances. Notice, however, that the similarity between instances of the same category must be *substantially* more regular than what would be expected by chance if the goal is to build usable knowledge about an entire category of FEAR just by experiencing a few of its instances. Scientific questions might also focus on the relations between a category and other categories (e.g., relations between FEAR and ATTITUDES), understood either as a predictable pattern or dependency in the flow of events (i.e., a correlation) or as a cause-effect connection between them (i.e., causation).

In cognitive science, a *concept* is a set of beliefs about a category. A concept for fear, for example, is composed of beliefs about the category FEAR. If a category simplifies variation and creates meaning by finding the similarities among different instances that render them equivalent, then a concept is the summary of that meaning. A concept includes the category's features of equivalence — the

intension that makes category instances functionally interchangeable (e.g., the neurons that cause instances of FEAR, the physiological changes in the body during instances of FEAR, the facial expressions for FEAR, or the tendency to escape or freeze during fear, etc.); the causal relations between those features and the category's other features (e.g., the pleasure of a rollercoaster or a horror movie, or an act of aggression towards another creature); the causal impacts of the category's instances (Murphy, 2004; E. E. Smith, 1989); and some abstract feature that refers to the concept, usually including the word that names the category (e.g., "fear"). Knowledge is a set of *justified* concepts (i.e., beliefs that are "true," "approximately true," legitimate, or merely usable). Traditionally, categories are said to exist in the world, whereas concepts exist in an individual's head. The knowledge claims contained in each *Handbook* chapter describe the scientific concept for the corresponding category and its extensions, including its web of relations to other categories and their corresponding concepts.

The hypothetico-deductive method runs on concepts to build reliable and trustworthy knowledge about the categories that are assumed to structure reality. The method itself is a system of interrelated concepts.

A High Dimensional Space Of Features

It can be helpful to think about the extensions of categories (i.e., the instances that share the same intension or features of equivalence) using the idea of a geometric space from mathematics. Imagine a space defined by some massive number of dimensions (e.g., *n* dimensions, where *n* is very large), one for every feature of every category that can ever exist in nature. Every object, process, event, etc. — every instance of anything — that exists anywhere in space at any time in any context is represented as a point in this hypothetical space. (For any mathematicians reading this chapter, the hypothetical space has been discretized and any changes over time have been removed.) The coordinates at a given location, a sequence or tuple of *n* values, represent the extent to which an instance possesses each of the *n* features. (Features that are irrelevant for a given instance would have the value zero.) For example, individual animals in the everyday category CAT can be described with various physical features: A cat has whiskers, a tail, and fur of a particular length and thickness. Some cats are hairless. Different cats come in various sizes and have fur and eyes of various colors. Each animal has a genome containing some genes but not others. Each cat behaves in certain ways: It eats meat, moves with a range of agility and speed, might or might not purr when stroked, might be affectionate or aloof, and so on. Each feature listed is one of the *n* dimensions in the imaginary feature space, and each individual animal occupies a point in the space according to the amount of each feature that describes it. Likewise, every instance of every other category, be it an instance of the category ELECTRON, or the category FEAR, or any psychological category discussed in this *Handbook*, is a point in this hypothetical high-dimensional space, defined as a series of values, one for each dimension.

Proximity in this hypothetical feature space reflects similarity. Distance reflects difference. A category, such as CAT, for example, is made of instances (individual animals) that share similar values on a subset of their features (i.e., they are close to one another in a subset of the larger space). These same cats might differ in other features and will be further apart on those dimensions. And so it goes for all objects, processes, and events located in this hypothetical space. When the instances are similar enough on some set of features, they group together in portions of the feature space as an equivalence class — a category. Accordingly, an experience of one category instance should allow useful inferences about the others.

In psychological science, *categorization* occurs when a perceiver groups instances together based on one or more features of equivalence within this high-dimensional feature space, such as when a newly encountered object, person, or event is assigned to an existing category (for more on categorization, see e.g., Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1984; E. E. Smith, 1989). Categorization is the means by which perceivers infer something about the present by drawing on similarities from the past (Radulescu et al., 2021). A brain downplays differences between the members of a category, as well as magnifies differences between members and nonmembers, so that categories seem to have firm boundaries (for a discussion see Goldstone et al., 2017).

Psychological science, because of its unique history (Barrett, 2009), distinguishes features that are considered *physical* (i.e., in the world) from those that are considered *mental* (i.e., in a person's head). This dichotomy is false because every mental feature has a physical basis in a brain, even if it's not yet known. Also, a brain computes, represents, or resonates to (depending on the metaphor) everything that is sensed, making all features that humans deal with mental features of a sort. Still, the distinction is useful. Both physical and mental features are part of the hypothetical *n* dimensional space as it is discussed here.

Physical features directly involve motor changes that occur in a specific time and place. For example, the physical features of an instance of FEAR might include a particular pattern of organ and tissue changes inside the body (e.g., heart beats, respiration, hormonal changes, inflammation, etc.) and bodily movements (e.g., muscle contractions, postural changes, etc.) as well as particular vocal acoustics, electrical and chemical changes within a particular assembly of neurons and glial cells, and so on. Animal bodies have sensory surfaces that detect these changes using physical signals that hit the retina in each eye, the cochlea in each ear, the olfactory receptors in the nose, the glucose receptors in the gut and muscle cells, the stretch receptors in muscle tendons, aortic and carotid bodies to detect pH changes in the blood, and so on. For physical features that human scientists cannot directly sense (e.g., vibrations in the ground or infrared light), they use specialized technologies or instruments that translate those signals into others that they can sense directly.

Mental features, in contrast, exist in a brain that is equipped to compute them (and therefore have a physical basis, as already noted). Mental features range at one end from high-dimensional, concrete sensory features that are close to signals transduced in the body's sensory surfaces (e.g., a map of the retina in the superior colliculus and in primary visual cortex) to low-dimensional, abstract features, such as goals, affect, threat, appraisals (describing whether a situation is experienced as novel, goal congruent, etc.), value, relational themes, words, and so on (for a discussion, see Adolphs et al., 2019; Hoemann, Wu et al., 2020). Ditto for motor features, ranging from those same low-dimensional abstract features at one end to the high-dimensional, concrete features in the brainstem and spinal cord that specify particular muscle contractions and other movements of the viscera (i.e., "visceromotor" movements) and the skeletomuscular movements they support. For example, the abstract mental features of an instance of FEAR might include unpleasant or pleasant affective feelings, appraisals such as experiencing the situation as threatening or uncontrollable, and so on.

Categories Enable Inference

Inferences are educated guesses that bestow meaning. Inferences ae made possible by categories and their concepts, whether they involve scientific knowledge or planning the most basic movements an animal can make. Categories and their concepts permit generalizations from past instances (based on the features that render the extensions equivalent), allowing a brain to predict

how to deal with a new instance in the present, even when other features vary. For example, every neurotypical human has a concept for WATER that, when applied to a never-before-encountered sample of clear, colorless liquid, allows a person to drink, anticipating a reduction in thirst. A particular glass of water might vary in taste from prior samples of water, for example, depending on the mineral composition and temperature, but all things being equal, these features are unimportant to its ability to prevent dehydration. The categorization of any new instance of anything is rooted in its similarity to past instances of categories (i.e., past extensions) that have been directly experienced or learned about. Categorizing from what has been learned from one's own sensory experience belongs to the philosophical tradition of *empiricism*, whereas categorizing based on what has been learned independent of experience, such as from an authority figure or from mathematics, is known as *rationalism*.

People routinely infer beyond the information given (Bruner, 1973). For example, in the case of WATER, a person may reasonably speculate that it is possible to shower, swim, and baptize a baby in instances of water. Similar actions would be ill-advised with other clear, colorless liquids, however, such as those in the category VODKA. Inferences of new features, particularly mental features that are not directly observable, such as threat, goals, appraisals, etc., can contribute to or depend on the fundamental nature of the category and in such cases help to explain the category's existence or its causal impacts (Lombrozo, 2009; Vasilyeva & Lombrozo, 2020).

Neurotypical human brains also routinely *create* new meaning by inferring similarities using bits and pieces of past extensions combined in novel ways, called *conceptual combination* (e.g., Costello & Keane, 2000; J. A. Hampton, 1991; Medin & Shoben, 1988; Wu & Barsalou, 2009). Cognitive scientists consider conceptual combination to be one of the most powerful capabilities of the human brain.

Categories, Concepts, And Words

Strictly speaking, words are not necessary to categorize an instance of an object, a process, or an event that exists at a particular time in a particular context. A person needn't consciously use the word "water" to label a glass of clear liquid before they drink it, for example. A person's brain automatically prepares a motor plan to imbibe the liquid; this is evidence of categorization, as is the quenching of thirst that results tens of minutes before the ingested water reaches the blood and changes its osmolarity, signaling the brain of a change in hydration (e.g., Zimmerman et al., 2016). Nonetheless, words are powerful invitations to learn categories and develop concepts about them (e.g., Giffin et al., 2017; Vouloumanos & Waxman, 2014; Waxman & Gelman, 2010; Waxman & Markow, 1995). Words quickly and efficiently relay many features of an instance (e.g., a sample of WATER) with just a couple of phonemes, automatically evoking those features in another person's brain. Imagine how time consuming it would be for a person to communicate that they drank some water at dinner if they had no word for "water" and were forced to enumerate its features every time they spoke of it ("The liquid was wet and colorless and transparent and quenches thirst and is necessary for life…").

Research suggests that a word prompts a search among instances for some underlying sameness (features of equivalence) that transcends any noticeable differences (e.g., Booth & Waxman, 2002; Gopnik & Sobel, 2000; Graham et al., 2004; Nazzi & Gopnik, 2001; Welder & Graham, 2001). For example, a young child probably hears the word "water" spoken incidentally at mealtimes (e.g., "Are you thirsty? Would you like some water?"), as the child is bathed by caregivers (e.g., "Let's splash some water around…"), as other people drink water (e.g., "Ahh, I really needed this cold water!"), and

so on. From repeated observations over time, a child learns the category WATER, in effect building a concept that allows them to infer and communicate about instances of the category.

A word is powerful as an inferential tool because, in part, it may function as an abstract feature of equivalence (i.e., the intension of the concept for a category; Barrett et al., 2007; Doyle & Lindquist, 2018; Gendron et al., 2012; Lindquist, 2017; Lindquist, Satpute, et al., 2015; Satpute & Lindquist, 2021; as discussed in Barrett, 2017b). Psychological science defines the word "abstract" in several ways, but in this discussion, to "abstract" means to generalize from different complex patterns of concrete sensory and motor features to the same mental feature, e.g., to generalize from sensory patterns and motor movements to "threat" or "value" or "pleasure." Different patterns of many sensory and motor features share the same abstract feature. Many detailed sensory and motor features (patterns of high dimensionality are summarized into fewer abstract features (of lower dimensionality).

An abstract feature is never directly sensed and is not directly measurable by physical means. It is always created from physical features that are themselves sensed and measured (i.e., from concrete mental features). A brain creates abstract features as it summarizes over the temporal and spatial correlations in the ensembles of signals that it continuously receives from the sensory surfaces of the body (i.e., it removes redundancies in single ensembles) to create *compressed abstractions* of the concrete sensory and motor features. Some scientists assume these compressed abstractions are amodal (i.e., propositional representations), whereas others assume these compressed abstractions are *multimodal summaries* (Katsumi et al., 2022, 2023; this is the embodied view of cognition; e.g., Barsalou, 2008; Mahon & Hickok, 2016; Varela et al., 1992). Signal compression increases signal quality while decreasing processing cost. The result is efficient summaries of the sensory and motor signals. These summaries create similarities from differences. (Other examples of compression are MP3 audio files or the results of dimensionality reduction analyses using principal components analysis or multidimensional scaling).

Instances that exist at a specific time in a specific place will vary in their concrete sensory and motor features. Over time, if those variable, high-dimensional signal ensembles are compressed to share a low-dimensional abstract feature, then those instances form a category. Perceivers are then equipped to use their encounters with prior instances — how they acted and what they sensed — to predict and explain a new instance that shares the same abstract feature, even when past instances looked different, sounded different, tasted different, required different motor movements, and so on. This arrangement affords the ability to simplify the variation in many, many complex patterns of details in a manner that transcends physical differences. Over time, this simplification supercharges a brain to generalize from past to present and efficiently accumulate knowledge for use in the future. The capacity to experience a concrete world of physical features in terms of non-observable abstractions is a hallmark of the human mind (S. A. Gelman, 2023).

The turbo-boost in efficiency provided by words plays an important role in the development and communication of scientific knowledge. Scientists call it "jargon," such as the terms "intension" and "extension" introduced earlier. Jargon is only efficient within a community of experts who are already familiar with the concept's intension and extensions. Jargon is a barrier to communicating with non-experts who have yet to learn the concept, however. Words are a Faustian bargain in other ways as well, a topic discussed in the next section.

Categories And Concepts In Hypothesis Testing And Knowledge Claims

Categories and concepts, because of their inferential power, are the secret sauce in the hypothetico-deductive method (Feest, 2010). Concepts and categories permit two types of logical inference. *Deduction*, or deductive inference, starts with a generally true premise about a category (i.e., a concept that is considered true) and infers something about new instances of the category. The inferences necessarily follow from the premise: If the premise is true, the inferences are guaranteed to be true. Their truth is guaranteed by the truth of the original premise because no auxiliary assumptions are required, meaning the hypothesis contains no new information beyond that original premise. (In philosophy, this is called a non-ampliative inference.) As a consequence, a deductive inference *extends what is already known* about a category to future category instances. *Induction*, or inductive inference, on the other hand, starts with experiencing instances of a category and generalizes to a conclusion that *likely* holds for all category inferences, with some probability of being correct. Inductive inferences extend what is known (the philosophical jargon is "ampliative"), but they are always uncertain, unlike deductive inferences, which are deterministic and conclusive.

In principle, a scientist's success in producing knowledge claims that can be justified by scientific evidence depends on their proper use of deductive and inductive inference, as shown in the following five steps. Failures to understand the requirements and implications of these two different forms of inference have important consequences for the credibility of any scientific product arising from the hypothetico-deductive method.

Step 1: Formulate A Hypothesis

The hypothetic-deductive method begins with a topic (i.e., a scientific category) and an idea about that topic (i.e., the corresponding scientific concept). Scientists aim to hone or further develop (i.e., justify) the concept using the hypothetico-deductive method. What a scientist believes about the category at the outset — their starting concept — serves as the *initial premise* to form a hypothesis. (To make scientific observations without prior knowledge or beliefs is called *naïve induction* — as if a person comes to observations as a blank slate — and is generally excluded from the hypothetico-deductive method.)

Deduction in hypothesis formation

Sometimes an initial premise is a *scientific law*, which is a scientific concept that is considered fully justified (i.e., "true") and therefore is assumed to represent or correspond to reality regardless of time and context. A law is not necessarily something lofty. It's simply a pattern of fully deterministic relations (Braithwaite, 1953; Gardiner, 1959; Nagel, 1961; K. R. Popper, 1959/2005; e.g., Newton's laws). A hypothesis premised on a scientific law is a *deductive inference*. Some philosophers claim that this deductive inference gives the hypothetico-deductive method its name.

In the hypothetico-deductive method, hypotheses have the potential to be unambiguously falsified only when they are initially deduced from a scientific law or other premise whose bedrock truth can be confidently assumed. *Only when the initial premise is certain will the falsification of a hypothesis be an open and shut case* (see Figure 1A). Otherwise, it's a problem of estimating probabilities (Hempel, 1958; see Figure 1B).

Here is an example from the science of emotion. The psychologist Nico Frijda published a paper (Frijda, 1988) and book (Frijda, 2006) entitled, *The Laws of Emotion*, asserting that emotions behave

in a deterministic (or highly regular) manner. Frijda's Law of Situational Meaning states:

> Input some event with its particular kind of meaning; out comes an emotion of a particular kind … in goes a personal loss that is felt as irremediable, and out comes grief, with some probability. In goes a frustration or an offense for which someone else is to blame and could have avoided, and out comes anger — almost certainly. (Frijda, 2006, pp. 349–350).

Frijda continued in the next paragraph, "a number of subsidiary laws — for the elicitation of fear, of anxiety, of joy, and so forth — can be subsumed under it." (Frijda, 2006, p. 350). If Frijda's proposed laws represent or correspond to the reality of emotions in everyday life, then any hypothesis derived from them would be a deductive inference, and deductive inferences can be unambiguously falsified with observations (Steps 3 and 4, discussed momentarily; e.g., observing anger rather than fear when a threat is uncertain, or observing a person's sadness when they are offended).

Remember that deductive inferences extend bedrock, true knowledge about a category to its future instances. Such inferences work best when a category's instances are similar enough across contexts, people, and cultures for future instances to be equivalent to past instances (past extensions), either deterministically or with a very strong probability. This sort of similarity occurs when all category instances (past, present, future) share a common intension (set of features) that make those instances interchangeable. For example, if Frijda's proposed law for FEAR is an actual scientific law, then situations that involve threat of some sort will always (or almost always) produce a readiness to protect oneself, which is an instance of FEAR. Here, the intension of FEAR is an experience of a situation as threatening (called an "appraisal") and a motivation to behave in a self-protective way (called "action readiness" or "action tendency"). Other scientists have proposed other intensions for the category FEAR, such as a neural circuit (e.g., Panksepp, 1988; Tracy & Randles, 2011) or a distributed pattern of neural activity that functions like a biomarker (e.g., Zhou et al., 2021).

Deductive inferences can be descriptive, employed to test the robustness of a lawful scientific concept for a group of similar instances across varying conditions. For example, do instances of FEAR, as a readiness to protect oneself, specifically and regularly follow an experience of being threatened? Is this relation observed even in cultural contexts that have a very different concept of the category FEAR, such as people in Bali (Mesquita & Frijda, 1992; Russell, 1991a)? Deductive inferences can also be aimed at explanation. For example, does the urge to protect oneself *causally create* an unpleasant readiness to flee, freeze, faint or even fall asleep (as the Balinese are observed to do; Bateson & Mead, 1942; cited in Mesquita & Frijda, 1992)? A scientist can also start with the intension of a scientific concept (the features of equivalence) and attempt to generalize or extend that concept to predict and explain new (non-intensional) features of the category that are caused in a lawful way by the intension. For example, most instances of FEAR feel unpleasant, but can fear feel pleasant as a threat is being vanquished? How about deliberately cultivated threat for entertainment, such as viewing horror films or riding a rollercoaster?

Induction in hypothesis formation

In practice, scientists rarely deduce hypotheses from a concept as certain as a scientific law. For the most part, psychological science has no general laws (see Corbett, 2015; Roediger, 2008; Teigen, 2002). Even Weber's Law in psychophysics, which describes how people experience perceptual changes, is not a bonified law, because "law-like" responses are only obtained with very particular

experimental constraints (e.g., when yes/no answers are required; for more discussion, see Danziger (for more discussion of this point, see Danziger, 1990, pp. 136–141). Similarly, a close read of Frijda's writings indicate his use of the word "law" is metaphorical; his laws are more like concepts in which he had a great deal of confidence. It's probably not an overstatement to suggest that hypotheses in psychology are always derived from a starting premise that is considerably more uncertain than a scientific law. Every hypothesis derived from an uncertain premise is an *inductive inference* — a statement premised on a conjecture — and therefore its truth is not guaranteed by the original premise.

In psychology, a hypothesis can be inductively inferred from unexpected observations in a prior experiment — apparent anomalies — that are tentatively explained with the best available or most plausible theoretical conclusion, called an *abductive inference* (e.g., Peirce, 1903/1998); technically, an abductive inference is a special case of induction and, like other inductive inferences, less certain than a deductive inference. For example, that people do not typically scowl when they are angry, frown when they are sad, widen their eyes and gasp when they are afraid, and so on (summarized in a recent meta-analysis, Duran & Fernandez-Dols, 2021), has prompted a variety of hypotheses to explain these unexpected observations while clinging to the hypothesis that emotions are universally expressed: hypotheses involving contextual shaping, normal stochasticity, differences in the triggers for emotion, emotion regulation efforts, and cultural norms for expressing emotion (such as, for example, display rules or cultural dialects) have all been proposed as moderators that tweak or otherwise add variation to universal expressive tendencies (e.g., Cordaro et al., 2018; Elfenbein, 2013; Levenson, 2011; Matsumoto, 1990; Roseman, 2011; Tracy & Randles, 2011). Abductive hypotheses are also formed from life experiences outside the lab (e.g., noticing that prejudicial behavior against immigrants is often associated with fear); from common sense (e.g., McDougall, 1908/1921, 1923); from observations and concepts in a closely related scientific field (e.g., premising a hypothesis on the scientific concept of natural selection); or, from mathematical modeling (e.g., premising a hypothesis on a mathematical result). Hypotheses can even be abductive inferences from the opinions or writings of esteemed scientific authorities; in the science of emotion, Charles Darwin is a favorite source of inductive inspiration. When things go well, hypothesis by abduction produces clear, probabilistic expectations, but what can ultimately be learned from testing these hypotheses against observations is less certain than when testing deduction-based hypotheses, as discussed in Step 5 below.

Whether a hypothesis is derived by deduction or induction is crucial for scientific disciplines, like psychological science, that have no scientific laws to work with. The choice determines what can and can't be inferred when observations are inconsistent with hypotheses, when they are not repeatable or replicable, and when they are consistent and robust.

Step 2: Design An Experiment And Formulate A Prediction

A scientific experiment is a context that a scientist designs to curate instances of a category to experience them (a.k.a. "observe" them) to test the corresponding concept (from which the hypothesis derives). To test a hypothesis about the category ELECTRON or FEAR, for example, instances of ELECTRON or FEAR are curated in carefully controlled situations (think of the Large Hadron Collider or socially immersive scenarios, respectively). Much of the time, scientists use fancy tools to translate signals they cannot sense into those they can (e.g., a telescope, an electron microscope, an fMRI scanner, etc.). If all goes well, scientists' experiences are consistent with their concept, and they'd eventually make inferences about the entire category of interest.

A hypothesis is a generic supposition, but observations always occur in a specific spatiotemporal context, so the hypothetico-deductive method requires a scientist to transform their hypothesis into specific predictions about the experiences that they will curate (i.e., the measured variables that will quantify their experiences) in the specific circumstances of their experiment. For example, if scientists hypothesize that people across the globe reliably and specifically express instances of FEAR by widening their eyes while gasping, then the corresponding scientific predictions would involve inferences about specific facial muscle movements and their expected frequency in particular situations (e.g., the probability of certain facial muscle movements given exposure to a spider, heights, or social evaluation, and conversely, the probability that a person is exposed to a spider, heights, or social evaluation given the presence of certain facial muscle movements; see Barrett et al., 2019). Based on what is known or surmised about past extensions of the category (i.e., the concept for the category of interest), a scientist predicts how new instances will behave in a particular experimental context. This transformation from hypothesis to prediction is yet another deductive inference. (For some philosophers, *this* is the deductive inference from which the hypothetico-deductive method takes its name.)

Notice that this particular deductive inference does not depend on the hypothesis alone. It is also conditioned on *additional details* of design decisions, such as the circumstances in which the observations (i.e., experiences) will be cultivated (usually referred to as "context" or "initial conditions") plus auxiliary assumptions entailed in the measurements and other features of the study. Experimental design, measured variables, and assumptions therein are usually considered background assumptions, but they are also part of the scientific concept for the category of interest (think of the latent construct causing the observed measurements).

As noted earlier, the details of a study's design must be made public, so that others have the opportunity to curate the same observations (i.e., experiences) to test the same hypothesis; these details, which are usually reported in the methods section of published reports, are what make scientific observations public. The assumptions entailed in those design details are often left unstated, although they, too, should be made public. (One major advantage of creating a mathematical model of a hypothesis, by the way, is that mathematical formalization requires assumptions to be spelled out explicitly.) For example, when testing whether the expression of FEAR is universal, what criteria are used to confirm the presence of fear? By what method are facial movements measured? Are emotion words, such as the word "fear," part of the experiment (e.g., do they appear in the consent form, in response scales, etc.)? How variable are the stimuli that have been chosen to evoke instances of FEAR? What other conditions does the experiment contain (e.g., instances of ANGER, of SADNESS, of HAPPINESS)? Are the trials from different conditions blocked or randomized? If the hypothesis is true, then the predicted experiences should materialize, conditioned on these and other aspects of the experimental design and the assumptions they entail. Of course, it's also possible that contextual conditions and assumptions influence what scientists observe even when the hypothesis is false (e.g., Hoemann, Crittenden, et al., 2019; for discussion, see Barrett, 2022; Dubova & Goldstone, 2023).

Step 3: Observe

Scientific observations (i.e., experiences) are the "facts" of reality against which predictions and their hypotheses are tested. When a participant in an experiment contracts certain facial muscles with a certain velocity and vigor, for example, this is a fact in nature. So, too, is the variation in heart rate produced by respiration (called respiratory sinus arrhythmia or RSA) and fluctuations in the

concentration of oxygen in blood (which is the basis of blood-oxygen-level-dependent (BOLD) imaging). The *meaning* of these physical changes, on the other hand — as an expression of emotion, as changes in "stress," or as "neural activation," respectively — is up for grabs. Scientific predictions never deal directly with meanings or causes. They deal only with observations as experiences (i.e., as facts of the matter). The internal validity of a particular experiment rests on the extent to which scientists have curated extensions of the correct category (and only that category) to test their hypothesis about their scientific concept.

Step 4: Compare And Infer

When hypotheses are deductively inferred from a scientific law

Observations (as facts of reality) that do not match a prediction (hypotheses plus initial conditions, auxiliary assumptions, etc.) falsify the hypothesis from which the prediction was derived, but only if the hypothesis was deductively inferred from a scientific law. This is why the philosopher Karl Popper, who was a fan of deductive inference, claimed that hypotheses could be refuted in an all-or-none manner (K. Popper, 1959/2005). He argued that a single inconsistent observation falsified a hypothesis because he assumed that hypotheses were deductive inferences, premised on a scientific law. A mismatch between experience and hypothesis irredeemably falsifies the hypothesis. A match does not. And that's as far as it goes. Here's why.

A scientist's experience of sampled category instances, cultivated within specific experimental conditions, can be consistent with (i.e., *corroborate*) a prediction and the hypothesis from which it is derived (e.g., Newton-Smith, 1981), but such experiences increase confidence in the prediction and hypothesis only because a *hidden inductive maneuver* lurks within the hypothetico-deductive method: a willingness to generalize what is observed from a small sample of category instances to form a knowledge claim about the entire category (i.e., about the larger population of all possible category instances, including future instances). The running example includes a willingness to infer something about all or most instances of the category FEAR, past, present, and future based on comparing a small sample of experiment-based instances to a hypothesis and its associated prediction. This inference, called an *inductive generalization*, assumes that the instances curated in an experiment are similar enough to be treated as equivalent to all possible future extensions of category. In effect, to make an inductive generalization, scientists must make assumptions about the kind of category they understand themselves to be investigating –– the nature of its intension and the degree of variation in its extensions.

Hidden inductive inferences

Many scientists and philosophers routinely sneak this inductive generalization into the hypothetico-deductive method. It combines inductive inference with estimates of statistical probability to justify increasing confidence in a hypothesis that is not yet disconfirmed by observation (a technique often associated with the logical empiricist philosopher, Carl Hempel). This weaker version of hypothesis testing with the hypothetico-deductive method, called *the inductive-statistical method* (Hempel, 1942, 1965; Hempel & Oppenheim, 1948), is probably what most psychological scientists are familiar with.

Popper emphatically rejected the logical utility of inductive inferences (following the philosopher David Hume; Hume, 1739/2004). The rationality of the scientific method, in Popper's mindset, depends *only* on deductive inferences from experiences that place concepts at ever-increasing amounts of risk of falsification. Period. No exceptions. Popper correspondingly rejected the notion that scientists tiptoe toward truth every time their hypothesis passes a test of experience. Here's why: No matter how many concept-consistent instances are experienced, concept-inconsistent instances still might materialize in the future (i.e., a law is always at risk of being disproven; for discussion, see Salmon, 1981). For example, the fact that fearful participants widen their eyes and/or gasp more than would be expected by chance in one experiment, 10 experiments, or even 100 experiments can never increase confidence in the corresponding hypothesis (i.e., that "people universally express instances of FEAR by widening their eyes and gasping"). Inductive inferences, therefore, never support a claim that a hypothesis is true, approximately true, or moving in the direction of truth, according to Popper, because a hypothesis can be felled by one future concept-inconsistent instance, provided that the hypothesis is derived from a scientific law (or a premise that is equally deterministic).

For scientists who rely on inductive inference, growing confidence in a hypothesis also supports a special sort of inductive inference called an *explanatory inference* — a provisional conceptual explanation for what was experienced (observed). This is how wide-eyed gasps become "fearful facial expressions" and scowls become "angry facial expressions," changes in RSA become "stress," and changes in BOLD signal during fMRI scanning become "neural activation." But there are still some tricky bits to deal with. How strong a match between experience and hypothesis is required to justify such inferences? And how much should confidence grow after each test? Pragmatic, social and political (i.e., non-epistemic) factors influence how scientists resolve these puzzles, giving philosophers reason to question whether the hypothetico-deductive method is a rational way of building justified knowledge (see Hacking, 1983; also see the section below on traditional scientific realism).

There is one other opportunity for explanatory inference when using the hypothetico-deductive method: when observations don't repeat or replicate. That is, when experiences of (what scientists take to be) instances of a given category are inconsistent with predictions, scientists may decide that the experiment was flawed rather than reject their hypothesis. Scientists may wonder, for example, if they inadvertently curated instances of other categories, explaining why their observations varied too much across time, situations, or people to produce robust results (e.g., stimuli that evoke emotions in the lab are too weak; e.g., Levenson, 2011). Or they may indulge in abductive inference to explain why the predicted observations did not materialize as expected (see examples of contextual shaping, emotion regulation efforts, display rules, and cultural dialects in Step 1, above).

Popper distained such inductive maneuvers (K. R. Popper, 1959/2005, 1963/2002). As scientific inquiry moves forward, scientists will of course refine their hypotheses. This is to be expected. But one thing a scientist should never do, according to Popper, is

> react to the falsification of one conjecture by cooking up a new conjecture that is designed to just avoid the problems revealed by earlier testing … We should not make ad hoc moves that merely patch the problems found in earlier conjectures. (Godfrey-Smith, 2021, p. 83).

For Popper, increasing boldness in testing hypotheses was the name of the game.

A not-so-hidden inductive inference

The discussion so far has focused on how the hypothetico-deductive method treats a hypothesis that was *deductively* inferred from an established scientific law. What happens when a hypothesis was *inductively* inferred from a more uncertain initial (conceptual) premise? Readers who work in a scientific field with no scientific laws, such as psychology, should pay close attention here.

By the formal rules of logic, a mismatch between experiences and a hypothesis can *never* falsify an inductively derived hypothesis. The squishy inductive inference that gave rise to a hypothesis is insufficiently deterministic (or probable) to support predictions that can be adjudicated by observations with any degree of confidence. For example, say a scientist begins with the starting premise that emotions are inherited, innate reactions to situations that risk life and reduce reproductive success. Fear, in this scenario, is assumed to be a mechanism that evolved to protect an organism from predatory threat (e.g., Mobbs et al., 2020). This idea is not a scientific law. It is a conjecture that requires numerous auxiliary assumptions in addition to the original premise to arrive at a hypothesis.

Continuing the example, the scientist, based on their conjecture, may hypothesize that subcortical neurons within the amygdala (a small cluster of neurons deep in the temporal lobe of mammalian brains) and the periaqueductal gray (a small cluster of neurons in the midbrain) innately create instances of FEAR (e.g., Fanselow, 2018; i.e., the neural circuits are hypothesized to be the intension of FEAR). Many assumptions are embedded in this (ampliative) hypothesis that go beyond the original premise from which it was inductively derived: the notion that organisms are passive vehicles for retaining genes in a gene-pool across generations; that genes are the only conduits of inheritance and provide a blueprint for building a brain; that evolution occurs primarily through small random changes in genes that are selected to allow organisms to reproduce more and/or make offspring more likely to survive to reproductive age; that behavior is controlled by genes in this way, and so on. These assumptions belong to a research tradition in evolutionary biology called the *modern evolutionary synthesis* (e.g., Dawkins, 1976/2006; Huxley, 1942; Mayr, 1982); it is common in psychology but is outdated when compared to the more recent *extended evolutionary synthesis*, which shows these assumptions to be false (e.g., Ball, 2024; Jablonka & Lamb, 2014; K. N. Laland et al., 2014, 2015; Lewontin, 2000; Müller, 2017; Noble & Noble, 2023; Richerson & Boyd, 2006).

The premise in this example — that emotions are inherited, innate reactions to situations that risk life and reduce reproductive success — could therefore be false, and in fact has been contested by other evidence (for discussion and selected references, see Barrett, 2017b; Barrett et al., 2019). When the premise of a hypothesis is false, then *the hypothesis is also necessarily false, even if corroborated by observations.* Even if the premise turns out to be justified, experimental observations still cannot logically evaluate the hypothesis against the facts of experience with any degree of confidence because the original premise of the hypothesis is uncertain — it contained auxiliary assumptions over and above the premise (i.e., the hypothesis is ampliative, it contains additional information beyond the original premise), and any one of those assumptions can be in error. Without the rock-solid foundation of scientific law as the initial premise, an inductively derived hypothesis that is inconsistent with observations cannot be confidently rejected. In fact, the history of science instructs that such hypotheses can, in fact, end up being justified even when incorrect auxiliary assumptions are made (see section on justifications and critiques of traditional scientific realism and also Appendix, Section 1). Nor can a hypothesis that is consistent with observations be confidently retained. Observations can be fully reproducible and still be wrong (Leek & Peng, 2015). Many examples in the history of science attest that, for inductively derived hypotheses, corroborating observations can be a poor indicator that they are true or approximately true (see Appendix, Section 2). This latter point is discussed in more detail in the next section.

One final note about inductive inferences, at least for the moment: When inferences to hypotheses arise from beliefs about the workings of genes or the opinion of an eminent scientist (such as Charles Darwin), a scientist is practicing a version of rationalism. Rationalism assumes that the most reliable knowledge is *given*, rather than learned, and therefore such knowledge must originate from sources that are independent of (and maybe even a priori to) personal experience. Philosophers who favor rationalist assumptions also tend to prefer the rock-solid deterministic logic of deductive inference (like Popper). This loose association between rationalism and deductive inference can cause problems: If scientists begin by making inductive inferences from sources other than experience to derive their hypotheses (i.e., they behave like rationalists who derive their hypotheses from beliefs about innateness, how genes work, the writings of Darwin, etc.), they may mistake their inductive inferences for deductive inferences, and attribute an unearned certainty the former that only the latter actually affords.

Step 5. Repeat Or Pivot

When a scientist retains a hypothesis, they should repeat the entire hypothetico-deductive method again. And again. And yet again, each time testing their hypothesis in a manner that is *riskier* than the last (K. R. Popper, 1959/2005, 1963/2002). At least that's what's supposed to happen. Ideally, other scientists join the effort to falsify the hypothesis by cultivating experiences with experiments that increasingly put the hypothesis at risk. To this end, experiments should sample different instances of the relevant category on different occasions, using different participants, in different contexts, with different experimental methods, to be experienced and compared to the corresponding concept. If the hypothesis was premised on anything other than a law, then the hypothesis should be tested under a variety of auxiliary assumptions; for example, what might a hypothesis about emotion look like if premised on the assumption that genes are not a blueprint for endowing emotion circuits into a brain?

If all goes well, hypotheses repeatedly survive such tests, increasing everyone's confidence that the associated scientific concept provides the best available summary of the category of interest (except for those who think like Popper; Appendix, Section 3). Scientists might never be one hundred percent certain that their hypotheses and associated scientific concepts are True with a capital T, but the probability of truth inches forward. Over time, accumulating successes invite scientists to accept the concept as justified knowledge, approximately true knowledge, or at least provisionally not falsified knowledge about the category of interest: reliable, factual knowledge about a category's features of equivalence (intension) and their causes, including how category instances will behave across different contexts and how they will impact instances of other categories. And everyone lives happily ever after.

What is another name for reliable, factual knowledge? *A scientific law*. When all goes well, scientists either test an existing scientific law (if a hypothesis was deduced from one in the first place) or they discover a scientific law (if a hypothesis was initially premised on inductive logic). That scientific laws always involve inductive inferences at some point constitutes a philosophical Achilles heel for the hypothetico-deductive method: Even rock-solid deductive inferences that are premised on laws initially involved squishy inductive inferences of some sort. Where building knowledge is concerned, it's inductive turtles all the way down.

This is how the hypothetico-deductive method works when things go well. What about when they don't? Is it reasonable to reject a hypothesis that fails the test of experience? The answer is: It depends on the initial premise. If the hypothesis in question was deduced from a scientific law, then

yes. If it was inductively inferred from any other source, which will be less deterministic than a law, then no. What does it mean to a failure to reproduce or replicate observations, then? Well, here is where things get interesting.

Consider the hypothesis that instances of FEAR are caused by a circuit in the amygdala. A scientist wanting to test that hypothesis may attempt to cultivate fear in participants by exposing them to stimuli that are supposed elicit threat while scanning their brains. The scientist predicts that the intensity of blood-oxygen-level-dependent (BOLD) signal increases in the voxels containing amygdala neurons will correlate with intensity of perceived threat and experienced fear (a voxel being a three-dimensional volume in the depiction of a brain). Instead, no such correlation is observed. So, what happened? It's possible that the hypothesis is not justified. Or maybe there were some shenanigans in the original experiments that produced increased BOLD signal during exposure to threatening stimuli. But there is a laundry list of possible explanations for why the experiment "failed" (meaning, the hypothesis did not pass the test of experience) that have to do with faulty auxiliary assumptions involved in creating the hypothesis from the original premise. Maybe the stimuli weren't intense enough to elicit authentic fear. Maybe the experimental situation was too artificial. Maybe the BOLD signal is not a measure of generic neural activity (Theriault et al., 2023). And so on. Or there could have been an error in assumptions about the nature of the FEAR. Perhaps category instances are more contextually determined and therefore not similar enough to be equivalent or interchangeable across contexts, people, and cultures. For such categories, similarity is based on a *particular function* in a particular *situation*. The hypothetico-deductive method is less suitable for building justified knowledge about such categories: If a category's existence and nature is not independent of or merely modified by context but is fully contingent upon context, then it is virtually impossible to experience a small sample of category instances and generalize that experience to the entire category of past, present and future extensions.

Every category is a simplification of the variation and uncertainty in an ever-changing world. A category represents the equivalence among instances that exist in a particular time and space — instances that vary in their concrete, high-dimensional features. A category reduces the dimensionality of the hypothetical feature space into a smaller subspace of relevant features. Too much simplification produces fictional categories that do not exist in nature, categories with limited utility for acting on a specific instance in a given spatio-temporal context. Too little simplification produces a category with limited generalizability and, again, with limited utility. The trick is to find the sweet spot, the *optimal constraint of variation in the name of sameness*. This conundrum has been called the complexity paradox (Levins, 1966, 1968; Oreskes, 2003). On a continuum from simplicity and sameness at one end (in which a categories instances share identical features regardless of context) to massive complexity and variation at the other (where every instance is its own category), the hypothetico-deductive method makes certain assumptions about the location of that sweet spot. These assumptions concern nothing less than the *nature of reality*. When scientists use the hypothetico-deductive method to study categories that do not meet these assumptions, they are fundamentally misunderstanding the nature of reality. The consequence: A replication crisis of major proportion.

Unexpected variation, in the guise of failures to replicate observations across studies, can be evidence that scientists are attempting to build knowledge about categories that are overly simplified fictions. In such as case, they misunderstand the very nature of the phenomena that they are attempting to know. The next section explores this idea while introducing the various kinds of categories that have been assumed to exist in science writ large. When assumptions about the

nature of reality do not match reality as it exists, this mismatch is a barrier to building robust, justified knowledge in science, even when methods are executed with maximal methodological rigor.

CONVENTIONAL ASSUMPTIONS ABOUT THE NATURE OF REALITY IN PSYCHOLOGICAL SCIENCE

To develop justified knowledge about reality, every scientist necessarily conditions their efforts on beliefs about the part of reality they are trying to know. Questions about the nature of reality fall within the domain of ontology, and beliefs about the nature of reality are known as *ontological assumptions*. Such assumptions, even if implicit, guide all scientific work. They lurk within the scientific questions asked, the scientific practices and inferences employed to answer those questions, the claims to knowledge that scientists published, and the claims of others that they decide to cite (or not to cite). Ontological assumptions are even related to beliefs about the nature of knowledge and what makes it trustworthy (i.e., epistemic assumptions and practices).

The history of philosophy is full of ontological debates about reality that boil down to fundamentally different ideas, or concepts, about the nature of the categories that constitute reality. Philosophers call them *kind concepts*. Every scientific method involves a kind concept, as does every experiment and knowledge claim. In psychological science, knowledge about human nature presumes something about the kind of categories that make up a human mind.

This section summarizes two kind concepts commonly used in psychological science: *natural kind categories* and *empirical kind categories*. Each kind concept is associated with a broader philosophical tradition and philosophy of science in psychology. This section describes the main assumptions that each kind concept entails, including how well each works with the hypothetico-deductive method, and the sorts of scientific inferences each supports. Examples from various scientific domains, including the science of emotion, are used throughout to illustrate when a given kind concept is an asset or an obstacle to building justified scientific knowledge.

Ontological assumptions usually run strong and deep. Kuhn understood this when he remarked, "the proponents of competing paradigms practice their trades *in different worlds*… [in which] scientists see different things when they look from the same point in the same direction." (Kuhn, 1962/1970, p. 150). When scientists find a particular kind concept ridiculous, implausible, or intellectual naïve, that is a pretty good clue that they inhabit a different conceptual world, using a different kind concept, and therefore are committed to some *other* set of beliefs about the nature of reality.

Natural Kind Categories

One kind concept, called a *natural kind category*, is the commonsense view of reality. A natural kind's features of equivalence (the category's intension) — whether concrete and directly observable or abstract and inferred — are assumed to exist in the natural world and remain stable and unchanging across time and context (i.e., the features of equivalence are *immutable* and *universal*). Natural kinds are all about simplicity and sameness. Standard examples of proposed natural kinds in science include ELECTRONS and QUARKS in physics, elements of the periodic table in chemistry

(HYDROGEN, HELIUM, etc.), and types of cognition and emotion in psychology. To a psychological scientist who believes in natural kinds, LOVE is as real in nature as ELECTRONS, GOALS are as real as HYDROGEN, and ATTITUDES are as real as CATS.

The notion of a natural kind category is itself a human-invented concept: a set of assumptions that there exist natural groupings in the world. Such categories are assumed to correspond to, represent, or track groupings in the world that themselves are not dependent on humans. This is why a natural kind is considered *mind-independent*, *perceiver-independent*, *observer-independent*, or *perspective-independent*, all of which mean that the similarities between instances within a category, and the differences between categories, are presumed to be independent of any human concepts. In short, natural kinds neatly divide reality into universal groupings like dictionary definitions (more or less).

From these characteristics of natural kinds, others follow. All extensions of the category possess the same features of equivalence, making those features necessary and sufficient. The category's intension also governs and explains, in a deterministic, law-like manner, how category instances behave and interact with instances of other categories. These characteristics give a natural kind its inductive, deductive, and explanatory power. For example, a wide swath of research programs in the science of emotion assume that ANGER, FEAR, AWE, and a handful of other emotion categories named with English words — upwards of 25 or so — are universal, psychological groupings. These programs include basic emotion theories (e.g., Cowen & Keltner, 2021; Elman & Cordaro, 2011; Panksepp, 1998), discrete emotion theories (e.g., Izard, 1977, some functional theories (that characterize an emotion category according to its purpose (e.g., Adolphs & Anderson, 2018) and certain appraisal theories (that define appraisals as causal mechanisms; e.g., Lazarus, 1991, 2001; Roseman, 2011; Scherer, 2001) (for discussion, see Barrett, 2017a, 2017c, 2022; Barrett & Lida, 2024; for historical references, see Gendron & Barrett, 2009 and references therein). These research programs are considered *typological* approaches to emotion in which each category is a "type," a family of biologically prepared states that are functionally equivalent by virtue of sharing highly similar pattern of diagnostic physical and mental features (intensions) with similar causal impacts and consequences.

The epistemic goal of a typological approach to science is to discover the natural categories that exist in reality to better predict, understand, and deal with reality. Philosophers as early as the 1840s debated the extent to which scientific categories (such as SPECIES, GENUS, FAMILY, etc.) were a biological reality *discovered* by humans vs. *invented* by them (Hacking, 1991). Since then, philosophers have written quite a bit about natural kinds, full of nuanced distinctions, the gist of which serves our purpose here. Interested readers are referred to the original texts (Boyd, 1979, 1984; Dupré, 1981, 1993; Hacking, 1991; Koslicki, 2008; Kripke, 1972/1980; Locke, 1690/1964; Mellor, 1977; Putnam, 1975b; Quine, 1977; Schwartz, 1977, 1980; E. E. Smith & Medin, 1981; Sober, 1980; see also, Bird & Tobin, 2024).

In the hypothetical feature space described earlier, the necessary and sufficient features of a natural kind (its intension) are unique enough to create a grouping with sharply delimited boundaries. This grouping is therefore categorically distinct from other groupings (which themselves share their own distinct sets of dimensions/features and clear boundaries), so that instances of a natural kind category cluster close together but remain distant from other such groupings. In fact, the instances of a given natural kind category cluster so closely that they occupy only a small subspace of the much larger *n*-dimensional feature space. Hence the familiar metaphor to "carve nature at its joints."

The world — rather than human goals, interests, or concepts — compels a scientist to group phenomena in this one particular way.

Essentialism

When features of equivalence are universally shared by all members of a category, are unique to those members, and are presumed to be objective (i.e., mind-independent), they are called the natural kind's *essence*, in accordance with a philosophical perspective called *essentialism*. Essential features are said to define the category, giving members a unique and characteristic identity. They should explain in a universal and law-like manner all instances and causal relations of the category. For example, in a basic emotion approach, certain emotion categories, such as FEAR, ANGER, SADNESS, and so on, are thought to constitute evolutionarily preserved, biologically inherent (i.e., genetically prescribed), universal groupings, the instances of each issuing from a specific, dedicated assembly of neurons in human or mammalian brains (e.g., Panksepp, 1998; Tracy & Randles, 2011). The neural assembly is the essential feature of equivalence (the intension) that supposedly makes all instances of the category FEAR, for example, a useful adaptation for reacting to and dealing with threat from a predator. For the purposes of definition, prediction, and explanation, instances of a natural kind category are interchangeable due to their shared essence. (Instances may vary in non-essential features that are irrelevant to the way that the kind behaves.) This category structure allows a scientist to curate and experience a small sample of instances of a category such as FEAR in study participants, aggregate across those instances, and make inferences about the entire category FEAR. Essences, in this view, correspond to the concept of the category (Barnes, 1981).

The notion of an essence is usually traced to the Ancient Greek philosopher Plato, who proposed that every natural kind category was defined by a universal abstract "form" that existed beyond time and space (for review, see Kraut, 2022). For example, Plato believed that all instances of WATER, no matter how much they vary from one another in temperature, viscosity, taste, etc., derive their nature from the perfect form of "water." Likewise, all instances of FEAR, no matter how much they vary from one another, should derive their nature from the perfect form of "fear." And so on. Plato's student Aristotle replaced the notion of abstract, universal forms with the notion of an essence: the defining features that give category instances their nature. So, putative essences of WATER or FEAR are the features that are necessary and sufficient to instances of WATER and FEAR, without which an instance would not belong to one of these categories. This is why essences are assumed to function like dictionary definitions.

In cognitive science, natural kind categories with essences have all the elements of *classical categories* (E. E. Smith & Medin, 1981). The concept for a classical category is considered "semantic," or a "type," meaning that there is one concept per category and its features of equivalence (intensions) and members (extensions) are stable across time and place. When organized into a scientific classification scheme — a *typology* — scientific concepts for natural kind categories should reflect the real similarities and differences between and among category instances in nature (hence the observation that certain research programs in the science of emotion are "typological" approaches). The only way that two natural kinds can share overlapping features, or that an instance can belong to more than one category, is when one category is a subset of the other (Ellis, 2001, 2002; Kuhn, 2002).

Functional Kinds

One version of a natural kind category, called a *functional kind*, has played an important role in many domains of science, including psychology. A functional kind is a group of instances that are similar in what they *do* — their function — rather than in how they are physically constituted. For example, in biology, macromolecules such as proteins are classified by the functions they serve, not by their concrete physical properties (Slater, 2009). Functions are features that cannot be directly observed via the senses; they must be inferred, meaning they are abstractions, by definition. A functional feature manifests in multiple concrete ways (called "degeneracy" or "multiple realizability"), so that instances of the same category, such as PROTEIN, differ in their physical features (which are not part of the category's intension). Similarly, instances of the category FEAR are said to differ in their physical features across species, even as they share the same functional intension (an internal "state" for reacting to and dealing with threat; e.g., Adolphs & Anderson, 2018).

A stable, universal, functional feature of equivalence is often an example of *psychological essentialism*: the intuition that a category's members share a deep, essential similarity, *even when none has been identified* (S. A. Gelman, 2003; Medin & Ortony, 1989; Prentice & Miller, 2007). An example is the idea that racial and gender categories have biological essences, beliefs that have been linked to racial and gender prejudice and stereotyping (for discussion, see Haslam & Whelan, 2008). Scientists sometimes employ a hypothetical placeholder for such an essence — an imagined set of features that gives a category its identity and distinctiveness and explains the kind's behavior. For example, scientists assumed for many years that each category of mental illness had an as-yet-undiscovered specific and unique set of genes. That is, the ensemble of genes presumed to cause every instance of a category such as MAJOR DEPRESSIVE DISORDER would be distinct from those causing instances of POST TRAUMATIC STRESS DISORDER. This belief persisted for many years despite repeated failures to identify the essential gene(s). Many scientists now take a transdiagnostic approach to mental illnesses, assuming that cross-cutting symptoms, such as instances of NEGATIVE AFFECT, POSITIVE AFFECT, AROUSAL, and so on, each have a distinctive biological essence. This belief fuels the search for genetic essences despite recurring failures to identify them. Some of clearest examples of psychological essentialism are found in typological approaches to emotion, such as unseen and hypothetical affect programs (e.g., Ekman & Cordaro, 2011; Tomkins & McCarter, 1964), hypothetical neural circuits (e.g., Panksepp, 1998), and hypothetical "central emotion states" (Adolphs & Anderson, 2018; D. J. Anderson, 2022). The idea that an evaluative mechanism called an "appraisal" causes a descriptive mental feature of the same name is yet another example of psychological essentialism (e.g., that experiencing a situation as relevant to your goal issues from a causal mechanism that evaluates goal relevance; e.g., Scherer, 2009; Scherer & Moors, 2019).

Probabilistic Feature Clusters

Must every natural kind have an essence? This question is a philosophical point of ongoing contention. Certain scientific categories once assumed to be quintessential natural kind categories, such as animal species, do not have essences (e.g., Boyd, 1991, 1999; Mayr, 2004; Millikan, 1999; Quine, 1969; Sober, 1980). A category of animal species has no features that are shared by all individuals of that species and by only individuals of that species. Even if such features were identified, they would be easily and continuously modified by mutation, recombination, random drift, and other processes of evolution. These considerations, plus others, led the evolutionary biologist Ernst Mayr to claim, "It took more than 2,000 years for biology, under the influence of Darwin, to escape the paralyzing grip of essentialism" (Mayr, 2004; also see Ereshefsky, 2018; Sober,

1994; Wilson et al., 2007). The eventual discovery that animal species do not (and really, cannot) be defined by a genetic essence prompted a crisis: Either important categories in science lack essences and therefore are not natural kind categories, or natural kind categories don't require essences. Either outcome was a problem for a science aiming to build justified knowledge about entire categories by observing small samples of their instances.

Ultimately, the solution was to broaden the concept of natural kinds by redefining it — yes, redefining what exists in reality — to include phenomena for which essences are unlikely or implausible, thereby permitting animal species to remain natural kinds (e.g., Boyd, 1991; Ereshefsky, 2018; Sober, 1994; Wilson et al., 2007). Features of equivalence were redefined to be probabilistic rather than universal and obligatory, variable rather than rigidly uniform, and related to one another rather than independent (for discussion, see Hacking, 1991). The intension of a natural kind category came to be defined as a tight cluster of features that reliably co-occur in variable subsets of category instances rather than an obligatory, fixed set of necessary and sufficient features that provide absolute and uniform similarity among instances.

In this probabilistic account of natural kind categories, natural kind-ness is not due to individual features but to the *relationships between co-occurring features*. Features of equivalence are described as "homeostatic property clusters" (Boyd, 1999) or "stable property clusters" (Slater, 2015). A given cluster of features is said to co-occur either because the features depend on each other probabilistically or because they share an underlying common cause. No single feature of equivalence is considered intrinsic or individually necessary to the identity of a kind, and recurring clusters of features can also be somewhat more dynamic with time and context. The similarity between instances that unites them as an equivalence class is family resemblance (Wittgenstein, 1973), not uniformity (again, see Hacking, 1991).

Notice the fancy philosophical maneuver here: The inferential success of a category became evidence that the category must be a natural kind. A grouping of similar instances is assumed to be real in nature if their features cluster together in a way that permits probabilistically reliable inductive and explanatory inferences about that category. This logic is a flavor of the maxim, "if it works, it's true," which turns out to be a major argument for the broader philosophy of science to which natural kinds belong, called *traditional scientific realism* (discussed below in more detail; this justification of traditional scientific realism is called "inference to the best explanation"). Note that natural kinds, as groupings of instances with stable property clusters, are now defined *functionally* by what one can do with them (i.e., by the epistemic practices that they support) rather than by (ontological) assumptions for how the world is. Moreover, this shift relaxes any claim about natural kinds being mind-independent because a natural kind is now defined by how a human mind uses it (for a similar point, see Reydon, 2016).

When the definition of similarity is more probabilistic, features of equivalence can vary across extensions of the category, making the presumed groupings more like *prototype categories* than classical categories (Mervis & Rosch, 1981; Rosch, 1975; Rosch & Mervis, 1975). In cognitive science, a prototype category is a grouping of individual instances that can vary somewhat in their features of equivalence, creating a distribution of similarity or family resemblance. It is assumed that the similarity distribution is summarized by a single instance, called the prototype (Rosch & Mervis, 1975), and together the distribution plus the prototype are the concept for the category. The concept is considered "semantic," like a classical concept, stable across time and place. The category's prototype might refer to any of several feature clusters: those belonging to the single most frequent instance of the grouping or the most typical instances (J. Hampton, 1979; Rosch & Mervis, 1975) or

the most representative or ideal instance (Borkenau, 1990; Chaplin et al., 1988; Loken & Ward, 1990; Lynch et al., 2000; Read et al., 1990; Voorspoels et al., 2011, 2013) even if it has never been observed and must be inferred (e.g., Posner & Keele, 1968). The instances within the grouping are assumed to share a family resemblance with one another because of their family resemblance to the prototype.

Prototype categories don't "carve nature at its joints" because there are no joints to carve. Instead, they have fuzzy boundaries by design, meaning instances from distinct categories can share some features. To return to the hypothetical feature space, each prototype category would exist in its own subspace of features within the larger feature space. Within-category instances would cluster to reflect their family resemblance. The subspaces for different categories might overlap somewhat (fuzzy boundaries), but their prototypes should not. The concept for each category would be found at the position for the most typical (frequent), most representative, or most ideal instance of the categories.

Some typological approaches to emotion offer a good example of scientific efforts to identify natural kind categories as prototypes, without essences (e.g., Cowen & Keltner, 2021; Ekman, 1992; Roseman, 2011; Shariff & Tracy, 2011; Shaver et al., 1987). A prototype category of emotion is defined as a grouping of instances with a family of related features. Across situations, people, and cultures, the features of instances within the same category are assumed to vary to some extent, creating a distribution (or family) of related features. Distributions for different categories can overlap (i.e., have fuzzy boundaries), meaning some instances from different categories can share the same features. Even if these prototype categories don't have a singular essence, they are still natural kinds if their prototypes do not overlap and are sufficiently regular (and therefore diagnostic enough) to permit inductive and explanatory inferences. The category's prototype, as a conceptual representation of the entire distribution (i.e., its intension), is assumed to share a family resemblance with all extensions of the category — a pattern of features that is similar enough to the other category instances in the distribution, and different enough from the prototypes of other categories, to permit reliable inferences (for specific quotations, see Ekman, 1992, p. 197; Ekman & Cordaro, 2011, p. 364; Levenson, 2011, p. 379; Scarantino & Griffiths, 2011, p. 448-449).

Prototype approaches to natural kinds of emotion admit that the features of emotional instances vary but propose that this variance is caused by moderating factors, a.k.a. "hidden moderators" (also called hidden assumptions; e.g., Zwaan et al., 2018 as cited in Framework for Open and Reproducible Research Training; https://forrt.org/): contextual influences that are thought to be epiphenomenal to the main mechanism that causes instances of a given emotion category and are therefore separate from the underlying emotion prototype itself. Proposed moderating factors have included contextual shaping, normal stochasticity, differences in the triggers for emotion, emotion regulation efforts, and cultural norms for expressing emotion (such as, for example, display rules or cultural dialects; e.g., Cordaro et al., 2018; Ekman & Cordaro, 2011; Elfenbein, 2013; Levenson, 2011; Matsumoto, 1990; Roseman, 2011; Tracy & Randles, 2011). In the prototype view of natural kinds of emotion, for example, a particular momentary fearful expression may involve prototypic gasping with widened eyes, whereas in other moments could involve crying or even smiling, but these variable instances nonetheless remain expressions of the underlying emotion prototype if the variation can be explained by these other moderating factors.

To hypothesize the existence of a moderator is an abductive inference, as discussed earlier in Step 4 of the hypothetico-deductive method. Knowledge claims about possible moderators may be bona-fide discoveries but could also be an attempt to patch up a problem with the original hypothesis of natural kinds of emotion types, something Popper warned scientists should never do. To this point,

a recent survey (Ekman, 2016) found that a substantial number of psychological scientists who are experts in the science of emotion (80% of respondents) accept some form of a natural kind view, despite the large magnitude of variation observed in instances of the same emotion category and recurring similarities in instances of different categories (e.g., Azari et al., 2020; Doyle et al., 2022; Durán & Fernández-Dols, 2021; Gendron et al., 2018; Gendron, Hoemann, et al., 2020; Guillory & Bujarski, 2014; Halgren et al., 1978; Hoemann, Crittenden, et al., 2019; Hoemann, Khan, et al., 2020; Jack et al., 2016; Khalaf et al., 2020; McVeigh et al., 2023; Raz et al., 2016; Reynolds & Berridge, 2008; Siegel et al., 2018; Srinivasan & Martinez, 2021; Wilson-Mendenhall et al., 2015). This acceptance signals a *philosophical commitment* to the natural kind status of certain emotion categories. This commitment may help explain why an abundance of scientific effort continues to be devoted to an empirical search for prototypic emotion features that reliably characterize and distinguish each emotion category as biologically and psychologically distinct (for a discussion, see Barrett, 2022) rather than follow Popper's directive to subject a hypothesis to ever bolder attempts at falsification.

Evidence for a philosophical commitment to the natural kind status of certain emotion categories can also be found in experiments designed to test their prototype structure (discussed in Azari et al., 2020; Barrett, 2022; Westlin et al., 2023). For example, stimuli are not sampled randomly but in a manner that presumes the existence of prototypes, often guided by pre-determined or commonsense beliefs about emotion. Sampling many participants with few trials (i.e., shallowly) is assumed to yield similar observations to sampling fewer participants with more trials (i.e., densely), because category instances are assumed to be relatively similar across people and contexts. Measurements are aggregated across instances of the same category within a person, and even across people, in search of commonalities rather than modeling structured variation within or across categories. When experiments are laden with the same concepts as those that give rise to the hypothesis being tested, the distinction between scientific concepts and the observations curated to test them is weakened (for discussion, see Dubova & Goldstone, 2023). This problem of *concept-laden scientific work* is not unique to the science of emotion. Philosophers of science have long cautioned that observations can never be a neutral test of a hypothesis because they are not free from the scientific concepts they are testing (e.g., Feyerabend, 1975/2010; Hanson, 1958; Kuhn, 1962/1970; Quine, 1951b). On occasion, an experimental setup actually creates the very observations that are needed to confirm a concept where they otherwise would not exist (for an example, see Hoemann, Crittenden, et al., 2019; see Appendix, Section 4).

Finally, it's important to note that not all prototype emotion categories that have been proposed are considered natural kinds (Barrett & Lida, 2024). Prototypes allow for but do not require the categories in question to be natural kinds. Since the 1980s, critics of the typological perspective who take a constructionist approach have also suggested that emotion categories are structured as prototypes (e.g., Fehr & Russell, 1984; Russell, 1991b, 2003), suggesting that emotional meanings are learned during development rather than innate. This is a psychological constructionist theory of emotion, not a typological theory, and hypothesizes that emotional meanings are constructed on the spot when an occurrence of high-dimensional concrete features that exist in a specific time and space is categorized with a prototype concept. To be a natural kind, the probability structure of the features must be sufficiently universal to permit inductive, deductive, and explanatory inferences about the entire category from observing specific spatiotemporal samples of its instances.

The Role Of Natural Kind Categories In Psychological Science

Crack open almost any North American or Western European introductory psychology textbook and take a gander at the table of contents. Most likely, it's a typology of presumed natural kinds, including kinds of knowing (categories such as PERCEPTION and COGNITION), wanting (categories such as GOALS, INTRINSIC MOTIVATION, ANGER, SADNESS, HAPPINESS, etc.), and doing (categories such as AUTOMATIC vs. CONTROLLED PROCESSING, HABITUAL vs. DELIBERATE BEHAVIOR, REFLEXIVE vs. WILLFULL or VOLITIONAL ACTION, and so on). Some version of this typology has organized the content and professional activities of psychology since it transitioned from a branch of mental philosophy into a formal science (Barrett, 2009). The modern form of this typology originated in 17th century European mental philosophy known as *faculty psychology* (Klein, 1970), although its historical tendrils can be traced to Plato and Ancient Greece (Danziger, 1997). Sometimes this typology is called *folk psychology* or common-sense psychology (e.g., Churchland, 1981; James, 1890/1931; Wundt, 1897/1998), on the assumption that a universal human mind can be described and studied as categories that derive from everyday human experiences. (Mainly Eurocentric experiences, most likely.)

Here's the gist of a psychological science committed to natural kinds: Certain mental categories are assumed to accurately and objectively describe a *single* human nature shared by all (neurotypical) humans. Every neurotypical human is presumed to think. Instances of THINKING are presumed to exist independently of any thoughts about thinking. Every human is presumed to feel. Instances of ANGER are presumed to exist independently of any feelings about anger, even for people in cultures that do not possess a concept for anger. Instances of GOALS are presumed to exist independently of any ideas about goals or motivation. Psychological faculties, as natural kinds, are also assumed to be categorically distinct from one another. Humans are assumed to think and feel as separate abilities, even in those cultures that do not have separate concepts for thinking and feeling (e.g., Balinese culture and Ilongot culture in the Philippines have words for mental activity that are best translated as "thought-feeling."; W. M. Reddy, personal communication, September 16, 2007; Rosaldo, 1980; Wikan, 1990). An instance of EMOTION is not an instance of COGNITION. Nor is an instance of VISUAL PERCEPTION to be confused with an instance of EPISODIC MEMORY. An instance of IMPLICIT MEMORY is distinct from an instance of ENDOGENOUS ATTENTION. Instances of EMOTION are either inputs to or outcomes of instances of STRESS, ATTITUDES, or EMOTION REGULATION. If a person confuses one faculty for the other, say, an instance of MEMORY for an instance of PERCEPTION, or an EMOTION instance for a physical symptom, the logic goes, then something must be seriously (or medically) wrong with that person. Papers that question the biological separation of faculties are less common and sometimes considered provocative (e.g., M. L. Anderson, 2014; Barrett & Satpute, 2013; Churchland, 1981; Hutchinson & Barrett, 2019).

The clearest modern examples of presumed natural kinds in psychological science are found in nativist theories of language by Chomsky and like-minded others (e.g., Chomsky, 1980), in the science of emotion (as discussed above, also see Scarantino & Griffiths, 2011), and in evolutionary psychology more generally (e.g., Cosmides & Tooby, 1992; Pinker, 1995; Tooby & Cosmides, 1992). These approaches assume that the human mind evolved a single, universal architecture of specific mental modules (which are natural kind categories), each one well-adapted to deal with a specific recurring challenge, giving humans survival and fitness advantages. By this argument, when the ancestors of modern humans developed this kind of mind with these sorts of categories, it must have given them competitive advantages, allowing them to pass on their genes more successfully and outcompete other species with different kinds of minds.

A large swath of psychological science ignores its ontological commitment to natural kinds even as the scientific products (peer-reviewed papers, lectures, textbooks, etc.) rely on and embody those assumptions. Scientific tools and practices for generating knowledge claims are not ontologically neutral (Barrett, 2006a, 2009, 2022; Danziger, 1988; Kuhn, 1962/1970) and contain clues to how reality is conceived. Here are a couple of clues to psychology's enduring preoccupation with the simplicity and sameness that are the hallmarks of natural kinds.

Reliance on psychological essentialism

The ontological commitment to natural kinds is clearly on display, whether acknowledged or not, when psychological scientists begin with a folk psychological category and assume that its instances are caused by a specific and distinct psychological process implemented by specific and distinct biological causes. For example, consider how some psychologists and neuroscientists use the conceptual framework for information processing devised by neuroscientist David Marr (1982/2010). The concepts describe three levels of information processing in engineered systems that have been applied to organic systems. There is a strict parallel across all three levels: folk categories constitute the phenomena of interest (e.g., instances of fear), the processes that cause those phenomena (e.g., "fear processes"), and the underlying biology that implements them (e.g., a "fear circuit"). These are Marr's computational, algorithmic, and implementational levels, respectively, which are similar to philosopher Daniel Dennett's three levels of abstraction when understanding, explaining or predicting behavior: the intentional, design and physical levels, respectively (Dennett, 1987). A word like "affect," for example, is assumed to refer to the instances of a category (e.g., feelings of pleasure or discomfort), the psychological process that is presumed to cause those instances (i.e., affective processes), and the biological features presumed to implement the process (e.g., genes, assemblies of neurons, etc.). This parallelism applies to "appraisal," "self," "control," "mentalizing," "goals," and most words that name psychological categories. In some versions of psychological essentialism, the implementation details, even though they are not known, are assumed to function as the category's essential features that explain the causal properties of its instances. Examples of psychological essentialism in practice include the many attempts to localize psychological phenomena to a specific neuron or assemblies of neurons (as a specific circuit, brain region, network, distributed pattern of brain activity, a.k.a. "biomarker") in a one-to-one mapping (for discussion, see Clark-Polner et al., 2017; Westlin et al., 2023). Other evidence of this essentialist commitment manifests in research programs that search for genes, chemicals, and physiological changes that are hypothesized to cause the instances of a particular psychological category uniquely and specifically and explain its causal impacts on other phenomena. Sometimes an essentialist commitment can be found in attempts to justify an inference of an unobservable (i.e., inferred) function that is shared by all category instances (e.g., Adolphs & Anderson, 2018; D. J. Anderson & Adolphs, 2014).

Use of generic language

The use of generic language in published scientific reports is another clue to lurking assumptions about natural kinds. Simple generalizations about categories (e.g., birds fly; fish swim; women are more emotional than men) have been empirically linked to essentialism (S. A. Gelman & Roberts, 2017; Rhodes et al., 2012; for a similar point regarding the use of nouns without context, see Barsalou et al., 2010). Generic language communicates broad conclusions about entire categories based on limited observations of small and often unrepresentative samples of their instances. For

example, a textual analysis found that 89% of over 1,000 psychology articles published in 11 well-respected journals during 2015-2016 used generic language in their titles, abstracts, or research highlights (DeJesus et al., 2019). The statements covered a wide range of categories, including people, women, children, emotions, decision-making, parental warmth, the brain, brain areas such as the amygdala, and so on. Additional studies by DeJesus et al. indicated that generic language was associated with readers' tendency to overgeneralize the reported findings.

The concept of an experiment

Further evidence of a philosophical commitment to natural kinds comes from the methods and practices scientists choose. Historically, the definition of an experiment has varied widely, even in the natural sciences (Danziger, 1990, 2000; Feyerabend, 1975/2010; Hacking, 1983; Oreskes, 2019). In psychological science, one standard design for a laboratory experiment is widely considered the gold standard in scientific practice. This experimental setup focuses on cultivating sameness in observations for the purposes of discovering simple, deterministic forms of causation. Its elements were originally borrowed from 19th century natural sciences and became the tool of choice in psychological science in the early 20th century due largely to pragmatic and political (i.e., non-epistemic) considerations rather than scientific ones (Danziger, 1990; for more details on its historical roots, see Danziger, 2000; Winston, 1990; Winston & Blais, 1996). A laboratory experiment is ideal for observing the sort of stability and sameness that should characterize natural kind categories, particularly when studied with the hypothetico-deductive method. In cartoon form, the standard experiment setup goes something like this:

1. Curate and observe category instances in participants in a manner that is objective. Category instances must exist independently of all human observers (i.e., scientists), consistent with the notion that a perspective-independent reality exists out there to be studied.

2. Minimize the impact of context. Limit sources of variation rather than cultivate and model variation.

3. Control all but a handful of variables, e.g., X, Y and Z. Variations in Y, called the dependent variable, should be caused by a force that is decomposable into a few basic elements, X and Z, called the independent variables, that can be experimentally isolated and separately manipulated (Danziger, 1990, 2000). Variables X and Z should have strong independent, additive, or simple linear interactive causal impacts that govern the outcome of interest, Y. Philosophers call this approach minimalist idealization (Weisberg, 2013) — isolate one or two causal forces, e.g., X and Z, that are assumed to be central to the category being studied. The causal relations between the variables X, Y, and Z should work like pieces of a machine, an idea cleverly called the machine model of causation (Lewontin, 2000) or more opaquely, linear determinism or efficient causation (Juarrero, 2023). The internal constitution of X and Z (as instances of specific psychological categories), their features, their causal powers — their nature — should not change because of their interaction.

4. Quantify and observe curated instances in a series of independent sequences of stimulus followed by response.

5. Treat contextual factors, which cause unexpected sources of variation in Y alone or via their influence on the causal potency of X and Z, as nuisance factors or "hidden moderators" (i.e., contextual factors that, unbeknownst to scientists, causally influence observations during a

replication attempt; also called hidden assumptions; Zwaan et al., 2018 as cited in Framework for Open and Reproducible Research Training; https://www.forrt.org/).

6. In the search for stability and sameness, sample enough participants to ensure reliable observations and inferences. Focus on averages to strengthen the variance that is reliable: average observations of Y across trials for each participant. Average the within-person averages across many, many participants so that stochasticity, hidden moderators, and other sources of unwanted variation will wash out.

7. Compare aggregated observations to the originating hypotheses that gave rise to the experiment by way of the hypothetico-deductive method given certain assumptions.

This experimental approach reveals just some of the ways that an ontological commitment to natural kind categories manifests in the practice of psychology, whether intended or not. It allows inferences about the "average" person — a summary of a population of people — an assumption that is in line with the search for stable, fixed relations to allow for inferences about inherent properties and simple causal relations. If all goes well, observations of category instances in a particular sample or samples of participants permit inferences to all other instances of the same category, and vice versa, regardless of time, individual, or context. The ultimate product is a stable, simple scientific law that provides justified knowledge about the category.

Creating the social psychology experiment: Festinger, Lewin and the Allport brothers

A particularly useful example of how to conform experimental setups to this one standard mold is evident in how the social psychologists Leon Festinger, Gordon Allport, and their like-minded colleagues "reinterpreted" the more complexity-based ideas of Kurt Lewin. As psychology was first transitioning from a discipline of mental philosophy into a bona-fide science, many psychology experiments examined individuals *embedded* in physical and social environments, such as dyads or groups (see Danziger, 2000; Greenwood, 2000). The social psychologist Floyd Allport (1919, 1924), Gordon's older brother, fundamentally changed the ontology of social psychology when he reduced the unit of analysis to an individual, independent from but interacting with separate and independent elements of the environment — creating a popular concept in experimental social psychology: the person x situation.

This concept reduced a group of people to the perspective of one individual, and the others served as "social stimuli" or "the social situation" (e.g., Dashiell, 1935). "Social behavior comprises the stimulations and reactions arising between an individual and the social portion of his environment," wrote Allport (1924, p. 3), followed a few pages later by, "The true locus of cause and effect … [is] the behavior mechanism of the individual" (1924, p. 9). Individuals were assumed to be social atoms reacting to the social stimuli around them. The consequence was that all psychological categories of scientific interest came to rest inside a person, separate from the contexts they inhabited. Biological factors that contributed to the complexity of human behavior, such as the time a person last ate or how much they slept, were ignored. Any influences that were not short term, local, and easily decomposable (i.e., not easily studied within the constraints of an experimental setup) were relegated to the scientific dustbin (Danziger, 2000). The special environment of the experiment became invisible, its assumed impact negligible.

Shortly thereafter, the social psychologist Kurt Lewin (1935, 1936, 1938, 1927/1992, 1949/1999) and others re-introduced the idea that social behavior emerged from complex social embeddedness and

could therefore not be ontologically reduced to more basic, individual parts (Danziger, 2000). Lewin conceived of an experiment holistically as a "field" or "total situation" (similar to William James's earlier notion of the "stimulus situation"; James, 1894/1994). The laboratory experiment was a means of constructing complex social realities and observing the emergent results through the conceptual lens of a theory. "For Lewin (1939/1952) instances of individual behavior had no intrinsic meaning," wrote historian Kurt Danziger, "Their significance could be assessed only in terms of the situation in which they occurred" (Danziger, 2000, p. 341). For this more holistic vision, Lewin drew on principles of Gestalt psychology and was inspired by the philosopher of science Ernst Cassirer who, among his many accomplishments, wrote philosophical works on relativity theory and quantum mechanics, crafting an early relational view of reality (Cassirer, 1921/1923, 1910/1923, 1936/1956) that is sometimes referred to as neo-Kantian. Specifically, Cassirer suggested that an experiment is a simplification of a much more complex reality (Cassirer, 1910/1923), what is known today as a *theoretical kind* (see Appendix, Section 5). Lewin, in effect, seems to have been practicing a philosophy of science called *relational realism*, discussed later in this chapter. This philosophy is articulated in some of his writings, although without the explicit label.

The complexity of Lewin's relational philosophy was no match for the simpler and more appealing mechanistic ideas of Floyd Allport and like-minded psychologists, particularly those who understood physics as the simple determinism of Newtonian mechanics. From their perspective, Lewin's experimental approach looked like a morass of uncontrolled variables (e.g., Festinger, 1953, p. 138). In Danziger's words, "Social psychological theory was simply unable to bear the weight that a Lewinian approach to experimentation would place on it" (Danziger, 2000, p. 341). The "experimentalism" of Floyd Allport was championed by his younger brother Gordon (G. W. Allport, 1935, 1954) and by Leon Festinger (e.g., 1953, 1980; for review, see Danziger, 2000; Greenwood, 2000).

In the end, Lewin's relational ideas were grist for the reductionist mill. In went social reality as complex phenomena and out came Allport's notion of behavior as the additive and interactive effects of an elemental "person" and one or two variables in the "situation." The laboratory experiment was solidified as the consummate condition for isolating specific causes (Festinger et al., 1950) "under 'pure' conditions" (Festinger, 1953, p. 139), an *ideal* rather than an *idealization* (i.e., a theoretical kind). (In fairness, though, some of Lewin's own wording suggests he may have veered in that direction at times, e.g., see Lewin, 1931). The work of Solomon Asch (1956, 1952/1962), who examined conceptual understanding as part of the web of complex influences on behavior, met a similar fate. The ensuing years continued to transform complexity into simple mechanistic interactions between separate, independent causes (see Allen, 1965; Danziger, 2000; Steiner, 1966) in the name of building a social psychology of natural kind categories that could be studied, and replicated, using the hypothetico-deductive method.

Construct validity

Psychology's commitment to natural kinds is also baked into the conceptual foundation of its scientific practices, even when scientists do not claim this commitment, and even when they explicitly claim otherwise. This includes every scientist who has ever measured anything in a study and then estimated the reliability and validity of their measurements using classical measurement theory (Novick, 1966). Classical measurement theory derives from the history of psychological testing and assessment and is standard pedagogy for psychological scientists (at least in North America and Europe). Most psychological scientists have used statistical methods rooted in this theory, even if they never explicitly learned it. It's summarized with the equation:

$X = T + E$

in which observations (the actual numbers generated during measurement, denoted as X) are partitioned into two parts: that which is consistent and repeatable (the "true score," denoted as T) and that which is random and not repeatable ("error," denoted as E). X might be the speed with which a participant presses a button after exposure to a stimulus (i.e., reaction time), the choice from an array of response alternatives (i.e., forced choice), a change in heart rate, a change in blood-oxygen-level-dependent (BOLD) signal, a number selected on a Likert-type scale to convey a subjective experience, and so on. The *reliability* of a measurement is the proportion of variance in X that is accounted for by the consistent and repeatable T.

The true score T corresponds to a hypothetical or *latent construct* (MacCorquodale & Meehl, 1948) — a category of scientific interest that is allegedly real but whose existence can never be directly experienced or measured. That is, T is the inferred essence of any experience a scientist cultivates for themselves in a psychological experiment, making it a prime example of psychological essentialism. A latent construct cannot be estimated merely by a mathematical combination of some set of measurements (i.e., it is not an intervening variable or abstractive concept). Instead, the latent construct T must be mathematically inferred from the observed statistical relations between measurements (X's), and it cannot be reduced to those relations. The latent construct contains what the philosopher Otto Reichenbach (1938) called "surplus meaning." Psychologist Paul Meehl called it a "fictitious substance or process or idea" (MacCorquodale & Meehl, 1948, p. 96).

The *construct validity* of X, which establishes the psychological essence T, rests on an explanatory inference: the unobservable construct *causes* the observed measurements. Accordingly, observed measurements —experiences of the construct as cultivated in the experiment — are assumed to correlate with one another. By this reasoning, if instances of some category of interest are captured using multiple measures (different X's), and if those measures covary substantially with one another, then it is reasonable to infer the existence of the latent construct (T) on the assumption that T is the common cause of those covariances. More formally, strong associations, which can be understood as shared information between the various instances of X, permits the inductive inference of an underlying, unobserved (or unobservable) hypothetical essence T.

Another interpretation of T, related to this essentialist interpretation, comes from *generalizability theory* (Cronbach et al., 1963, 1972). The goal of measurement remains minimizing error to reveal a fixed and stable true score, but here, T is conceptualized as the certain estimate of the characteristic in question, equivalent to the average score obtained for a given participant across all possible measurement events. This definition of T is philosophically pragmatic because the focus is on estimating with as much certainty as possible, which is how objective truth was defined by the pragmatist philosopher Charles Sanders Peirce (1877, 1878; See Appendix, Section 6). It also is similar to the concept of an expected observed score in item response theory (Raykov et al., 2019).

In all three psychometric approaches — classical measurement theory, generalizability theory, and item response theory — the statistical abstraction is considered real in nature, whereas the variation across measurement instances (be they items or trials) is considered epiphenomenal or outright error. This sort of typological thinking implies a belief in natural kind categories. Its contrast, called *population thinking*, assumes that variation across measurement events is real and the statistical abstraction is a simplified fiction. Population thinking is discussed in more detail in the third section of the chapter.

Traditional Scientific Realism

The idea that reality exists as typologies of natural kind categories is an ontological commitment that typically belongs to a broader philosophy known as *traditional realism*. Traditional realists are committed to a conventional view of reality: natural groupings of objects, processes, and events exist out there in the world separately from any perceiver. Justified knowledge represents, corresponds to, or coordinates with the true, objective reality of those natural kinds, and it is possible to build justified knowledge about those kinds, including their deep, unobservable parts, by making inferences from what is observable (for review, see Chakravartty, 2017; Hacking, 1983). Traditional *scientific* realism applies these philosophical commitments to the practice of science: a realism for the targets of science (i.e., an ontological commitment to the existence of natural kinds) and for the products of science (i.e., an epistemic commitment to build true or approximately true knowledge about natural kinds).

Traditional scientific realists often replace "truth" with coherence (Godfrey-Smith, 2021) — a search for observations that cohere or co-vary with one another across different measurement methods, contexts, and occasions. In psychology, it is standard practice to assume that if different instances of a category consistently cohere, then a scientist can be confident that the category exists in nature. Coherence is the basis of *convergent validity*, which is part of establishing the construct validity of a latent construct (Cronbach & Meehl, 1955; MacCorquodale & Meehl, 1948). An example of truth as coherence can be found in typological approaches to emotion. A natural kind of emotion category is characterized as having a unique, reliable suite of coordinated features (e.g., in peripheral physiology, motivation, facial expressions, action tendencies, etc.) that is supposed to serve as an evolved adaptation to a recurring fitness-relevant challenge (e.g., Shariff & Tracy, 2011).

Traditional scientific realism is the common-sense view of science. Physics is said to have discovered the reality of electrons, quarks, and so on. Chemistry has similarly discovered chemical elements, such as oxygen and hydrogen, and molecules such as water (H_20). Biology has supposedly provided the knowledge that each human has a body made of cells that are built with proteins synthesized using oxygen and other bits and bobs according to a blueprint provided by DNA molecules. Electrons, oxygen, DNA, and all the other material stuff that scientists have discovered are assumed to exist in objective reality, independent of any human perceiver.

Traditional scientific realism admits that scientific efforts are not concept-free. Human concepts are acknowledged as important tools of science, and any tool can be used well or poorly. Different people living in different places or at different times had different concepts for what is real and used those concepts to structure their inquiries into nature. But the assumption here is that good scientific practice eventually corrects any misguided concepts and persistently nudges scientists toward the one universal typology of natural categories that persist in space and time. Concepts may change, but the natural categories they describe do not. These categories are just plain facts, objective bits of reality. These facts of reality were true even when humans lacked concepts for them.

In principle, a scientist can be a realist about the categories of scientific study but not about the processes and products of science, or vice versa (for accessible discussions, see Godfrey-Smith, 2021; Hacking, 1983), but usually the two are so deeply entwined that one is used to justify the other. The discussion of natural kinds was, in effect, a treatment of ontological realism, so the rest of this section focuses on the epistemic commitments of traditional scientific realism.

Traditional epistemic realism

Generally speaking, traditional epistemic realism in science seeks "True Facts" about mind-independent reality, with capitals "T" and "F" (for discussions, see Godfrey-Smith, 2021; Hacking, 1983). A traditional epistemic realist aims the hypothetico-deductive method at scientific concepts that objectively describe, predict, and explain the natural kind categories of reality. In equation form,

natural kind categories + the hypothetico-deductive method = the possibility of discovering "facts" of reality

(or results that approach the facts of reality). It's helpful to consider each term in this equation.

Natural kinds are assumed to play a central role in the predictive and explanatory success of science. Their instances should have the sort of clarity and regularity that make them particularly well-suited for inductively deriving scientific laws, for making deductive inferences that support generalizations from those laws, and for making abductive inferences that provide explanations (e.g., Koslicki, 2008). Accordingly, it is possible to sample just a few instances of a category to test hypotheses and eventually yield "factually true" knowledge about the entire category and its role in the fundamental, deep structure of reality. Once a scientific law has been established, it becomes possible to deduce truths about future category instances and causally explain unexpected observations about any category instances, thereby resolving apparent anomalies.

The hypothetico-deductive method is assumed to be the most rational way to learn about the world, because the fate of ideas (i.e., hypotheses) as true or false (Chakravartty, 2007) depends on how they fare when exposed to what really exists in the world (the facts of scientific observations). This process of testing hypotheses against the facts of experience is assumed to work, more or less, independent of anything else a scientist thinks, feels, or believes. When used properly, the logic goes, the hypothetico-deductive method produces the robust, replicable observations that furnish facts about the world — facts that are experienced reliably and directly. In such cases, the method also permits inferences from those experiences about aspects of reality that are unobservable but are assumed to be factually true.

A traditional psychological realist assumes that there is one universal human nature consisting of natural kinds of psychological phenomena (i.e., psychological faculties) and aims to develop an objective picture of those categories and the scientific laws that govern them. Psychological scientists who use the hypothetico-deductive method to make inferences and create knowledge claims about entire categories after observing relatively small samples of category instances are probably practicing traditional scientific realism, automatically assuming that the categories under investigation are natural kinds. The same is true for scientists who expect that observations for any given sample of category instances should be repeatable, regardless of enormous variations in spatiotemporal context (i.e., variation across trials within an experiment, across participants in a given study, and across participants in different studies). Sure, isolated aspects of context might tweak the instances of a category, introducing a bit of variation, but by and large, contextual factors are considered moderating rather than fully causal (e.g., see the prior discussion of typological approaches to emotion). Concluding that a hypothesis is false (a la Popper) when similar experiments do not yield consistent observations given rigorous scientific practices further suggests the practice of traditional scientific realism. In fact, the widespread belief that psychological science

is suffering a "replicability crisis" or "credibility crisis" of grand proportions may in fact be evidence of widespread commitments to traditional scientific realism.

Justifications and critiques of traditional scientific realism

There are three common justifications in support of traditional scientific realism. These are presented below, along with common objections. Following these are four particularly knotty complications for traditional realism in psychological science.

The best scientific theories are empirically successful

If a theory works well enough, it therefore must be true. This is called "inference to the best explanation" and "the no miracles argument" (Boyd, 1999; Chakravartty, 2007; Putnam, 1975a; Smart, 1963). *Critiques*: (1) A theory can be relatively "less false" than other yet require years of research or methodological innovations before it can be empirically successful (see Appendix, Section 1, for more on justified knowledge or "truth" without empirical success). (2) A theory can be successful yet be false (see Appendix, Section 2, for more on empirical success without justified knowledge or "truth"); in fact, empirical successes always include some falsehoods by design (see Appendix, Section 5, on theoretical kinds). (3) All experimental tests are theory-laden, sometimes so much so that empirical success is unintentionally manufactured (see Appendix, Section 4, on the use of choice-from-array methods in the science of emotion). (4) What constitutes "working well enough"? (see Appendix, Section 10, on effect sizes in psychological science).

The most trustworthy knowledge is perceiver-independent

Scientific observation is limited by human senses. The most trustworthy knowledge of reality derives from sources that are independent of, or exist prior to, personal experience. This is rationalism. (1) Mathematics, also called "pure reason," is the thought by some rationalists to be the most reliable way to discover "truth." If a hypothesis can be successfully modeled, it therefore must be true. *Critiques*: Mathematics is, itself, concept-laden and not independent of nor a priori to human experience (e.g., Brown, 2017; Lakoff & Nuñes, 2001; Livio, 2009). Machine learning cannot surmount the concept-laden models and observations that train the system (e.g., Barrett, 2021a, 2022). (2) Some knowledge about reality is encoded in human genes and therefore part of innate, human nature. Any theory consistent with genetic evidence must therefore be true (e.g., Carruthers, 1992). *Critiques*: The concept of a "gene" is, itself, theory-laden and biologists do not agree on its definition (Ball, 2024). Even if the commonsense idea of a gene is taken at face value, humans have the kind of nature (genes) that require nurture (experience) to develop working knowledge about the world, so knowledge requires more than evidence about genes alone (e.g., Ball, 2024; Jablonka & Lamb, 2014; K. N. Laland et al., 2014, 2015; Lewontin, 2000; Müller, 2017; Noble & Noble, 2023; Richerson & Boyd, 2006). As a consequence, the human brain can produce more than one kind of human mind (e.g., Hoemann et al., 2024; Robbins & Rumsey, 2008; also see Appendix, Section 9). (3) Accomplished scientists of the past, such as Darwin, are vested with authority by virtue of their empirical successes. As a consequence, their opinions stand in for, or come to serve as, evidence that certain ideas are true. *Critique*: Truth by authority is in conflict with the period of time known as the Scientific Revolution. All scientific works by Darwin or otherwise are historical

and must not be read as modern scientific texts and understood in the context of their original historical context.

The hypothetico-deductive method is rational

The hypothetico-deductive method allows a scientist to test a hypothesis against the reality of lived experience and therefore is the most rational way to build justified and trustworthy knowledge. It is the best guide to objective truth. *Critiques*: (1) The hypothetico-deductive method is a normative account of science, not an actual description of how scientists practice science. Science is more of a social enterprise and social influences help to determine which knowledge claims are transformed into knowledge. Scientific facts are made, not discovered (e.g., Bloor, 1991; Collins, 1992; Hacking, 1983; D. L. Hull, 1988; Latour & Woolgar, 1979/1986; Laudan, 1977, 1990; Longino, 1990, 2001; Shapin & Schaffer, 1985/2011; for a subtler view, see Kuhn, 1997; also see Appendix, Sections 7 and 12); (2) Observations never determine the truth of a hypothesis, a scientific concept, or a larger theory of networked concepts (Duhem, 1954/1991; Quine, 1951b; also see Appendix, Section 11 on why knowledge is under-determined by evidence); (3) Scientists confuse inductive and deductive logic when they use the former to draw conclusions and make knowledge claims with the confidence that is reserved for the latter; this point is discussed below, in more detail, with reference to psychological science.

The Faustian bargain of words in psychological science

Words are powerful and efficient ways to communicate. The word "pizza," for example, allows people to communicate multiple features with just a few phonemes: A pizza is dough rolled flat and baked with tomato sauce and topped with cheese, maybe a few olives and a little oregano and basil. A word also invites humans to learn a new category by searching for features of equivalence. By the same process, a word can also invite psychological essentialism. A word can lead people to *invent* said features if they do not otherwise find them, *creating sameness amidst sensory and motor variation* while also *creating an explanation* for that sameness: some deep, unobservable, or even unknowable quality that should be responsible for the equivalence of category members, giving them their true identity (for a similar point, see Brick et al., 2022). This is the downside of a category's explanatory utility and, for scientists, a danger zone when using concepts to search for natural kind categories with the hypothetico-deductive method. Following classical measurement theory, abstractions are sometimes assumed to point to the deep, unobservable (i.e., latent) causes of things (Cronbach & Meehl, 1955). Words can be "essence-placeholders" (Xu, 2002) that tempt a person to invent an essence-like intension for the category, particularly if the words are nouns (Barsalou et al., 2010) used in generic sentences; together, this combination implies generalizations about categories (e.g., birds fly; fish swim; women are more emotional than men). William James made this observation more than a century ago, when he wrote, "Whenever we have made a word […] to denote a certain group of phenomena, we are prone to suppose a substantive entity existing beyond the phenomena, of which the word shall be the name" (James, 1890/1931, p. 195). Another name for this is *reification*. The words that help people learn concepts (by serving as features of equivalence) can also trick people into believing that the associated categories have essential features. A belief in essential features might not be a problem in day-to-day life (e.g., believing neuroticism causes anxiety), but the invitation to essentialism leads to problems when developing scientific knowledge, as discussed later.

An insidious problem arises when different communities of scientists use the same word to point to different categories that do not share other features of equivalence or extensions. When the meanings are different enough and lack features that can be compared, the result is incommensurability. For example, the engineering definition of a "state," which is used in dynamical systems theory, means something very different than the definition of a "state" in traditional psychology. Published papers in psychology routinely refer to all sorts of "states," such as "perceptual states," "attentional states," "cognitive states," and "emotional states," each of which is thought to be implemented in the vigorous firing of particular sets of neurons. In this typological usage, a brain can be simultaneously in many "states" at once, perhaps interacting with one another, and one state is said to mechanistically cause another. Modeling the brain as a dynamical system, in contrast, leads to the idea that there is one single brain state that can take on many values, and whose behavior over time corresponds to the fluctuations of physical quantities in the brain and traces out a trajectory in its state space. This approach has profoundly different implications for understanding what a mind is and how a brain creates it. Similarly, in the science of emotion, the word for an emotion category, such as "anger," is used so incommensurately (as a feeling, as an action, as a psychological process, as a brain circuit, as a specific instance of the category ANGER, as the entire category, etc.) that it's no wonder the field is rife with irreconcilable differences (for a broader discussion on this problem, see Sullivan, 2016).

An inconvenient truth at the heart of the hypothetico-deductive method

As discussed earlier, traditional epistemic realism has a hidden assumption discussed earlier: The hypothetico-deductive method works best when hypotheses and predictions about natural kind categories are initially deduced from scientific laws (i.e., deduced from a premise that is deterministic and true; e.g., Braithwaite, 1953; Gardiner, 1959; Nagel, 1961; K. R. Popper, 1959/2005). For this reason, the hypothetico-deductive method is also called the *deductive-nomological model*, meaning "deduction from laws" (Esfeld, 2009). Most psychological scientists encounter the word "nomological" in Cronbach & Meehl's (1955) famous paper on construct validity, which explains the importance of establishing a "nomological net" (or law-like network) for confidently creating inferences about the psychological meaning of latent constructs. Law-like universal regularities are the basis of scientific prediction and explanation within traditional scientific realism. If the premise of a hypothesis is indeterminate or weakly supported, such as a casual observation, a supposition, or an unexpected experimental result from a previous study, or if a premise has been shown to be false, then by the rules of logic a scientist cannot confidently reject their hypothesis in the face of unexpected (anomalous) observations; nor can they confidently claim that a deduced hypothesis is justified or approximately true, *even when that hypothesis matches the facts of experience* (i.e., is consistent with experimental observations).

This last point is of considerable importance. It implies that every psychology experiment has a huge, invisible asterisk on it. Even for a scientist who is comfortable making hidden inductive inferences within the hypothetico-deductive method, observations that are predicted by a hypothesis *cannot be considered evidence that the hypothesis is true* unless the category of interest is a natural kind governed by a scientific law. What does this mean for psychological science, which has no known scientific laws or starting conjectures that predict highly probable regularities? In place of laws, psychologists routinely derive hypotheses and predictions premised on conjectures from a version of evolutionary theory (the modern evolutionary synthesis) that many biologists consider to be outdated, from common-sense, folk beliefs, or from observations from prior studies that merely crossed some threshold of statistical significance. In such cases, as discussed earlier,

hypotheses are derived by inductive inferences, which offer considerably less certainty than deductions from scientific laws (Oreskes, 2003). The more uncertain the original premise, the weaker the resulting inductive inference, even when observations match hypotheses.

A traditional scientific realist might assume that the way out of this morass is for psychology to get its act together, improve its methodological rigor, and finally discover some psychological laws. This is the obvious solution when viewed from within the logic of the hypothetico-deductive model. But philosophers of science have pointed out a little problem with the logic of the hypothetico-deductive model itself: Scientific laws are always *simplifications* of reality. A scientific law on its own never works in the real world because it does not describe the complexity of the real world; nor does a law ever make clear predictions about what will be observed in specific situations (Cartwright, 1983; Giere, 2010). Contextual features and other auxiliary assumptions must be added to a law for it predict anything in the real world. A law strips away context and complexity in the service of generality and sameness. A scientific law is a fiction by design.

Consider, for example, Newton's universal law of gravity (Newton, 1687/2009): Every particle attracts every other particle with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between their centers. This law, like all scientific laws, is an idealized statement that approximates gravitational force in a variety of circumstances only because it *ignores important contextual details*, namely all other forces that affect objects (e.g., electrical forces). Moreover, Newton's law fails as an approximation when objects are microscopic, such as electrons and protons, or massively large and dense, such as stars. This kind of simplification is signified by the phrase *ceretis paribus*, meaning "all other things being equal" (Cartwright, 1983). A law describes universal regularities *only* when all other factors are ignored. In the real world, all things are never equal. Contextual influences always matter. In the philosophy of science, simplifications that remove contextual variation are called *theoretical kinds* (Giere, 2010; Teller, 2019). Every scientific law is a theoretical kind — an idealization that does not exist in nature (see Appendix, Section 5).

There are two important implications here. First, a scientific law, as a simplification, is a *template* that requires additional context to make precise descriptions, predictions, and explanations about the natural kind categories that are assumed to structure reality. These contextual details are not "hidden moderators" of a supposed universal truth. They are necessary to transform a simplification — a fiction — into something usable in the real world. Contextual details are introduced as auxiliary assumptions in the hypothetico-deductive method when any hypothesis is transformed into a scientific prediction. Predictions are always premised on a combination of (1) the contextual conditions in which observations will take place and (2) the relevant law applied to those conditions. This contextuality is built right into the hypothetico-deductive method, although it usually goes unstated, just as ambient humidity and altitude are always contextual ingredients in baking bread, even though they are rarely mentioned in a recipe. What does this contextuality imply about the reality of natural kind categories that, by definition, are supposed to be governed by lawful, universal relations of cause and effect? Natural kinds, like the laws that govern them, may be theoretical kinds — simplifications that do not exist in reality.

The second implication is related to the first. If scientific laws are theoretical kinds — idealized simplifications of reality — rather than universal truths in a boots-on-the-ground sort of way, then traditional scientific realism, particularly as it is practiced in psychological science, misunderstands the hypothetico-deductive method and its role in scientific practice. The method does not churn out "True Scientific Facts" in a mind-independent manner. There is no reason to assume that

observations will *de facto* replicate under their own steam across experiments in the absence of contextual details, even when hypotheses are derived from scientific laws, and even when predictions are tested with the highest degree of rigor. To yield predictable, reliable, and accurate observations, laws *always* require additional contextual details to coordinate with the phenomena in the world. Who decides which details to include and which to ignore? Scientists do. Using concepts. Scientists select specific details that particularize the law and make it progressively more concrete and complex. Empirical successes — the "facts of reality" delivered by the hypothetico-deductive method under optimal conditions — necessarily depend on those choices. Whether a scientist can confidently conclude anything about empirical successes and failures when using the hypothetico-deductive method depends on the concepts that a scientist starts with, which belong to what philosopher Willard Van Orman Quine called a *web of belief* (Quine, 1951b; Quine & Ullian, 1978).

In a very real sense, every hypothesis is an idealization. Before a hypothesis can be tested, it must be contextualized with various constraints and assumptions to create a prediction. In other words, no one ever tests a simple hypothesis (e.g., "Is the category FEAR a basic emotion?"). Instead, a *larger web of relations* surrounding the hypothesis is tested against the facts of experience (e.g., "Does FEAR give evidence of basicness in this particular experimental setup, when sampled in this particular way, with these particular subjects, using these particular stimuli and those particular measures," etc.). This is called the *Duhem-Quine thesis* (Duhem, 1954/1991; Quine, 1951b; Stanford, 2023). Observations emerge from the whole web, and therefore a hypothesis is always evaluated relationally in the context of this web, an idea known as *confirmation holism* (Quine, 1951b). A hypothesis, by itself, never has observable consequences that can be used to test it (see Appendix, Section 11).

By misunderstanding the philosophical preconditions that are necessary to make certain inferences, many psychologists continue to seek law-like, universal explanations about all humans, all mammals, all vertebrates, or what have you, for psychological categories that are assumed to be natural kinds. On the flip side of this coin, replication failures are interpreted as a crisis of credibility. If there is a crisis here, it may be partially created by a naïve realism that psychological categories have the kind of sameness and simplicity characteristic of natural kinds that are governed by scientific laws and are independent of context and human concepts.

Variation and complexity

The hypothetico-deductive method works best to build knowledge for categories whose extensions that share the same intension or intensions that are highly similar to one another, such as natural kind categories (with essences or as prototypes, respectively). It's very difficult — perhaps impossible — to use the hypothetico-deductive method to build justified knowledge about categories whose extensions are highly contextual and therefore highly variable across situations (Mayr, 2004). Here's why: The extensions of such categories have cause-and-effect relations that emerge from *complex ensembles of causal factors* rather than from simple, mechanistic sequences like those that are captured in a scientific law. In reality, no extensions of the category will necessarily be certain (deterministic) or even highly probable. A prototype is more like a stereotype than a representative or typical instance. Category instances can be infrequent or even downright unique yet still be extensions of the category. This is what it means for a category to have an abundance of contextual variation.

When categories have highly variable extensions, the hypothetico-deductive method often produces anomalous-looking results, including what appear to be failures to reproduce and replicate observations (Van Bavel et al., 2016). Two experimental setups might look the same on the surface, but if they contain subtle differences in one or two weak causal factors, factors that are necessary but ignored by scientists, then the resulting experiences (i.e., observations) will be different. What looks like a failure to replicate may be a discovery that the category of interest has a different structure than first assumed.

An apt example comes from the science of molecular genetics (described in Lewontin, 2000). A standard method for demonstrating that a gene is the source of a phenotypic characteristic, such as wings on a fly, is to identify mutations that disrupt normal development. Many studies replicated the observation that flies with a particular genetic mutation developed curly wings rather than typical straight wings. Inside the lab, this result is robust. Outside the laboratory, however, most flies with the same genetic mutation develop straight wings. Why? The mutation produces curly wings *only* in a lab where causal factors, such as temperature and humidity, are hidden in the "background context," carefully controlled, whereas outside the lab, temperature and humidity vary across a broad range of environments. The mutation is a necessary but not sufficient causal element. Genes are turned on and off by signals from the environment in an ongoing, complex dynamic, so that alone, a genotype is usually insufficient as a blueprint of any phenotypic characteristic (Lewontin, 2000; for similar examples in rodent studies, see Reardon, 2016).

This sort of contextuality is endemic to psychology experiments. A growing number of studies give evidence of the profound and surprising ways in which so-called contextual factors play a causal role within experiments. A smattering of examples includes time of day (Hahn et al., 2012; May et al., 1993), ambient odors (Leleu et al., 2015; Sorge et al., 2014), CO_2 concentration in the air (Scully et al., 2019), room temperature (C. A. Anderson, 1989), participants' quality of sleep (Prather et al., 2013), and the race or gender of the experimenter (Sattler, 1970). A routinely overlooked aspect of context in most psychology experiments involves the signals associated with sensory conditions of a participant's body. For example, heart rate, respiration rate, and other signals informing the brain on the energetic state of the body (called interoceptive signals) routinely interact with the brain's processing of signals from the world (exteroceptive signals), starting where the signals are transduced in the sensory periphery (see these papers and references therein Barrett, 2022; Engelen et al., 2023). In fact, respiration and olfactory signals may play a role in the most basic aspects of memory (see discussion and references in Buzsáki & Tingley, 2023; Shaffer et al., 2023). And it is well established that contextual factors influence even the most basic aspects of movement, perception, and object perception (e.g., Bar, 2004; Brandman & Peelen, 2017; Castelhano & Pereira, 2018). Even whether a participant saccades to a stimulus versus passively views a stimulus is associated with different signal patterns in primary visual cortex (e.g., MacEvoy et al., 2008; Zirnsak & Moore, 2014). Findings like these suggest that psychological categories are strongly contextually variable (see also Richard et al., 2003). Variation may be a routine part of the phenomena psychologists are trying to know rather than an obstacle to be removed on the road to knowledge, meaning situated variation should be cultivated and modeled for its structure rather than controlled away (Barrett, 2022; A. Gelman, 2015; Westlin et al., 2023). This suggestion has been made numerous times (e.g., Cronbach, 1975; Gergen, 1978; Leary, 1987; Lewin, 1935, 1936, 1938, 1927/1992, 1949/1999; McGuire, 1973). In the words of psychologist Jay Van Bavel and colleagues, psychological categories are "orderly in their complexity rather than lawful in their simplicity" (Van Bavel et al., 2016, p. 6458).

The upshot is that unexpected variation in psychological observations is never grounds for confidently rejecting the associated psychological hypothesis (because the hypothesis is never premised on something as deterministic as a scientific law, and even if it were, the law would be a fiction requiring the addition of contextual details to make it useful). Nor is unexpected variation in psychological observations necessarily grounds for suspecting that study was executed poorly or that the scientists are untrustworthy. Unexpected variation in observations may in fact be a sign that the category of interest is more contextually situated in reality than previously assumed, that instances emerge from complex causation rather than simple, mechanistic causation, and therefore a different sort of experimental setup and scientific method is called for.

Objectivity and consensus in traditional scientific realism

In psychological science, as in everyday life, humans routinely observe the actions of humans and other animals and make inferences about unobservables, such as the animals' internal states (for a review, see Gilbert, 1998). Such inferences, from observed to unobservable, are a routine part of the hypothetico-deductive method and depend on scientists' concepts. How, then, can science ever be objective? The answer, for some psychological scientists, is to pursue consensus and call it objectivity.

In the science of emotion, for example, Adolphs and Anderson (Adolphs & Anderson, 2018; D. J. Anderson & Adolphs, 2014) assert that FEAR, ANGER, and other emotion categories each constitute a distinct functional state in an animal's brain. This is one version of the philosophical practice called *functionalism*: instances of the category FEAR are thought to derive their unique nature from the purpose the fear state achieves, not the internal constitution of each instance's physical details. Animal species differ in their ecologies, their nervous systems, and their development. Accordingly, the inputs to each functional state (the triggers), the circuitry that creates each functional state, and their outputs (the actions that result) will differ. When faced with a species-relevant threat, a fly (faced with a flyswatter) might hop, a rat (faced with a cat) might freeze, and a human (faced with uncertain social evaluation) might widen their eyes and sweat or close their eyes with a racing heart. Adolphs and Anderson assert that these actions, which differ in their physical details, are caused by a universal functional state of fear that is common to flies, rats, humans and other animals, and moreover, that this state can be observed *objectively* across different species. A functional state is always *inferred* — it cannot be directly observed. A function is an abstract mental feature, created in each scientist's brain, and serves as the intention of the scientist's concept. How is this objective?

Adolphs and Anderson's answer is scientific consensus. They publish detailed criteria for making careful measurements so any other scientist can repeat the observations. They also publish criteria for when to impose a function on what they observe; these are basically rules for inferring what cannot be directly experienced (i.e., they transparently specify their concepts and invite other scientists to use those concepts under certain conditions). This transparency allows scientists, should they so choose, to reliably make the same abductive inferences as Adolphs and Anderson. The similarity between hopping flies, freezing rats and gasping humans lies not in the movements observed, but in the explanatory inferences that scientists make. The rules for inference, if followed, should produce scientific consensus. From there, scientists are free to claim — if they are traditional scientific realists — that the neural circuitry supporting an animal's actions in threat is the neural circuitry for FEAR in that animal (for discussion, see Adolphs et al., 2019; Barrett, 2012).

To be sure, it's good scientific practice to define clear criteria for repeatable observation and explicate concepts for inferential transparency. Other scientists could take a lesson from the rigor of Adolphs and Anderson in this regard. Nevertheless, acculturating other scientists to one set of concepts and persuading them to use those concepts is not objectivity. Observational and inferential consensus, as derived via the consistent application of concepts, reflects the reality of *how humans, including scientists, create meaning by social consensus*, called social reality (Barrett, 2009, 2012; Berger & Luckmann, 1966/2011; Searle, 1995). That the application is transparent and deliberate does not transform consensus into a mind-independent, objective reality that includes FEAR as a species-general natural kind category.

The history, sociology, and anthropology of scientific practice suggests that scientific knowledge is less a body of objective facts about a mind-independent reality than a set of knowledge claims invested with confidence and authority via social activities in a scientific community (for references, see Appendix, Section 12; for reviews, see Godfrey-Smith, 2021, Ch. 7; Longino, 2019; Oreskes, 2019, Ch. 1). Scientific consensus is much more powerful (and more justified) when it emerges from *a diverse and self-critical scientific community of experts* (Oreskes, 2019), rather than a group of like-minded scientists. A more diverse community will share some ontological commitments but not others and *de facto* bring more perspectives to bear. Diversity also allows scientists to make discoveries that are not available to others whose lives are situated differently (Harding, 2015; Longino, 1990). Science is more likely to be self-correcting when scientists with diverse views challenge each other through the mechanisms and avenues of professional practices, or what philosopher Helen Longino calls *transformative interrogation* — the collective give-and-take of critical examination (Longino, 1990). Objectivity, in this case, is not truth per se, but something closer to "critically achieved consensus of the scientific community" (Oreskes, 2019, p. 55; see Appendix, Section 7). Put another way, objectivity in science is relational.

Summary

Traditional scientific realism, with its attendant assumption of natural kind categories, appears to be widespread in psychological science. It motivates the scientific search for a single, universal human nature, which is assumed to consist of objectively true, perspective-independent categories that can be discovered by using the hypothetico-deductive method to test scientific concepts against scientific observations (usually in laboratory experiments). In other words, many hypotheses in psychological science, and the experiments designed to test them, assume the existence of natural kinds *at the outset*, putting boundaries on what can ever be discovered. Some psychologists deny this ontological commitment to traditional scientific realism even as they behave in accordance with its assumptions. Here is a partial list of the behaviors that indicate traditional scientific realism is in play:

1. Ask questions without reference to a context or perspective.

2. Infer the existence of latent constructs.

3. Use a standard experimental setup the removes or holds constant most contextual factors to control rather than observe and model variance. Manipulate only one or two variables.

4. Estimate the reliability and construct validity of measurements using classical measurement theory.

5. Model simple (usually linear) relations between independent and dependent variables.

6. Favor experimental replication (i.e., internal validity) over generalization to the outside world (i.e., external validity).

7. Interpret findings without reference to the conditions of the experiment and all the contextual factors controlled.

8. Use generic language to report results.

Traditional scientific realism is intuitive, but it carries a host of assumptions and disclaimers that were discussed in this section. Notably, the hypothetico-deductive method works best to discover categories whose instances have the degree of sameness and simple law-like determinism that characterize natural kinds. Not surprisingly, psychological experiments are also designed to discover and build knowledge about natural kinds. And this is where problems arise. Since psychological hypotheses cannot be premised on anything with the high degree of certainty as a scientific law, psychological scientists must substitute premises of which they are less certain, ideas they feel strongly about or have confidence in, such as beliefs about evolution or consensus views. They mistakenly continue to expect the sort of predictable experimental outcomes (i.e., robust replications) that only a law for a natural kind category can provide (only after providing reliable contextual details). According to the traditional logic of the hypothetico-deductive method, any test of a hypothesis will be inconclusive without the law-like certainty (or even just strong predictability) in its premise because the hypothesis or its premise could be false.

Psychologists are left with a gulf between the requirements of the hypothetico-deductive method on the one hand (i.e., a law-like premise and natural kind categories — both of which assume a degree of sameness and simplicity of causation) and the complexity and variation of psychological phenomena on the other. Concerns about failures to repeat or replicate observations across studies — failures to observe sameness and simplicity — rallied efforts to improve methodological rigor but failed to question the suitability of this scientific method given its ontological presumptions. Improving methodological rigor is always a good thing, but even if psychological methods were perfect, experimental results would remain inconclusive. Even after dealing with poorly run experiments, p-hacking, and other undesirable behaviors, scientists will still be left in doubt as to whether they disproved a hypothesis or made incorrect auxiliary assumptions (i.e., a disconfirmation dilemma; Greenwald & Ronis, 1981). All things being equal, the strongest evidence for psychology's dependence on traditional realism may be the field's failure to examine its unwavering dedication to the hypothetico-deductive method and its assumptions of natural kinds.

With the additional realization that a scientific law is a theoretical kind — a mirage of generality and sameness — comes further complication: Experimental results will remain inconclusive even if the hypothetic-deductive method is used as intended, even if psychological science could furnish premises with law-like certainties. A law is nothing more than a simplification. It always requires auxiliary assumptions and contextual details to be useful in the real world. When it comes to building robust, justified knowledge, the scientific devil is in those details.

Empirical Kind Categories

Traditionally, when scientists and philosophers assume that the instances of a category share features that make them equivalent in a way that does not change very much across different

situations and times (e.g., instances of FEAR are interchangeable regardless of situation or the animal in which fear occurs), they understand this sameness and stability in one of two ways. Either the *instances* share something in common, like an enduring essence or a prototype, in which case the instances belong to a natural kind category, or the *human perceivers* share something in common, such as the same sensory systems or the same concepts, allowing the detection or inference of similarity amongst those instances. This latter assumption — that the similarity among a class of instances is *perceiver-dependent* — belongs to a group of philosophical traditions collectively called *anti-realism*. Anti-realism sets aside or outright rejects the existence of a mind-independent reality. The most recognizable flavor of philosophical antirealism is *idealism*, the idea that there is no reality beyond what can be conjured by human mental activity. Lots of ink has been spilled on the debate between realism and idealism, much of which is important to the history of philosophy, but idealism as a philosophy is not that useful to modern scientists. There is another anti-realist philosophy that psychological scientists are almost certainly familiar with, called *empiricism*. Empiricism involves more than simply using methods that involve observation to test hypotheses. As a philosophical tradition, empiricism is the stance that only sensory experiences can reveal true, justified knowledge about the nature of reality. In psychological science, empiricism provided the philosophical bedrock for *behaviorism*. Most students of psychology encounter behaviorism in an introductory textbook and learn about it as a historical blip that collapsed under the weight of its own limitations. Nonetheless, it's useful to briefly examine the philosophical foundations of behaviorism for their similarities and differences when compared to traditional scientific realism. Amidst important differences, both scientific approaches emphasize a search for categories with timeless and unchanging regularities as the way to understand nature. This focus on sameness and simplicity provides additional insight into the ontological assumptions of modern psychological science.

Empirical Kinds As Regularities In Human Sensory Experience

Behaviorism came in several flavors, all of which employed the concept of an *empirical kind* category. An empirical kind is like a natural kind — its features of equivalence (the category's intension) are assumed to exist in the natural world and remain stable and unchanging across time and context (i.e., the features of equivalence are immutable and universal). The *one critical difference*: An empirical kind's features must be concrete and directly observable by human senses; they cannot be abstract. Latent constructs, functional states, abstract mental features, and the like are considered out of scope in empiricism because they are unobservable and inferred and therefore epiphenomenal to science. No psychological essentialism is allowed. The category WANTING, to an empiricist, includes those instances when an animal expends physical energy to secure an outcome. For example, wanting chocolate would be defined as spending time, money, or effort to obtain it. Some scientists still define WANTING in this manner today (Berridge & Robinson, 2016). SEEING equals eye movements called saccades (e.g., moving the fovea of each eye to the position in space where the chocolate is sitting). Instances in which observable behavior changes following exposure to a stimulus (i.e., classical conditioning) or preceding an outcome (i.e., instrumental conditioning) belongs to the category LEARNING. Empirical kind categories, as groupings of instances that are physically similar across time and space, are assumed to describe and organize the part of reality that is accessible as patterns of human sensation and therefore the only part of reality that can be known. Each empirical kind is literally its measurement. Economists still reason about "utility" in this manner today (e.g., as "revealed preference"; Samuelson, 1938, 1948)

In the science of emotion, for example, a traditional scientific realist stipulates that instances of the category FEAR all share an "internal state" that functions to protect an animal from impending threat, measures fear in a variety of ways in different situations and animals (e.g., as the absence of movement or "freezing," widening eyes and gasping, etc.), and engages in scientific activities that presume FEAR is a latent (unobservable) state with an ensemble of varying physical features. That scientist expects an animal to freeze when exposed to a potential threat, and when they observe the animal doing so, they feel entitled to use their observations to make inferences about a latent state of fear — even the bits they did not or cannot observe. In contrast, when an empiricist expects a fearful animal to freeze, they are stipulating that fear literally *is* freezing. There is no surplus meaning. A sentence such as, "The rat is fearful" literally means "the rat is freezing" because observing the rat freeze verifies the sentence as true (an idea in empiricism called *verificationism*; Ayer, 1936/1952; also see Godfrey-Smith, 2021). And statements like "these neurons are the circuit for fear" or "a fear circuit is innate" literally mean (and only mean) "these neurons make a circuit that controls freezing" or "the circuit that controls freezing is innate." Such statements cannot be generalized to any other physical manifestation of fear, which would be considered a different category, hence the need for multiple narrow categories of FEAR and additional assumptions to organize them into typologies of FEAR categories (for examples of typologies, see Fanselow, 2018; Gross & Canteras, 2012).

Perception Vs. Conception

In the hypothetical feature space, empirical kinds would inhabit only the portion defined by concrete feature dimensions. The features of equivalence that ground these categories necessarily depend on human sensory systems, which is why empirical kinds are considered perceiver dependent. The assumption is not that an empirical kind category is brought into existence by human perception, but that sensation and perception (a.k.a. observations) are the only reliable and trustworthy basis to build knowledge about the objects and processes of reality. Most scientists who use the concept of empirical kinds in their scientific efforts assume that perceptions of similarity and distinctiveness track real similarities and distinctions in nature because human perceptual systems evolved through natural selection. By applying a scientific method correctly, the logic goes, scientific concepts should increasingly come to reflect natural groupings in the world via the regularities of human perceptions and actions. But even in such cases, the parts of reality that humans cannot contact, which are outside what humans can sense and act on, are out of bounds, strictly speaking, and therefore scientists should resist the conceit that they can build scientific knowledge about those parts.

Empirical kinds are perceiver-dependent, but like natural kinds, they are *concept-independent* (notice the assumption here that sensation and perception are process-pure and proceed independently of concepts). The category ELECTRON is assumed to exist regardless of human knowledge of electrons. FEAR is assumed to exist regardless of any human concept of fear. Empirical kinds, like natural kinds, are also assumed to be immutable and universal, not because the categories have essences or property clusters, but because all neurotypical humans have the same sensory and motor systems; so neurotypical humans should agree, more or less, on the category structure in the portion of reality that they can sense and act on. Notice the nod, here, to social consensus that is standing in for objectivity (see Appendix, Section 7). Concepts might differ from (neurotypical) brain to brain, person to person, but such variation is thought to be unlikely in sensations and perceptions.

Ontological Vs. Epistemic Reduction

A scientist could have different reasons for aiming their scientific efforts at empirical kinds rather than at natural kinds. A scientist may be pretty confident that reality is structured as natural kinds (traditional ontological realism) but assume that scientific practices can most reliably build concepts for the parts of reality that can be experienced directly through the senses (epistemic empiricism). A scientist may consider private mental events irrelevant to the practice of science because those events cannot not be verified by observation (similar to a modern form of empiricism called constructive empiricism, see van Fraassen, 1980). Certain varieties of behaviorism were outright suspicious of the existence of mental phenomena, and mental categories with abstract mental features of equivalence were thought to result from a fancy but unscientific sort of introspection. In such cases, redefining something mental as something physical meant defining the mental out of existence, a full-fledged *ontological reduction*.

Consider the behaviorist psychologist B. F. Skinner. At times, Skinner allowed for the existence of private mental events. At other times he likened psychological "processes" to fictions of the highest order. He referred to cognitive science and its search for mental processes (i.e., the categories of mental causation) as "the creationism of psychology," akin to religious efforts to "establish the existence of and the nature of a different creator, in this case, spelled with a capitol 'C'…" (Skinner, 1990). In effect, Skinner was accusing psychological scientists of making stuff up, in the same manner that an atheist would accuse a believer in intelligent design of inventing things they cannot directly observe to bring the world in line with their concepts or expectations. His solution to this problem reduced mental categories to physical categories that derived from combinations of variation and selection (natural selection, operant conditioning, and cultural evolution).

It's worth calling out an important point here. Ontological reductions are stipulations about how the world is. They are conventions whose truth necessarily depends on the consensus of scientists. They cannot be tested against evidence (Barrett, 2012). There is consensus among physicists, for example, that the presence of an electron is indicated by a trail of ionized gas particles in a cloud chamber. There is consensus among psychological scientists who study "fear learning" that an animal is in a state of fear when it freezes after exposure to a neutral tone that preceded a painful stimulation in previous trials. These definitions are *not* operational. An operational definition allows a scientist to observe and then make statements about the latent, theoretical category that they infer but cannot directly experience (Bridgman, 1927). Traditional scientific realists operationalize when they believe they are building knowledge about natural kind categories. For empiricists, the same maneuver is ontological reduction. For a scientist who believes they are studying empirical kinds, latent categories either do not exist or are beside the point.

Describing Vs. Explaining

Scientists who aim their scientific efforts at empirical kinds usually argue that *the goal of science is to describe and predict experienced regularities*, not to explain them. In principle, future instances of an empirical kind should be predictable from experiences with prior instances, as long as all instances share concrete sensory and motor features of equivalence. Empirical kinds should likewise support the inductive discoveries of scientific laws. What a scientist must never do is invent fictional features (i.e., infer abstract features), such as psychological essences or probabilistic feature clusters, to explain why these concrete features hang together. Empirical kinds can never reveal the hidden causal structure of reality. The intension of a natural kind category refers to the deep features that

make instances similar enough to be treated as equivalent. The intension of an empirical kind category, by contrast, is purely descriptive; the deep features of causation and explanation are considered irrelevant. So, the Faustian bargain of words must be strictly avoided. Words should never be considered invitations to search for functional, abstract features that create sameness amidst sensory and motor variation. Sameness in sensory and motor features within all instances of an empirical kind are the rules of the scientific game.

Reduced Inferential Potential

Empirical kinds create the same dilemma as natural kinds when it comes to scientific laws: A law is a theoretical kind — a simplification that does not exist in reality — and the price of generality is reduced deductive potential. The additional contextual details required to make a law usable in particular situations are auxiliary assumptions that could be wrong, weaking the power of the initial deduction.

Empirical kinds have reduced inductive potential when compared to natural kinds precisely because they only describe observable regularities. Observations for one sample of an empirical kind support inferences to other instances that share the exact physical features of equivalence. If the category FEAR, for example, is defined as a cessation of movement (i.e., "freezing"), then the sentences "The rat is fearful" and "The person is fearful" mean precisely and nothing more than the rat and the person are not moving. Any accumulated knowledge about the environmental contingencies that control freezing behavior (i.e., any knowledge about FEAR as freezing) cannot be used to infer anything about any other possible manifestations of fear, such as when a human cries in fear, when a rat attacks a predator in fear, when a fly rubs its legs in fear, etc. If the category FEAR is an empirical kind, then freezing is its intension, and inductive inferences only apply to extensions of the category with that intension (i.e., that have the same concrete, physical features).

Empiricism As A Philosophical Tradition

Empiricism is mostly a set of normative, epistemic beliefs about the right way to build true, trustworthy facts about the world. In modern philosophy of science, empiricism first took root in a philosophical mindset known as *positivism*, which then morphed into *logical positivism*, which in turn became *logical empiricism*. The differences among these different flavors of empiricism are less important for this chapter than their similarities, whose gist is that justified knowledge adequately predicts only those aspects of the world that are directly observable. A virtual melting pot of empiricist ideas kept philosophers doing what they do best — arguing with one another in service to those who practice philosophy in their scientific efforts. In the end, however, most versions of empiricism in philosophy collapsed under the weight of their own disagreements (Godfrey-Smith, 2021). Plus, they could not overcome mounting scientific evidence that observations are never concept-free. Most (if not all) aspects of sensation and perception depend on memory and other knowledge that scientists collectively refer to as "conceptual" (although for an attempt to refute this evidence, see Firestone & Scholl, 2016). Empiricists nonetheless tried to maintain their assumptions via a patchwork of philosophical maneuvers; for example, they argued that it's OK to use words to refer to unobservables if those words make it easier to detect patterns of experience that are, in principle, observable (see Hempel, 1958). Ultimately, empiricism in philosophy today is considerably diminished relative to its heyday in the 20th century. Behaviorism in psychology went the way of empiricism for various reasons, including, in true Kuhnian style, the resurgence of traditional

scientific realism during the cognitive revolution, which produced the interdisciplinary field of *cognitive science*. Nonetheless, a few points are worth considering.

Where traditional realists aim to discover knowledge about reality writ large, including its deep, latent structure, empiricists aim their efforts at reality only *as they experience it*. They presume that human experience is the only valid source of factual knowledge about the world and eschew inference to anything unobservable, such as theoretical constructs, latent states, or any deep structure to reality. For a traditional realist, a tree crashes to the ground with a loud thud whether or not a listener is present to hear it. Sensing and inferring are either not distinguished or one is thought to justify the other. For an empiricist, however, the question "When a tree falls in the forest and no one is there to hear it, does it make a sound?" is not meaningful. An idea can only be judged true or empirically adequate (which stands in for truth; van Fraassen, 1980) when it is consistent with something that is directly sensed (i.e., directly observed). All other hypotheticals are off-limits. For empiricists, "real" means whatever people experience through their senses, and nothing more.

Empiricism's anti-realist flavor and its emphasis on inductive inference is reflected in its complicated history with rationalism, the epistemic tradition of gaining knowledge from sources that are independent of one's own experiences, such as authority, superstition, religious theology, and political ideology, which tends to be favored by traditional scientific realists and those who favor deductive inference. In the philosophy of science, the dialog between empiricists and rationalists plays out in several ways. For example, rationalist arguments are sometimes used by traditional scientific realists to justify their inferences when using the hypothetico-deductive method, particularly when no law is available to serve as a premise for a hypothesis. Realists accredit mathematical proofs, nativist claims about natural selection, or the authority of a trusted scientist, like Darwin, to validate their inferences. A scientific empiricist actively avoids such maneuvers, denying the scientific value of anything that cannot be directly experienced. (The experience of reading Darwin's claims about natural selection is not admissible as evidence, for example; nor is the experience of solving a mathematical proof.)

The inductive-statistical model

Instead of using the hypothetico-deductive method to search for the categories that structure reality and build justified knowledge about them, behaviorists, like all empiricists, focused only on the regularities in human perception, which they studied with a weaker version of the hypothetico-deductive method called the *inductive-statistical method* (Hempel, 1942, 1965; Hempel & Oppenheim, 1948). As the name implies, empiricist hypotheses began as inductive inference from less certain premises rather than as deductive inferences from scientific laws. Knowledge claims developed by way of inductive inferences that accumulate over time with the help of statistics and probability estimates. Empiricists believe that the inductive-statistical method is a rational way to build knowledge because a hypothesis's meaning can be directly verified by personal observation, thereby allowing knowledge to accumulate by induction. By contrast, traditional realists like Karl Popper, to remind, argue strenuously against the use of inductive logic and instead believe that the rationality of the hypothetico-deductive method is firmly rooted in deduction from scientific laws.

Traditional realists also usually assume that scientific laws are true with a capital T (rather than being theoretical kinds which are simplifications and false by design) and therefore feel comfortable using laws to infer and explain unobservable features of reality. The explanatory power of the inductive-statistical method, by contrast, is effectively neutered. Fear cannot be said to *cause* freezing because fear *is* freezing. What, then, causes specific instances of FEAR? The empiricist's

answer: Wrong question. Behaviorists bent the assumptions of empiricism, however, moving beyond mere description of perceptual regularities (what most empiricists concerned themselves with) to offer explanatory inferences for those regularities (typically eschewed by most empiricists; e.g., Hempel, 1958; Schlick, 1918/1985). They inferred a few species-general, universal mechanisms with law-like regularities, all of which had to do with natural selection (a rationalist maneuver) or associative learning from experience (e.g., classical conditioning, selection by the situation (a.k.a. instrumental learning), or cultural inheritance). There were no category-specific, evolved adaptations here, however. A genetically inherited circuit to freeze in fear was just that and nothing more. Generic "evolved adaptations" to "deal with an impending threat" (i.e., fear), "communicate dominance and remove obstacles" (i.e., anger), or "avoid aversive foods" (i.e., disgust) were unobservable inferences — examples of psychological essentialism — and considered out of scope, epiphenomenal to science.

Instrumentalism vs. operationalism

At first blush, the empiricists' avoidance of unobservables seems to present a conundrum for scientists. Much of science concerns itself with parts of reality that go well beyond what scientists can observe directly. Committed empiricists had an answer to this problem: Words in science that seem to refer to unobservables are not hypotheses to be tested against observation, but tools for predicting what can be observed. This is called *instrumentalism* — the view that abstract concepts are useful tools for accurately identifying, systematizing, and describing observable parts of the world (Dewey, 1900, 1903; see Appendix, Section 6; also see Feest, 2010). Concepts are not, however, theories for discovering the hidden structure of nature or exploring, predicting, or confirming unobservable aspects of reality. For example, when logical empiricism is working well, scientists' concept for FEAR can help build knowledge about its observable aspects. A traditional scientific realist would understand this maneuver as "operationalizing" a latent construct: translating words for concepts corresponding to theoretical, abstract categories into observational language that is closely tied to the measurements and workings of a specific experiment (Bridgman, 1927). A logical empiricist would view operationalism as an error, however. Instrumentalizing FEAR as freezing behavior is not an operational definition. As an empiricist maneuver, instrumentalism is an ontological reduction.

Action tendencies and other unobservables

Logical empiricists figured out another angle for dealing with the conundrum of unobservables. They made surreptitious inferences about abstract, mental features while maintaining plausible deniability. They simply treated a movement as equivalent to, or interchangeable with, a hypothesized internal representation, such as a drive (C. L. Hull, 1931; Seward, 1956), a map (Tolman, 1948), or a disposition to act (Bull, 1945; Gray, 1935; Young, 1943). One example — dispositions to act — is explored here to highlight a few basic points.

A disposition is a state of readiness to perform a specific action. It cannot be directly observed and must be inferred, or to channel Skinner, invented. In 1895, for example, the philosopher John Dewey defined various categories of emotion as dispositions to act in specific ways (Dewey, 1895; also see Spencer, 1894; Titchener, 1896; for a detailed discussion, see Gendron & Barrett, 2009). In his view, an object instinctually causes a readiness to act in a particular way that is adapted to a particular situation, and this readiness sits at the core of instance of emotion. The notion of "action tendency" or "action readiness" was maintained by emotion researchers in the first half of the 20th

century who were committed to empirical kinds; they ontologically reduced a specific emotion category, like anger, to a specific disposition to make a specific action (e.g., "to strike out;" e.g., Bull, 1945; Gray, 1935). This meaning was later transformed from the likelihood of making a specific action into an even more abstract motivation that could be satisfied by various actions, such as "an impulse to aggress" in anger (e.g., Arnold, 1960a, 1960b; Frijda, 1986; Young, 1943). This seemingly insignificant definitional change appreciably altered the attendant philosophy of science. An empirical kind was transformed into a natural kind in all but name, making it scientifically acceptable to generalize from different observed actions to a single unobservable state (i.e., an inferred, abstract mental feature).

Hopefully, the problem is obvious. When scientists aim their efforts at what they assume are empirical kinds but then make inferences as if they are dealing with natural kinds, they erroneously generalize beyond what is philosophically justifiable and scientifically defensible, sometimes with tragic consequences. Consider what happens, for example, when certain scientists routinely refer to freezing in non-human animals as "fear" and the assembly of neurons that help to control this behavior as a "fear circuit." Other scientists discover that a certain drug can suppress this behavior or the circuit, leading them to claim generalized and explanatory knowledge about FEAR on the basis of evidence for specific physical actions (i.e., treating FEAR as a natural kind by making inductive generalizations and abductive inferences about a class of events that the research itself does not support). The result: A brain circuit that plays a role in learning to freeze in a specific situation is transformed into the circuit for instances of FEAR in all situations (Barrett, 2012), including those that cause psychopathology. With this philosophical slippage, scientists, drug companies, physicians, and patients struggling with "fear-related disorders" assume that the drug in question will tame anxiety-related symptoms, but it doesn't. The neuroscientist Joseph LeDoux has implicated such errors in the widespread failure of research using animal models of ANXIETY to develop medications for anxiety disorders, which led pharmaceutical companies to abandon further research and development (Ledoux, 2015).

Summary

Empiricism, and its attendant assumption of empirical kind categories, holds that anything unobservable is unknowable and therefore irrelevant to science. As a philosophy of science, empiricism has always been more about the epistemic norms for practicing science than a description of what scientists actually do. In the words of philosopher Peter Godfrey-Smith, empiricism

> has often been a story about how the mind (or language) works, along with a recommendation to do things in a way that best fits with this picture — not to deceive oneself about possible sources of knowledge, to be mindful of our fallibility, and to retain openness and flexibility" (Godfrey-Smith, 2021, p. 325).

Empiricists of all varieties were generally averse to anything unobservable. Assumptions about the true nature of reality (i.e., ontological assumptions) were to be generally avoided. This is a lost cause, however, because assumptions about reality cannot be avoided. All epistemic decisions entail ontological assumptions in one way or another. Even behaviorists who overtly avoided lofty questions about latent constructs and aimed their scientific efforts at empirical kinds were making implicit claims about the nature of reality via their actions. Ironically, one implicit claim makes behaviorism and empirical kinds, as an example of anti-realism, similar to the traditional scientific

realism of natural kinds: Both presume a conventional view of reality based on sameness and stability in a natural world that exists independently of humans and their concepts.

In both traditional scientific realism and the various forms of empiricism, categories that are assumed to exist in a perceiver-independent physical world, including the categories that describe human minds and actions, are further assumed to have fixed, unchanging intensions (features of equivalence), waiting to be discovered by human scientists with their sensory systems. Each extension (instance) is likewise assumed to be an enduring, independent object or process (i.e., an "entity" in philosophical jargon), possessing these inherent, unchanging features. In this conventional view of reality, any two entities (e.g., hydrogen and oxygen atoms, or an instance of cognition and an instance of emotion) are inherently separate. They can influence one another by simple cause and effect, like bits and pieces of a machine. One entity can cause a change in the other's behavior, or the two may interact to influence the behavior of some third entity, but the internal constitution of each entity, its features, its causal powers — its nature — does not change as a consequence of interaction. When a scientist observes individual interacting entities, they should be able to make inferences to other instances of the same category, and vice versa. Or so the story goes.

If psychological phenomena are contextually variable and complex in reality, however, then both natural kinds and empirical kinds and their shared assumptions about the nature of reality may be ill-suited to develop robust, justified knowledge in psychological science — knowledge that replicates in the lab and is useful in the real world. This possibility is considered in the third and final section of this chapter.

A RELATIONAL PHILOSOPHY FOR PSYCHOLOGICAL SCIENCE

Assumptions about the reality of natural kind or empirical kind categories gives rise to a certain way of doing science in psychology. Controlled experiments are designed to engineer situations in which hypothesized phenomena consistently arise, so as to discover knowledge that can be generalized to all humans. Attempts are made to control or reduce the impact of all but one or two factors of interest by holding all other "hidden moderators" constant across participants or by measuring their influence and statistically removing the variance they cause, by increasing sample sizes in an attempt to drown them out with the signal of interest (using the logic of classical measurement theory), or by ignoring them and hoping the (error) variance they contribute will be randomly distributed across observations.

Psychologists have periodically noted that this epistemic approach produces unsatisfactory results and have proposed best practices that have more rigor, as if additional rigor will solve the problem. Another option, however, is to question the approach's basic premise and utility for psychological science. The reality of a human mind might not be structured as a perspective-independent system of fixed, universal categories that are governed by simple mechanistic laws so that it is possible to build justified knowledge about each category individually without affecting the others. Conventional assumptions about the nature of reality emphasize the independence of categories from concepts, the sameness of instances within categories (rather than variation that comes from contextuality and situatedness), and the simplicity of their mechanistic impacts (rather than causal complexity). Such an approach may fundamentally misunderstand the phenomena of interest,

however. What if reality is *relational* and *situated*, continually in flux as moments of temporary structure from complex ensembles of causes? Objects and processes would then be dynamic *events*, not static fixed ornaments of reality.

These philosophical notions of a relational reality, rooted in context and complexity, have been suggested at various times throughout the history of psychological science. William James, in fact, planted the seeds of a fully relational philosophy of science at the very dawn of psychological science in a series of essays. Everything that is real — a category, its instances and features, its causal impacts — is assumed to exist *only in relation to something else*. And much of the time, particularly in psychological science, that "something else" includes people: namely, the meaning-making capacities of perceivers who actively interact with one another and the rest of the ecology they live in. This philosophy is called *relational realism*.

The situated nature of reality, emerging from relational meaning, fundamentally challenges the conventional notion of reality. It jettisons the deep ontological commitment to fixed, universal, perspective-independent categories as collections of fixed, contextless objects and processes in ways that echo relational approaches to quantum mechanics in modern physics. The categories that result, called *relational kinds*, are temporary groupings of *relational events*, constructed with functional (abstract or unobserved) intensions to guide action in specific situations. This section briefly reviews some early relational ideas in psychology, followed by an example of relational realism in the science of emotion. The section ends with suggestions for how a relational shift might manifest in the philosophy of science that is practiced by psychological scientists.

One proviso before diving in: The relational view of reality being suggested here is not extreme relativism (Ashton, 2020; Bloor, 2011). "Relative" does not mean "completely subjective." Nor does it mean "arbitrary." "Relative" means only that reality is conditioned on some perspective (i.e., reality is "not absolute"). Absolutism is traditional scientific realism and empiricism, with their notion that reality is structured as universal, objective categories in a world of fixed, objective objects and processes. Such absolutism allows for universal, objective standards for justified knowledge about those entities and their categories. The basic ontological requirement for relational realism is not that reality exists only in a person's head (idealism). Nor are there many realities, all equally true. There is *one reality* that is inherently perspective-dependent and flexible, although not infinitely so.

William James's Radical Empiricism

Radical empiricism is a philosophical attitude introduced by William James (James, 1897/2009) that obliterates the idea of a mind-independent reality. In *Principles of Psychology*, James voiced concern that the emerging science of psychology was rife with folk psychology categories and common-sense essentialism, both of which he thought were ill-suited foundations for building a robust, useful science. James coined the term "the psychologist's fallacy" for the mistaken belief that perception, memory, and other psychological categories were physically distinct in some objective way. "The cardinal passions of our life, anger, love, fear, hate, hope, and the most comprehensive divisions of our intellectual activity, to remember, expect, think, know, dream, with the broadest genera of aesthetic feeling, joy, sorrow, pleasure, pain," wrote James, "are the only facts of a subjective order which this vocabulary deigns to note by special words" (James, 1890/1931, pp. 194–195). For James, a question such as "What is the difference between an attitude and an emotion?" was similar in meaning to "What is the difference between one situated instance of anger and another?" In James's words, "A science of the relations of mind and brain must show how the

elementary ingredients of the former correspond to the elementary functions of the latter" (James, 1890/1931, p. 28).

James further developed these ideas in his essays on radical empiricism, written near the end of his life. The essays are very difficult to understand because they violate common sense, but also because James had no lingua franca to describe his ideas. In addition, James was working with an outdated understanding of the body and brain. Still, a close reading of these essays in their historical context makes it clear that James was rejecting the existence of a mind-independent reality structured as natural kind categories. Instead, he suggested that reality emerges from continuous transactions between a person and the world. Radical empiricism is therefore not like other varieties of empiricism — it does not distinguish between perceivable (knowable) and non-perceivable (unknowable) parts of reality. James was suggesting something much more … radical: that reality *is* perception. Reality, according to James, always and necessarily requires a perceiver (James, 1905, 1897/2009): The objects, processes, and other entities of reality emerge from a continual transaction between whatever is out there in the world and the concepts in a person's head. Using more modern language: The objects that people hold, the events that they see, the processes they infer in the minds of others — these aspects of reality emerge when signals from the world interact with the concepts in a perceiver's brain (concepts which are, themselves, a bunch of signals).

To explain James's radical empiricism with a modern scientific example, consider the dark green leaves of an apple tree. "Dark green" is not a property that is inherent in a leaf or in any other object. It is a *feature of the relation* between signals of light at certain wavelengths (e.g., around 510 nanometers) in the midst of other reflections at other wavelengths, signals in the cones of an animal's retina that transduce this light, and signals inside the animal's brain. A person who is missing certain cones, leaving them unable to distinguish certain wavelengths of light, would experience the leaves as yellowish-green (i.e., chartreuse) and the tree's apples as a muddy green-brown rather than red. So, too, would mammals with neurotypical dichromatic vision (e.g., dogs, bulls, rabbits, and horses). Traditional realism would suggest that the leaves are objectively dark green and the apples are objectively red and these creatures just can't see reality (hence the term "colorblind"). In a more relational view of reality, red and dark green are not part of reality as it pertains to creatures with dichromatic vision, consistent with James' radical empiricism and the understanding of many modern scientists who study vision. To assume otherwise is to implicitly define objectivity by what is neurotypical for humans. If neurotypical human vision involved only two cones or four cones, then apples would not "objectively" be red and leaves "objectively" green — they would "objectively" be some other colors. Reality, as it pertains to neurotypical humans, would be different. So, red and green are relationally real, as are trees, apples, and all other objects, processes, and events that are part of human reality. They exist as relational transactions between complex ensembles of physical signals that make up a perceiver and the physical signals in the world.

James's radical empiricism is a flavor of ontological realism. Something exists out there in the world. Modern scientists might call it "matter," "quanta," or just "signals," meaning any quantity that can vary. James called it "stuff" and "pure experience." On their own, signals are not reality per se. To become reality, signals must interact with other signals. The world of solid objects, such as a physical book filled with pages, is an event; it exists as a relation between signals in the world and signals that make up a person. Book and perceiver are not two separate, interacting entities. They are one unified event that emerges from an ensemble of interacting signals. This view is identical to the relational view of quantum mechanics (Rovelli, 2021). In modern terms, the interacting signals

that create the object are categorized as an instance of a BOOK to a human reader, but they might be an instance of FUEL to someone building a fire and FOOD to a booklouse. To claim that interacting signals are inherently a "book" in a perspective-independent reality is to elevate one perceiver and their needs over the others.

Continuing with modern language, James was suggesting that reality — not "one person's experience of reality versus another's" but bona-fide reality — manifests as the stream of experience that emerges as sensory signals stream to a brain and interact with the signals there (or, in James's words, "stuff" and "pure experience" interact with a mind that has "apperceptive ideas" or "concepts"; James, 1907). Psychological meaning, which James called "knowing," emerges from the continual relations between sense data and concepts.

James's radical empiricism foreshadowed the concept of an *ecological niche* from the science of ecology: the environmental conditions (i.e., the physical signals in the environment) that are relevant to an animal's survival (Pocheville, 2015). Only those signals create reality as it pertains to that animal. Ecological psychology, in fact, has deep intellectual roots in radical empiricism (Heft, 2001). In his essays on pragmatism, James phrased things slightly differently, replacing survival with utility (referring only the physical signals that are useful to a person or that a person can act on; James, 1907). A modern neuroscientist might instead appeal to abstract features of "salience" or "value" (e.g., Rangel et al., 2008; Seeley, 2019). Signals that are useless or irrelevant to survival will always be noise (in relation to a perceiver).

A couple of additional points are useful. James's radical empiricism is meant to be both an evolutionary perspective, which he thought psychology lacked at the time, and an assault on Descartes's mind-body dualism, which he thought psychology had too much of. According to James, humans are biological creatures with particular sensory, motor, and conceptual abilities who evolved under the selection pressure of a dynamically changing physical world. These abilities maintain a reciprocal relationship with the physical world. To translate this idea into modern jargon: A brain inhabits a body and a physical world full of signals that wire and constrain it, and that brain, in turn, shapes its body and the world (a.k.a. niche construction; K. Laland et al., 2016). A mind, as a continual series of dynamically shifting mental events, including motor movements and their meanings, emerges from these ongoing relations. With radical empiricism, James was attempting to free psychology from various ontological dualisms that he deemed philosophically pernicious: objective vs. subjective, mental vs. physical, mind vs. reality (or representations of things vs. the things themselves), and the knower and the known.

Epistemic Critiques Of Experimental Social Psychology

The history of social psychology contains several critiques of controlled laboratory experiments as ill-suited for cultivating and observing social phenomena. Embedded in these epistemic critiques is the ontological suggestion that reality is complex and inherently relational. For example, contrast the views of Floyd Allport, Gordon Allport, and Leon Festinger versus those of Kurt Lewin and Solomon Asch, discussed earlier. Their debate, seemingly about methods, can also be understood as an argument about the appropriate ontological focus for a science of social psychology. Lewin and Asch were advocating for a view of social behavior as inherently relational, emerging from complex instances of social embeddedness; therefore, any behavior could not be ontologically reduced to its parts (Danziger, 2000). The Allport brothers and Festinger assumed that individuals functioned like social atoms who interacted with and reacted to an independent social world.

A similar debate resurfaced in the 1970s, this time couched in critiques of experimental social psychology. Some critics suggested that social psychologists routinely ignored the complex array of causal influences in standard laboratory experiments (e.g., Cronbach, 1975; Gergen, 1978; McGuire, 1973). Others highlighted the problems with studying individuals as social atoms rather than as the groups and communities they inhabit outside of the lab (Greenwood, 2000, 2004). At the time, as in the present day, shouts of a "crisis" in psychology were widespread (e.g., Elms, 1975; Pepitone, 1976; Sherif, 1977). Then, as now, the modal response was to double down on methodological rigor and enshrine the standard laboratory experiment as the gold standard for building justified knowledge — effectively ensuring that the relational embeddedness of psychological phenomena was outside the scope of study (Greenwood, 2004, Ch. 9). Consider, for example, the widespread use of "minimal group" manipulations in social and developmental psychology (e.g., Tajfel, 1970; for review, see Dunham, 2018): Do minimal groups really have the same relational properties as groups that exist outside the lab? Are scientists justified in assuming that knowledge claims about one will apply to the other? Then, as now, methodological rigor was improved (because there is always room for improvement), but the ontological assumptions went unquestioned.

Emotions As Relational Events

In the science of emotion, persistent epistemic critiques of the typological research tradition gave rise to constructionist accounts that understand emotions as inherently relational events. Historical critiques, beginning with William James (1884, 1890/1931, 1894/1994) onward (e.g., Dashiell, 1928; Duffy, 1934a, 1934b, 1941; Dunlap, 1932; Harlow & Stagner, 1932, 1933; Hunt, 1941; Landis, 1924; Sherman, 1927, 1928), concluded that ANGER, SADNESS, FEAR, and other English folk categories of emotion were not suitable scientific categories for accumulating justified knowledge. In study after study, instances of the same folk emotion category were observed to vary too much in their facial expressions, autonomic nervous system changes, and other concrete sensory and motor features for the category to be considered a natural kind. At the same time, a single concrete feature or pattern of features was observed to belong to instances of different categories. The solution to this many-feature-to-many-emotion mapping was to hypothesize that anger, sadness, fear, and other emotion instances only emerge when sensory and motor changes were made meaningful as such (see Gendron & Barrett, 2009 and references therein). All these ontological efforts emphasize the ways in which humans create meaning and, in so doing, construct instances of emotion that *intrinsically* depend on context.

Social constructionist accounts emphasize the relational importance of social and cultural meanings in creating instances of emotion. Early accounts focused on social affiliation (Schachter, 1959; Schachter & Singer, 1962) and the social meaning of situations (e.g., Dashiell, 1928; Dunlap, 1932; Landis, 1924; Sherman, 1927). More modern social constructionist accounts emphasize a causal role for beliefs, values, other people's actions, social roles, and various sociocultural artifacts in the creation of emotion (e.g., Averill, 1980; Boiger & Mesquita, 2012; Harre, 1986; Mesquita, 2022; for a historical and anthropological overview of social constructionism, see Reddy, 1997). All of these hypotheses in one way or another are concerned with the relational nature of emotional reality.

Psychological construction accounts propose that instances of emotion emerge when affective feelings are made meaningful via influences from psychological ingredients. Historical accounts of meaning-making (e.g., Harlow & Stagner, 1932, 1933; Hunt, 1941; Mandler, 1984, 1990) were less specific than more recent accounts, which have proposed that affect is shaped into emotion by a combination of attribution and categorization with emotion concepts as prototypes (Russell, 2003),

by abstraction over time (Cunningham et al., 2013), or by situation-specific categorizations that occur on the spot, in specific contexts, called situated conceptualization (Barrett, 2006b; Barrett et al., 2015; Lindquist, 2013; Wilson-Mendenhall et al., 2011).

The theory of constructed emotion (Barrett, 2012, 2017b, 2017c), an interdisciplinary constructionist view, proposes that concepts, categories, and their instances are momentary, contextualized events in a brain. An instance of an emotion category, such as ANGER, for example, is hypothesized to begin as *an ad hoc, situated category* — a brain-wide ensemble of signals that creates a pattern of abstract features associated with multiple patterns of possible concrete features (see Appendix, Section 8). The abstract features are the intension of the ad hoc category and define the function or goal that the category meets in that specific situation, rather than a stable function that remains the same across all individuals and situations as suggested by Adolphs and Anderson (Adolphs & Anderson, 2018; D. J. Anderson & Adolphs, 2014). For example, the goal of an instance of ANGER might be to remove an obstacle, punish someone, win a competition, or feel part of a group (e.g., Sinaceur & Tiedens, 2006; Van Kleef & Côté, 2007; van Zomeren et al., 2004), or even to avoid harm by attempting to appear powerful (e.g., Ceulemans et al., 2012; Sinaceur & Tiedens, 2006). The patterns of possible concrete sensory and motor features are possible extensions of the ad hoc category. They allow the individual to flexibly enact the function and meet the goal in that specific spatiotemporal context. The functional features make the possible concrete sensory-motor ensembles equivalent for meeting that goal in that situation. The specific skeletomotor movements that functionally enact a goal (e.g., to remove an obstacle) will depend on the specifics of the situation and the energetic state of the body. Accordingly, the person might scowl and shout in anger, cry in anger, freeze in anger, or even laugh — whatever action concept the brain has learned to construct to meet the goal in similar situations. The movements of the heart, lungs, and other tissues in the body (visceromotor movements) that support the movements of the body (skeletomotor movements) will also vary with the situated goal. Even the affective features of an instance of anger might vary (e.g., they might be pleasant or unpleasant; e.g., Harmon-Jones & Peterson, 2009). And the neural signals that construct the category, which are distributed across the brain and give rise to these varying features, can themselves vary from situation to situation, instance to instance, a phenomenon known as degeneracy (e.g., Doyle et al., 2022; Edelman & Gally, 2001; Marder & Taylor, 2011). This hypothesis explains how different ensembles of physical signals can have the same relational emotional meaning across situations and persons, and also how a single physical signal or pattern of signals can have different relational meanings in different situations, in the same way that any object or event can be conceptually categorized in a flexible, situated manner (for a summary of evidence, see Barrett & Lida, 2024).

Across a person's lifetime, assuming their brain is equipped to construct ad hoc ANGER categories, they will construct a population of *n* possible ANGER categories (i.e., a distribution of categories), each with graded similarity in their features. Think of this as a *vocabulary* of ANGERS. And, in principle, the vocabulary of ad hoc ANGER categories (with graded distributions of features) can vary across people, particularly if those people come from different backgrounds with different opportunities to culturally inherit variable knowledge about anger. This relational hypothesis holds for every other category that a human brain is equipped to construct.

In this view, category construction occurs automatically and continuously through an individual's life. It proceeds via prediction, selection, and correction, together known as predictive processing (for discussion and evidence, see these papers and references therein: Barrett, 2017c; Barrett & Simmons, 2015; Chanes & Barrett, 2016; Hutchinson & Barrett, 2019; Katsumi et al., 2022, 2023). Each psychological event *begins* as a category, constructed as an ensemble of interrelated, temporally

evolving physical signals that are assembled across the entire brain and *ends* as a specific instance of the category. The originating signals, known in the psychological literature as "prediction signals," "simulation," "perceptual inferences," "top-down influences," or simply "memory," constitute signal patterns of possible features. The brain continually checks these prediction signals against ongoing signals from the body's sensory surfaces (relaying information about the state of the body and world). Incoming sensory signals, along with attentional signals called "precision signals," help to select the pattern of signals (or features) that will result in a plan for coordinated motor actions and conscious experience. Once selection occurs, the incoming signals and the signals that control motor actions are said to be categorized, and the functional features of equivalence are said to explain actions and their associated sensations.

As part of category construction, a brain *adds meaning* over and above the incoming signals from the sensory periphery by constructing abstract features that impose functions on those signals that they do not have by their interactions alone. When multiple brains become wired to construct the same functional feature, concrete physical changes like facial movements, physiological changes, and other behaviors acquire new meaning, and the result is called *social reality* (Barrett, 2012; Barrett, 2017b, 2017c). The abstract categories of social reality, such ANGER or MONEY or RED, further require relations between signals *across* coordinating human brains for those categories to be useful in communication. The categories of social reality, then, are relationally real in nature — a nature that includes interacting signals in human brains and bodies.

If categories are situated events constructed in individual brains, how is it that people share the same relational kind categories? In brief, neurotypical humans have roughly the same bodies and roughly similar brain architecture with similar compression gradients that construct abstract, functional features. Also, people in a given culture learn to construct similar abstract features through language that aids the transmission of relational kind categories of social reality across generations, called *cultural inheritance* (e.g., Boyd et al., 2011; Richerson & Boyd, 2006). The evolution and development of the human brain creates the preconditions that allow for cultural inheritance (see Barrett, 2020, 2021b; Gendron, Mesquita, et al., 2020). Brains receive wiring instructions from their body and the world. Humans have genes that allow a child's brain development to be biologically shaped by (and coupled to) the sensory signals from their body and the sensory signals from the actions of others who interact with them (i.e., conditions of the social environment). Signals from the physical world, including those arising from social behaviors, are necessary inputs for a human brain to construct specific abstract mental features. During development and the processes that scientists call "socialization," the words (S. A. Gelman & Roberts, 2017) and actions of others (e.g., Atzil et al., 2018; Barrett, 2020; Gendron, Mesquita, et al., 2020; Mesquita, 2022) create recurrent situations that allow a brain to learn specific, situated meanings of particular signals in the natural and cultural ecology of a person's environment.

Words, in particular, are powerful invitations to learn categories (Waxman & Gelman, 2010), even for very young infants (Vouloumanos & Waxman, 2014) and are particularly useful for learning categories with abstract features of equivalence that are necessary for emotion categories (for discussion and references, see Barrett, 2017b; Hoemann, Wu et al., 2020). This arrangement creates opportunities for cultural inheritance to transfer emotion concepts and other cultural knowledge across generations (see Barrett, 2017b; Gendron, Mesquita, et al., 2020; Lindquist et al., 2022).

This explanation is simplified, but it captures the hypothesis that emotion categories do not exist as enduring, universal groupings of entities in the world. Categorizing, in this sense, does not involve "detecting regularities" but constructing a category to constrain the meaning of an interacting signal

event in a particular situation, rendering that event meaningful to the perceiver in terms of some function. Instances of categories are not encoded and then categorized. The brain constructs the category first, as an ensemble of prediction signals, to anticipate a future instance and contribute to its constitution. In short, an instance of a category emerges *from* a category construction event. From a relational perspective, a "category," a "concept," and "categorization" all refer to the same process: a brain's attempt to reduce complexity and therefore uncertainty. The boundary between observer and observed is dissolved.

Stated in more general terms, instances of emotions are hypothesized to emerge from a brain that is engaged in continual category generation (Barrett, 2017c, 2022; Barrett & Lida, 2024) or "conceptual acts" (Barrett, 2006b). In philosophy, these categories are *relational kinds*. They are constructed in real time, online to serve a function and guide action in a specific situation (Barrett, 2006b, 2017c; Barsalou, 1982, 1983, 1985, 1986, 1987, 1989, 1993; Barsalou & Sewell, 1984; reviewed in Barsalou, 1991; Casasanto & Lupyan, 2015; see Appendix, Section 8). The equivalence of instances stems from their shared function in a specific situation rather than similar sensory and motor features. Emotion categories, as ad hoc, relational kinds, are perspective-dependent in an inclusive way: They are dynamic, probabilistic, and variable groupings of relations that themselves are dynamic, probabilistic, and variable. Note that the word "variable" here does not mean random. It means structured by relations within situations (and "situation" here includes the condition of the body as well as the outside world).

Correspondingly, in this view, a folk category like ANGER or FEAR is vast and full of situated variety. Some of that situational context is in a perceiver's brain. Sensory signals from the world have no inherent psychological meaning. Incoming signals are *made* meaningful when constrained and categorized by signals in the brain. Ditto for changes in heart rate, respiration, vocal acoustics, facial movements, neural firing and so on; they have no inherent, biologically prepared emotional meaning. Different combinations of signals on different occasions can have the same emotional meaning. Likewise, the same signal can have different meanings. Not infinite meanings, but more than one. The variety of muscle movements that produce a scowl may variously be an expression of ANGER, CONCENTRATION, or a bout of INTESTINAL GAS. Emotional meanings are properties of the relation between signals in the body and the world (the sensory present) and the brain (the remembered present).

In this relational view of reality, emotion categories are not part of some perceiver-independent reality because there is no perceiver-independent reality. Emotion categories and the instances of emotion that they produce emerge in a brain as the interaction of intrinsic brain signals and the signal ensembles that continually arrive from the sensory surfaces of the body, which transduce signals in the body and in the world. Neurotypical human brains construct categories predictively, as ensembles of signals, to guide action in an efficient manner. In doing so, those signals give meaning to other incoming sensory signals to guide action and create lived experience. The resulting emergent event is an instance of emotion or of any other psychological category. Categories are in a brain that is wired by the body and the world. So are concepts. The distinction is meaningless. Channeling James' radical empiricism, the theory of constructed emotion proposes a theory of brain function that dissolves intuitive but artificial ontological dualisms such as mental vs. physical, the perceiver vs. the perceived, and ultimately, objective reality vs. subjective minds.

Relational Realism And Relational Kinds

Some of relational realism's core ideas can be found in radical empiricism, critiques of experimental social psychology, and constructionist views of emotion. Reality is assumed to be *variable* and *contextual*, not immutable and absolute. It is constantly in flux as contextual constraints change to create relational meaning. Objects and processes are relational *events*. So are categories. Features describe relations. These ideas, and the observations that support them, paint a very different picture of psychological phenomena than the traditional realism of natural kinds or the anti-realism of empirical kinds. A boots-on-the-ground example of relational realism in psychological science is presented in Appendix, Section 9. A quick comparison of traditional realism and relational realism is as follows:

1. Traditional realism: Reality is perspective independent. Knowledge about reality can be perspective-independent (i.e., consensus among scientists) or perspective-dependent. 2. *Relational realism*: Reality is perspective dependent. Knowledge about reality is perspective dependent.

3. Traditional realism: Categories exist in reality, separate and independent of concepts, which exist in a living creature, such as a human. *Relational realism*: Categories and concepts are synonymous. Categories exist in reality as momentary relations between the signals that constitute the living creature who is doing the categorizing and the signals outside that creature (a.k.a. in the "outside world").

4. Traditional realism: Reality is structured as natural kinds, enduring groupings of instances that are similar based on inherent features (i.e., each category has a specific and unique intension). Each grouping is unchanging and universal in time and space (i.e., its extensions remain stable across contexts). Its intension can be an essence or a highly probable cluster. All atoms in the universe belong to the category HYDROGEN, for example, if they contain one proton in their nucleus, and this feature explains their behavior in all circumstances. *Relational realism*: Reality is structured as relational kinds, dynamic, situated events whose features of equivalence are contextual. A category's extensions also vary with context. Variation is not random but structured by relations within situations. For example, atoms such as HYDROGEN can be grouped by the number of protons in their nucleus, as in the classic periodic table or by quantum numbers that describe the movement of their electrons, depending on a scientist's goals. The classical category HYDROGEN is an idealization that does not exist in reality; it is functionally inert and useless to chemists in any practical way until they add in contextual details.

5. Traditional realism: An instance of a category is a physical object, process, or event — something that exists on its own, independent of other things. For example, an atom is a very small particle or a wave. A star is a big ball of gas. *Relational realism*: An instance of a category is interactions or relations among signals. A signal is simply any quantity that varies in time and space. An atom is an interaction among signals, and so is a star. Every scientific observation of an atom or star is also an interaction among signals, one that employs fancy instruments to convert signals that humans cannot detect into those they can, and those signals interact with other signals in scientists' brains and bodies.

6. Traditional realism: A feature is a fixed, inherent property of something. A feature can be physical or mental, and mental features can be concrete or abstract. Every category instance shares certain inherent features with other category members (deterministically or with high probability) that make the instance the kind of thing it is. *Relational realism*: A feature is a property of the relation between signals. The velocity of an atom or a star, for example, is not an inherent property of the atom or star. It is a property of the atom or star in relation to something else. Contextually constrained interactions between signals in a spatiotemporal context give rise to an instance's

features, making the instance what it is. Any given instance will have many, many physical features (e.g., chemicals, movements, etc.) and concrete mental features (e.g., lines, edges, cones, etc.). Abstract features are compressed, multimodal summaries, such that different high-dimensional concrete features are rendered equivalent by sharing fewer low-dimensional, functional features. This is how many instances that are physically different can serve the same function in a specific situation. This function is the situated intension of a relational kind.

7. *Traditional realism*: Causation involves simple mechanisms. *Relational realism*: Causation is complex.

Relational themes are woven throughout psychological science, usually without the name. The importance of meaning-making (construal), complexity, and contextuality are prominent themes in Gestalt psychology, in social construction (e.g., Gergen, 1985), ecological psychology (inspired by James's radical empiricism, e.g., Heft, 2001), and embodied cognition (e.g., Barsalou, 2008; Varela et al., 1992). A relational mindset is consistent with several of the most basic ideas in social psychology, particularly its emphasis on context (e.g., Bruner, 1990; Markus, 2017, 2021; Markus & Kitayama, 1991; Mesquita et al., 2010) and the "principles" of situationism, construal and complex systems in social cognition (Ross & Nisbett, 1991), which basically suggest that reality is a series of momentary, relational events emerging from complex interactions that involve construal in context. There are also vast literatures on constructionist approaches to memory and perception, and on transactional approaches to development, as well as the rapidly growing number of predictive processing accounts of mind, brain, and behavior that have emerged in the last two decades, all of which can be easily understood within relational realism (for reviews, see Barrett, 2017b, 2017c; Hutchinson & Barrett, 2019; Pezzulo et al., 2015, 2021; Shaffer et al., 2022): Actions, and their accompanying mental features, begin as complex, top-down constructions in the brain, fashioned from past experiences, that are subsequently tested against incoming signals informing on the state of the body and the world. More generally, most of the major topics in psychological science are about meaning making but go by different names, such as perception, memory, categorization, attention, and decision-making, as well as generalization and latent causal inference (Gershman et al., 2010). Every scientific category discussed in this *Handbook* can, in fact, be understood as a relational kind in relational reality.

The constant challenge in a variable and contextual reality of interacting signals, whether in life or in science, is to constrain the variation enough to create meaning that is useful, but not overly constrain it to create a simplification that is misleading and potentially useless. The sweet spot between complexity and simplicity offers maximal utility.

Key Assumptions

In relational realism, the objects and processes of experience, as they occur in life or as curated in experiments, are not independent entities but the products of a web of causal influences. Every instance of every category is a collection of instantaneous events or ensembles of relations. Instead of "substances," think "interacting signals." An electron, for example, is not a particle or a wave — in relational quantum mechanics, an electron is understood as an event that consists of signals interacting at a particular time and place (Rovelli, 2021). A classical view of reality is an abstraction or simplification of relational reality. Brains and bodies also consist of interacting signals. In psychological science the word "signal" usually refers to meaningful information (i.e., the opposite of "noise"). In a relational view of reality, however, the meaning of "signal," which comes from

engineering, is more general than the psychological one: It's any quantity of anything that varies. When signals become entangled, they *manifest* information, as discussed below.

A set of interacting signals is sometimes called a "system" (Rovelli, 2021). A neuron's action potential, for example, is a system of interacting ATP concentrations, Na+ and K+ pumps, etc. Similarly, every variable that is measured in every experiment is a system of interacting signals. When a scientist observes one electron influencing another, they are really observing the relations between two systems of interacting signals. A neuron, as a system of interacting signals, can itself participate in even more complex systems (e.g., in an assembly of neurons in the brain). A brain is a dynamical system of signals that interact with incoming signals from the body (interoceptive signals) and the world (exteroceptive signals).

The boundaries of any system are flexible. The brain can be parsed into subsystems in various ways. The body can be likewise parsed into ensembles of signals in any number of ways. The brain and body together could be considered a system. A person interacting with other people could be a system. Scientists, using their own interests and goals, decide which boundaries to place and where to place them, parsing complex signal ensembles into a system vs. its inputs and outputs.

Complexity in relational events

The word "complexity" implies variation. (It's opposite, "simplicity," implies sameness.) When signals don't interact — when they exist independently of one another — complexity is abundant. For n independent signals, there are 2^n outcomes. This variation is nothing more than noise, however. Think William James's "blooming, buzzing confusion" (James, 1890/1931).

When signals interact, they *constrain* one another, resulting in fewer outcomes. Constraint produces order, referred to as *contextual coherence* (Juarrero, 2023). Coherence is a *reduction* in complexity from which new levels of organization *emerge*. The blooming, buzzing confusion becomes predictable moments of organization that create the perception of the classical world of enduring objects and processes (i.e., entities; Rovelli, 2021). As signals become entangled, therefore, reality becomes less complex.

Contextual coherence reduces complexity but does not obliterate it (i.e., it does not produce the sameness that is expected in the natural kinds of traditional scientific realism). More structure is apparent, but there is also variation — structured variation. This is because systems of interacting signals are themselves complex ensembles of many dynamically varying parts, each weakly causal on their own. When they interact nonlinearly with one another, however, they produce powerful collective outcomes, and usually more than one (i.e., heterogeneous outcomes). These outcomes are variable but, again, the variation is structured, not random. Hence the saying, "A system is more than the sum of its parts." (Contrast this with the earlier description of natural kinds as "carving nature at its joints.") New levels of organization within a complex dynamical system of signals cannot be reduced to the individual signals themselves.

In relational realism, a scientist's job is to discover and learn about structured variation. A traditional scientific realist starts with the assumption of sameness and simplicity and discovers unexpected variation piecemeal that must be explained post hoc (e.g., probabilistic clusters of features instead of essences). A relational realist expects complexity. Their scientific challenge is to observe and learn the constraints that produce structured regularities within.

Any interaction among signals gives rise to a set of *features* that describe relations. The thunderous thud of a falling tree is not a feature of the tree alone; it's a feature of the relation between the air and ground vibrations in the world as the tree lands, the signals in the human ear as hair cells detect those vibrations, and the signals in a human brain. The color red is not a feature of roses, cherries, sports cars, and Chanel lipstick. Red is a feature of reflected light waves, detected by a human retina, in relation to the signaling within the human brain. Deliciousness is not a feature of dark chocolate, French fries, and crusty bread. It is a feature of the relations between the molecules in those foods, the signals within a body (e.g., signals from the bacteria in the gut, signals flying up the vagus nerve, signals from tastebuds, signals to the olfactory blub, etc.), and the signal ensembles in a human brain. Paintings are not inherently beautiful, photographs are not inherently valenced, money is not inherently valuable, jokes are not inherently funny, and insults are not inherently threatening; beauty, valence, value, humorousness, and threat are not features of objects and processes, but features of relations.

Meaning (construal) in relational events

Features are *meaning*. When signals are not interacting with one another, they have no inherent meaning. When signals constrain one another, creating features, one signal reveals something predictable about the others. This reduces the variation and uncertainty that is possible during any event, producing *mutual information* (Shannon & Weaver, 1949/1964) as signals become meaningful in relation to one another (for a similar discussion in terms of systems, see Höhn, 2017; Rovelli, 2021). Meaning, here, is not defined as what a signal *is* but by what it *does* when interacting with other signals. Meaning is *function*, described by a system's emergent features. For example, the signals in a brain continually constrain signals from its body and the world, creating relational meaning. This is how a racing heart becomes an instance of EXCITEMENT, FEAR, or OVER-CAFFEINATION.

The interactions between signals — their features — also allow scientists to predict the behavior of the system; this is what the physicist Rovelli (2021) calls *relative information*. The features of a system, therefore, describe its capacity to interact with other systems of signals (i.e., the "information" the system can create with other systems). Relative information, therefore, is another source of relational meaning. Two people communicating with each other, for example, create relative information. Everything that exists in the social world is relative information. Relative information manifests reality. For a living creature, information in the form of increased predictability, described as features or relational meaning, sustains life.

Contextuality of instances

In the relational view of reality, a signal's momentary meaning is always relative to *all* the other signals it is interacting with. In psychology, this larger ensemble of signals is referred to as a *situation* or a *context*. Typically, "context" connotes epiphenomenal influences that might tweak or modify a phenomenon of interest. In a relational view of reality, however, context refers to all the signals that interact to create any given event that humans know as reality. Even relatively weak or ephemeral signals, the sort that traditional scientific realists relegate to the background, can be important to an instance of a relational kind.

A signal's meaning changes instantaneously as it leaves one ensemble and enters others (i.e., when the event changes). In this way, the muscle movements that create a scowl might have meaning as

an expression within an instance of anger, distaste, concentration, memory, confusion, stress, or discomfort from intestinal gas, depending on the larger ensemble of signals of which they are a part and in which they create relative information. Even the psychological meaning of an action potential is related to the action potentials of other neurons in the brain, called the neural context (McIntosh, 2004); for example, a given signal is considered a motor signal when received by motor neurons and a sensory signal when received by sensory neurons. By implication, a brain is a system of signals that creates contextual coherence with incoming signals from the body and the world, constraining them and reducing their complexity, giving them relational meaning. Any signal can be a context for processing signals from outside the brain (e.g., Azzalini et al., 2019; Engelen et al., 2023).

"Contextual" in a relational view of reality means contingent, conditional, and relative — one signal has meaning only from the perspective of another signal it's interacting with. In philosophy, such meaning is called *perspective dependent*. Humans participate in everything that is real *that pertains to them* because the signals in their bodies and brains are part of the ensemble of relations that give rise to reality. A perspective-independent reality that pertains to humans would not exist in any meaningful sense.

A "perspective" has a broad meaning as used here. In psychological science, perspective usually refers to a system of concepts shared by people in a culture, such as people who live in the same geographic area or people who share a scientific paradigm (Kuhn, 1962/1970), a research program (Lakatos, 1970/1978; Lakatos & Musgrave, 1970) or a research tradition (Laudan, 1977; also see, Carnap, 1950; van Fraassen, 2001, 2004; also see Foucault's discursive formations, referenced in Nickles, 2024). In principle, however, a perspective need not involve human concepts or human experience. *Conscious experience is not required* for interacting signals to emerge as reality. For example, a human garbed in a gorilla suit continues to exist in reality as it pertains to humans even when perceivers fail to consciously see the person stroll across a basketball court right before their eyes (Drew et al., 2013; Simons & Chabris, 1999).

A person can have a tumor that they don't consciously experience in any way, yet the tumor affects something in their body, such as their rate of oxygen consumption or degree of inflammation. That tumor exists in reality as it pertains to humans. Or consider various electromagnetic waves, which humans do not consciously experience but that impact human bodies. Neutrinos also are subatomic events that pass through human bodies undetected. Before their discovery, neutrinos were not part of human reality. Their discovery expanded reality as it pertains to humans. So, too, for dark matter. The concepts of gravity and light allowed physicists to discover the hypothetical substance "dark matter" as consistent anomalies in their observations. These anomalies discovery expanded reality, even if indirectly, as it pertains to humans.

On the other side of the galaxy, real events manifest as ensembles of interacting signals that humans have never experienced. Humans have concepts for some of those events (e.g., what scientists understand to be a "black hole"). Knowledge about reality always requires human concepts, but there could be real events that humans don't have concepts for and can't even conceive of (except using the notion of "mystery" or "enigma"). Philosophically speaking, it is straightforward to assume that instances exist in reality that are currently unknowable by humans. Perspective is not about experience. It's about relevance and meaning. Think of "relevance" as whatever is contained in an animal's ecological niche — the set of environmental circumstances that are relevant to survival (Pocheville, 2015).

Finally, there are cases in which "perspective" only peripherally involves a human. The temperature at which water freezes, for example, varies with atmospheric pressure. Water freezes at zero degrees Celsius at sea level on Earth, but the freezing point drops with ascent into the atmosphere. In a near vacuum, the freezing point can be as low as -273 Celsius. In space, liquid water would boil, not freeze. In this example, the proximal "perspective" is air pressure; the exact features of the interaction of water and temperature depend on it. The distal perspective involves human concepts, such as the concept of temperature, which was originally invented using and is ultimately grounded in human experiences of hot and cold (Chang, 2007).

In a perspective-dependent reality, the characteristics that describe an individual person, such as an attitude, a goal, or their ability to remember or resist distraction, always require human concepts and categories. These characteristics are not fixed, stable entities, true scores that are tweaked by this or that aspect of the context. For example, whether a person slept well or poorly the night before, whether they just finished a cup of coffee or a cup of water, whether they are anticipating a difficult meeting or a walk in the park after leaving the lab — all of these factors could influence their performance on a working memory task. A traditional realist would assume that these factors were hidden moderators or perhaps just noise, but certainly not part of the real cause of their working memory performance. If WORKING MEMORY is understood as a relational kind, however, then the person's working memory performance in any moment emerges from a complex ensemble of interacting factors that may include the amount of sleep, the caffeine they've had, and so on. Each characteristic is a distribution of complex, situated events. The variation is real. The mean of the distribution is a statistical abstraction. A traditional realist would assume the reverse, as noted earlier in the chapter — they'd assume the mean is real and the variation is epiphenomenal.

For humans, part of the perspective that contributes to reality comes from the categories they've inherited from their culture. Humans do not genetically inherit specific categories but instead they inherit a brain architecture for learning and constructing categories. Many of those categories are acquired as a brain's micro-wiring is continually shaped by incoming signals from the body and from the wider physical and social ecology. As a result, one flexible human brain architecture can make a variety of human minds. Human natures are not of infinite number (due to embodiment constraints; i.e., the physical facts of the body and brain, the way human eyes work, etc.), but there is definitely more than one. As children develop into adults and continue to interact with their world, they create some of the signals in the environment (by their words and actions) that serve as wiring instructions for the brains of the next generation. This is cultural inheritance in action.

One final note: Human meaning-making is constrained by signals in the world (and cannot exist without them). People cannot walk through solid steel walls, for example, but they can look at a solid steel automobile and change its function from a vehicle to a dwelling. Humans in Iceland can decide that the rotting flesh of a shark is food, while those in India can decide that the muscle from a healthy cow is not, but no matter how humans use concepts, they can never ingest ground glass as a nutritious meal. Signals in human brains and bodies, in consensus, contribute to reality in the same way as signals in the world, by mutual constraint.

Summary

To summarize the key ideas of relational realism: Relational reality is a complex web of variable, interacting signals that momentarily constrain one another, producing events with relational meaning. The contextual constraints inherent in these interactions momentarily reduce signal complexity, producing what humans experience as stable objects, processes, and events with stable

characteristics or features. Categorization is a further move towards simplicity and sameness and away from complexity and variation. Both the physical and the social aspects of reality are created in this way. Many of the categories that humans use to create reality are culturally inherited.

Relational Realism In Philosophy

Relational realism, as outlined here, belongs to a scattered philosophical tradition known as relationalism (Kaipayil, 2009), which has been around at least as long as people have wondered about the nature of reality and humanity's place in it. In Western philosophy, the threads of relational realism can be traced back to Heraclitus (as opposed to the atomists like Plato & Aristotle). The 18th century brought the scientist and philosopher Gottfried Leibniz's ideas that time and space are systems of relations between objects (Futch, 2008) and the 19th century brought the relational ideas of physicist Ernst Mach (Huggett et al., 2023). In modern times, some notable examples of relational realism include the relational interpretation of quantum mechanics (Di Biagio & Rovelli, 2021; Epperson & Zafiris, 2013; Rovelli, 2021; Van Fraassen, 2010); contextual realism (Bhaskar, 2008); natural constructivism (Gleiser, 2015); certain versions of perspectivism, which is the idea that perception and knowledge are always relative to the perspective of perceivers, such as perspectival realism (e.g., Giere, 2010; Teller, 2018, 2019); relational thinking (Candiotto, 2017); relational sociology (Donati & Archer, 2015); and, certain aspects of relativism (Baghramian & Carter, 2022). Relational realism is also similar to philosophical constructivism, constructionism and conventionalism (Goodman, 1978; Kuhn, 1962/1970; Latour & WooIgar, 1979/1986; for review, see Cohnitz & Rossberg, 2024; Longino, 2019; Mallon, 2019) or what the philosopher Peter Godfrey-Smith calls metaphysical constructivism (Godfrey-Smith, 2021, p. 233). A unifying assumption in this family of ideas is that reality is perspective dependent. The human mind is part of nature, and therefore, something that is mind-dependent is also natural and real (see Dupré, 1981). This notion has been called pluralistic realism (Case, 2001; Kitcher, 1984).

Relational Realism As A Philosophy For Scientific Practice

Relational realism is a conceptual framework and a set of tools for constructing justified scientific knowledge about reality. Some justifications for relational realism as a philosophy of science can be found in critiques of traditional scientific realism discussed earlier in this chapter. "The world as we know it and interact with it," writes Godfrey-Smith (2021, p. 234) "is partly a human construction. The way things are, or the facts, is dependent on our beliefs, language, theories" (and on human actions). To a relational realist, the relations of interest in science will always involve the signals in scientists' brains and bodies because those signals interact with and constrain all other signals of interest. To some extent, this relational context is extended in time and in space (to include the signals that make up other scientists past and present). This broader relational context fundamentally shapes what counts as reality and what constitutes justified knowledge about reality. These next sections consider some of the many epistemic implications of relational realism for deriving trustworthy knowledge from scientific practices, particularly in psychological science.

Scientific Categories Are Relational Kinds

The categories of science are relational kinds, not determined by physical reality alone because they also depend on scientists' concepts and interests. There are various ways to carve the world into

categories, each attuned to a particular purpose or use, and scientific categories always exist in relation to the scientists who employ them. "One of the amazing characteristics of nature," according to the celebrated physicist Richard Feynman, "is the variety of interpretational schemes" it admits (Feynman, 1965, p. 54). In the relational view of reality, scientific categories exist as interactions between the signals in scientists' brains and the signals from the periphery (including the experimental setup).

Any category can be considered a relational kind if its within-category similarities and between-category differences depend on the concepts of a perceiver. In the philosophy of science, this is sometimes known as the epistemic challenge of *boundary construction*: drawing temporary, situated lines that group certain instances as a category and separate them from all non-instances. Human brains place the boundaries that create similarity and difference. Relational kinds are not infinitely malleable, but neither are they fixed and universal. Natural kind and even empirical kind categories, from the perspective of relational realism, are considered theoretical kinds — simplifications that do not exist in reality and arise when contextual relations are ignored or hidden and therefore taken for granted (see Appendix, Section 5).

One example of relational kind categories in the natural sciences comes from the study of stars in astronomy. A star is a luminous ball of gas, mostly hydrogen and helium, that is held together by its own gravity. Stars don't exist neatly packaged into categories by nature. They vary continuously in features such as temperature, luminosity, and mass (the gravity pressing in on a star is balanced by the pressure that is created by the nuclear reactions in its core, which in turn are conditioned on its internal temperature and mass). Depending on their interests and goals, scientists draw one set of boundaries to group stars into categories by temperature and luminosity, called the Morgan-Keenan system. The granularity (resolution) of the categories depends on what is most useful to scientists in a particular circumstance (Ruphy, 2010). Scientists also have a second system for categorizing stars by their mass, which is orthogonal to the Morgan-Keenan system. Scientists might group stars by mass when they are interested in the internal structure of energy flow in stars, which varies with mass. It's also interesting that a star's luminosity is easier to compute than its mass, given human sensory systems and our measurement and modeling tools, so perhaps it is unsurprising that the Morgan-Keenan system tends to dominate in astronomy. Given a particular classification system (i.e., a particular conceptual context), the categories predict and explain the phenomena of interest well (e.g., within the Morgan-Keenan classification of stars, it is possible to build knowledge about the Sun as a G-type star; for discussion, see Carnap's concept of volitional impenetrability; Carnap, 1950).

Many other competing taxonomies of relational kinds happily co-exist in the practice of science. In physics, light can be a wave or a particle. In chemistry, scientists classify elements by the number of protons they contain or by their nuclear properties, and historically, other equally viable categories have been created, depending on scientists' commitments and interests (Chang, 2012).

In biology, the concept of a biological species often (but not always) creates categories based on evolutionary lineage, but the features of equivalence and the resulting categories can vary (e.g., Dupré, 1999; Mayr, 1992; Wilkins, 2009). When the concept of a species is defined by the relations between genomic populations, scientists decide the granularity of categories to use (i.e., magnitude of genomic difference that is required to claim that speciation has occurred). In neuroscience, a brain is variously understood as a single, complex network, as a system of subnetworks or brain regions, or as groups of neurons and glial cells within different parts of brain regions, depending on the interests of scientists. In ecology, the categories for types of wetlands or forests exist in relation

to the time of year because the needs or interests of scientists change with seasonal variation of environmental features (Cadenasso et al., 2003; Sismondo & Chrisman, 2001). Recently, there has been a proposal to change the category CANCER to include only instances of uncontrolled cells that are life threatening and require treatment (versus those that merely increase risk of illness and require no treatment; Esserman & Eggener, 2023).

When mutual information emerges from interacting human brains in consensus, additional meaning emerges that cannot be found in the signals outside of human brains alone. This mutual information is the definition of a socially constructed category (for a discussion of a socially constructed category, see Hacking, 1999). Any category X is social reality when "X is not determined by the nature of things; it is not inevitable" (Hacking, 1999). Some features of events are conditioned on embodiment constraints (the physical facts of the body and brain, the way our eyes work, etc.). Some features are arbitrary when it comes to physical signals in the world, however. These features are invented by humans who collectively impose functions on things that they don't have solely by their physical nature, and because humans agree that they have this function, they do (Searle, 1995). Through consensus, humans create categories that only exist as social reality: MONEY, GOVERNMENT, STREET NAMES, and so on. Other categories of social reality include ANGER, SADNESS, FEAR, COGNITION, PERCEPTION, and so on (Barrett, 2009, 2012). Such categories constitute knowledge about the parts of reality for which those categories are functional. Not all cultures have the same concerns; cultures with different concerns (i.e., different niches) need different vocabularies of categories defined by different functional features (e.g., Mesquita, 2022; also see Appendix, Section 9).

The very core of traditional scientific realism posits a firm boundary between a scientist — what is inside their head — and the rest of the world. Evidence that supports relational realism, from physics to sociology, suggests that this dichotomy is false. Interestingly, people of different geographic cultures conceive of this boundary as more porous than do many in the West (e.g., Markus & Kitayama, 1991, 1994; Shweder & Bourne, 1982; for discussion, see Danziger, 1997).

A high dimensional space of concrete features with vectors on manifolds

The hypothetical *n*-dimensional feature space that depicted natural and empirical kinds earlier in the chapter can also be used to discuss relational kinds. To review, the dimensionality of this space for natural kind categories includes physical features in the world as well as mental features in a brain, including both concrete mental features (e.g., lines, edges) and abstract mental features that are compressed summaries of the concrete features (e.g., threat, value). Each point in the space is an object, process, or event that permanently belongs to a particular grouping of instances that share similar values on a feature or set of features (the intension of the category), so that the instances cluster very closely to one another in part of the space (called a subspace). It is assumed that "threat," for example, is an amodal abstract feature that can be detected as combinations of physical features in the world in much the same way that a brain detects concrete mental features from lines and edges in the world. The abstract and concrete features are, in effect, independent of one another. For empirical kinds, the dimensionality of the space includes only physical features in the world and concrete mental features that are perceptual and non-conceptual. Abstract mental features are out of bounds because they cannot be directly sensed.

For relational kinds, the space is the same as for empirical kinds — the dimensionality of the space includes only physical features in the world and concrete sensory and motor mental features. Every instance of anything that exists anywhere in space and time is an event represented as a point in

this hypothetical space. The coordinates at a given location, a sequence or tuple of *n* values, represent the extent to which an instance possesses each of the *n* features. Many more features will be relevant (i.e., will have non-zero values) than would be the case for natural and empirical kinds because events are contextual in both time and space. Features of high dimensionality such as lines and edges, glucose levels, vibrations, and so on, are continually compressed in the brain into lower dimensional abstract features such as "threat" or any other low-dimensional functional feature. Each abstract, functional feature is a compressed *multimodal* summary of high-dimensional features (rather than an amodal summary) and therefore is equivalent to a *vector* through the space. An array of high-dimensional features, their vectors and the geometric relation among the vectors corresponds to a specific spatiotemporal occurrence that has been categorized with an ad hoc, situated category. Abstract features, as compressed multimodal summaries, are computed in a brain — they do not exist in the world outside the brain. "Threat," "reward," "pleasant," "unpleasant," and so on, are probabilistic properties of the relations between interacting, higher dimensional concrete features and signals that are intrinsic to the brain.

Different instances of signal interactions will vary in their high dimensional features but can share the similar values on the lower dimensional vector. The instances will therefore cluster closely so that their proximity can be described in simple Euclidean terms (called a manifold) within the much larger *n*-dimensional feature space. (In math-speak, the vector lives on the manifold; in three dimensions, this would be equivalent to saying that the points live on the surface of a sphere.) One important implication here is that different relational events (interactions of different signal ensembles) can realize the same value of the same functional feature (i.e., have the same meaning). This is how lower-dimensional features of equivalence that cohere a category at one level of abstraction can give way to reveal structured variation at a lower level of abstraction (i.e., structured variation in higher-dimensional features). For example, there is more than one ensemble of neural signals, bodily signals and signals from the world, that give rise to situated instances of FEAR (e.g., see Barrett, 2017c, 2018; Hoemann, Khan, et al., 2020; LeBois et al., 2020; Le Mau et al., 2021; Siegel et al., 2018; Wilson-Mendenhall et al., 2011; Wang et al., 2022; Table 1).

When specific spatiotemporal signal ensembles are similar enough on some set of low-dimensional features (represented as vectors on a manifold), they form an equivalence class — a category — with inductive and deductive power and utility distributed across the entire manifold of vectors. Accordingly, an experience of one category instance should allow useful inferences about the others, even when the instances differ in their high-dimensional particulars. The distance in vector values corresponds to the capacity to generalize from prior instances of the category to a new instance.

Scientific Questions Are Relational Questions

When practicing relational realism, questions about the nature of reality, scientific or otherwise, cannot be phrased in contextless absolutes, because reality is not absolute, it's perspective dependent. If an instance of EMOTION, MEMORY, or any other psychological category is relationally real, then questions must always be posed and tested relationally. Absolute questions like "What is an emotion?" are meaningless or at least incomplete. Questions about the nature of the category FEAR, for example, are questions about the nature of FEAR from a particular perspective (Barrett, 2012, 2017b, 2017c, 2022). Rather than ask whether a fly is fearful in some absolute sense, scientists should ask whether FEAR exists in a fly *from the perspective of a human observer*, maybe from the perspective of a scientist who believes in natural kinds. Or they might ask whether the fly

is in a state of fear *from the fly's own perspective*. These are different questions that realistically can have different answers. The question "Who is right?" is not scientifically (or philosophically) meaningful, however. The reality of the situation is relational. A scientist can experience the fly as fearful because the architecture and scaling of their human brain is capable of constructing a concept of fear that abstracts and generalizes across different sensory and motor circumstances. The fly's brain is *not* equipped to generalize in this abstract way, as least far as scientists know, and so reality, as it pertains to the fly, will be different.

More generally, from a relational point of view, scientists cannot build justified knowledge about FEAR, MOTIVATION, IMPLICIT ATTITUDES, and other categories as if those categories exist independently from a perceiver in a particular context. Situations are not merely moderators of stable personal characteristics. Their signals help manifest the phenomenon. When a scientist is interested in whether implicit attitudes predict behavior, for example, scientific efforts should be tuned to testing hypotheses about attitudes in a *context-specific way* to measure *context-specific behavior* in *individual participants*. Then it becomes possible to *discover* generalities as they exist (also see Westlin et al., 2023). This strategy may sound impractical, but it is less problematic than developing knowledge claims using group statistics that do not necessarily (or even usually) manifest or predict anything about individual subjects (Clark-Polner et al., 2017; Dunlap, 1935; Estes, 1956; Flournoy et al., 2024; Gallistel, 2012).

Hypothesis Testing Is Relational

Relational realism calls into question a fundamental premise of the hypothetico-deductive method: An observer is fundamentally separate from the reality that is observed, thereby making observations useful for evaluating whether scientific concepts are justified as knowledge. As noted earlier, psychological science has a long tradition of deliberate attempts to wall off the observer from the observed. These attempts date back to the field's origins in the 19th century, when neurologists, physiologists, and even physicists took the claims of mental philosophers and exposed them to the rigors of experience, structured as experimental observations (e.g., Danziger, 1990). The observer and the observed were separated in the name of scientific objectivity, in an attempt to observe psychological events from a third person perspective (i.e., from the outside). A relational view of reality challenges this fundamental assumption. In a relational approach to psychological science, any factors that constitute the scientific "background" (i.e., so-called "hidden moderators") may actually help to create the phenomenon of interest. Let's consider some of these so-called background factors.

Hypotheses and predictions are features of relations. Those relations are hidden from view in standard experimental practice, however. A hypothesis exists in relation to auxiliary assumptions and additional details that are necessary to derive it from its premise, even when using the hypothetico-deductive method under ideal circumstances (when the premise is a scientific law). To remind, every scientific law is a theoretical kind, a falsehood that lacks the contextual details that are necessary to use it in real life. A prediction exists in relation to the additional details of design decisions, such as the circumstances in which the observations take place (a.k.a., "initial conditions") and the assumptions entailed in the measurements and other features of the study.

Hypotheses and predictions, along with all other aspects of scientific practice, are also concept dependent. No hypothesis, no prediction, no experimental setup or analysis is concept-free. The "facts of experience" — the observations used to evaluate hypotheses and predictions — are never fully separate from the inferences being tested. Everything that a scientist learns about reality by

exposing hypotheses to the rigors of experience is conditioned on the very concepts they are trying to test (Brewer, 2012; Dubova & Goldstone, 2023; Feyerabend, 1975/2010; Hanson, 1958; Kuhn, 1962/1970; Quine, 1951b). When concept-laden observations are investigated with concept-laden analyses and inferences, it creates a confirmatory stance that gives the appearance of consensus. Examples in the science of emotion are discussed in Barrett (2022) and Westlin et al., (2023). Also see Appendix, Section 4. Observations in science can never offer a truly neutral test of any concepts of interest. That's why it is important to test a hypothesis derived from one research tradition using the methods of another (e.g., Gendron et al., 2014a, 2014b, 2015; Gendron, Hoemann, et al., 2020; Hoemann et al., 2019; Hoemann et al., 2023).

Initial conditions, auxiliary assumptions, and scientists' concepts are all part of the larger context of signals that contribute to hypothesis testing, impacting scientific observations and the inferences that result from comparing those observations to the originating concept. How these contextual signals are handled depends on philosophical commitments. Everything a scientist does, from choosing a category of interest to interpreting the meaning of observations is conditioned on certain philosophical commitments and interpretive points of view (Ayer, 1936/1952; Carnap, 1937/2002 V.(b); Duhem, 1954/1991; Giere, 2010; Kuhn, 1962/1970; Massimi & McCoy, 2019; Quine, 1951b; Stanford, 2023; van Fraassen, 2004; also called "pragmatic" or "natural" realism, see Putnam, 1981). The result is a web of relations (Quine, 1951a, p. 72).

Justified knowledge about the world cannot be obtained by testing a hypothesis against observations in the traditional sense because any observation tests *the entire web*. Quine called this idea *holism*. (The contrast to holism is reductionism, also called atomism; for critiques of reductionism, see Hoyningen-Huene, 1989; Kitcher, 1984; K. R. Popper, 1974; also see Jackman, 2020). A hypothesis is always tested holistically; it only can predict what is observed within the context of the entire web. When unexpected observations emerge from an experiment, then the entire web is implicated. And when things go well, it's still not clear that the hypothesis is justified (see Appendix, Sections 2 and 4). This predicament poses a serious challenge to traditional scientific realism and empiricism in any form because all tests within the hypothetico-deductive method, and even in the weaker inductive-statistical method, are always under-determined by evidence (see Appendix, Section 11).

In the science of emotion, for example, scientists' concepts about emotion routinely dictate how they design experiments and how they make, analyze, and interpret their observations. Certain experimental setups and sampling of stimuli create observations that appear to be robust, replicable evidence for the hypothesis that ANGER, SADNESS, FEAR and several other Western folk emotion categories are universal (i.e., each with its own inherent pattern of evolutionarily prepared facial expressions, peripheral nervous system changes, and so on). Slightly tweak the experimental setup, however, and the observations change, dissolving evidence for universality. This has been shown again and again, particularly over the past couple of decades (for extended discussions, see Barrett, 2022; Barrett et al., 2019; Gendron et al., 2018; Russell, 1991a).

Relational realism, as a philosophy for science, offers some ways to deal with this conundrum. The simplest one: Don't conflate measurements and inferences. Scientists expose human participants to stimuli such as movies and photos and expose non-human animals to stimuli such as the scent or some other depiction of a predator. That these stimuli are "fear stimuli" requires an inference that might not be correct. Similarly, scientists observe physical signals (e.g., increases in heart rate), not threat or fear. The presence of threat or fear is inferred psychological meaning, constructed as mutual information with respect to other signals in scientists' brain. By more clearly separating

description and observation from inference, scientists will make it easier to evaluate their knowledge claims.

Again, an example from the science of emotion is instructive. In peer-reviewed journal articles, facial muscle movements are routinely referred to as "facial expressions" or "emotional expressions." Movements are observations. To call them "expressions" is an inference, a priori, that they emit or carry information about an internal state, emotion or otherwise. Ditto for referring to facial movements as emotional "displays" or describing a particular configuration of movements, like a scowl, by its inferred meaning ("anger expression" or "anger face"; for a discussion and additional examples, see Barrett et al., 2019). Notably, meta-analyses estimate that people living in urban settings scowl in anger only about 35% of the time (Durán & Fernández-Dols, 2021) and at least half the time people are scowling they are not angry (Le Mau et al., 2021). Decades of research in social psychology show that movements are automatically and effortlessly experienced in terms of inferences about the psychological meanings of those movements (e.g., Gilbert, 1998; Lieberman et al., 2002; E. R. Smith & DeCoster, 2000; Vallacher & Wegner, 1987). These mental inferences are examples of meaning-making, thanks to the human brain's continuous category construction. They can sometimes wreak havoc in one's personal life, but their impact in a scientific context is even more pernicious.

It's also helpful to explicitly pose hypotheses and predictions in relation to the larger web of beliefs and contextual details in which they are embedded, including the categories and philosophical assumptions being made, how these concepts manifest in experimental design, and their impact on subsequent inferences and knowledge claims. Adolphs and Anderson's approach (Adolphs & Anderson, 2018; D. J. Anderson & Adolphs, 2014) is a good start to such transparency. This sort of specificity is required to formalize hypotheses and predictions in mathematical language. Writing ideas in formal mathematical equations can be difficult, but grappling with that difficulty is useful in that the requirements of math, as a language, calls for making assumptions explicit.

Experiments

All experiments are simplifications of the complexity in nature. But how much simplification is required to test a hypothesis (i.e., where's the sweet spot)? When scientists search for natural kinds of psychological categories and design experiments accordingly, they reduce contextuality and complexity, making it difficult, if not impossible, to discover with any degree of confidence that their ontological assumption may be in error and that reality is more relational. Failures to reproduce or replicate observations will be perceived as error or sloppy methods. There is no opportunity to discover and model structure in variation because the experiment was not structured to cultivate and observe variation. On the other hand, efforts aimed at relational kinds put contextuality and complexity front and center to discover (rather than prescribe) the optimal level of simplification. Experiments can be designed to identify, measure, or manipulate the fuller web of signals that give rise to category instances, to take account of the full context, or as much of the context as can be managed. This means observing variation and attempting to analyze and model its organization or structure rather than using experimental control to limit variation and risk removing part of the phenomenon itself. Obviously, this approach is much more complicated, expensive, and painful than assuming instances of any psychological category can be studied in a contextless, universal way. But a relational realist would suggest that, if the goal is to build justified psychological knowledge that can be used in the real world to help real people, then what is the alternative?

The price of oversimplifying the experimental context is high. Effect sizes in psychology experiments rarely surpass .30 (Funder & Ozer, 2019; Mischel, 1968; Nisbett, 1980), which, for a traditional scientific realist attempting to build "true knowledge," is arguably an embarrassment. Such modest effect sizes may, in part, be a consequence of mistakenly modeling complex phenomena as simple, linear mechanistic systems (as discussed in Barrett, 2022). Cultivating structured variation in experiments — embracing complexity, rather than controlling it away — may be a more realistic approach for building justified scientific knowledge in an inherently complex world full of inherently complex brains and bodies. Even Popper eventually accepted this. "As a philosophy, reductionism is a failure," he wrote. "We live in a universe of emergent novelty; of a novelty which, as a rule, is not completely reducible to any of the preceding stages" (K. R. Popper, 1974, pp. 269, 281; as cited in Mayr, 2004, p. 79). Sometimes controlled experiments might be necessary, however. In such cases, it is important to remember that an effect size observed in a specific experiment is a function of many causal influences, some of which have values that were held constant in the experimental set-up. These influences should not be ignored when effect sizes are interpreted. Holism and complex causation can make their appearance during inference if not during experimentation.

Participants

When practicing relational realism in psychological science, it is important formulate and test hypotheses as they pertain to *individuals* rather than the mythical "average person" (i.e., people in aggregate), as is routinely done in psychophysics, idiographic studies of emotion categories (e.g., Ceulemans et al., 2012; Flournoy et al., 2024; Hoemann, Wu et al., 2020; Kuppens et al., 2003; Singh et al., 2021), and studies of trajectories of change (e.g., Bonanno et al., 2024). This choice is significant from a philosophy of science perspective. Early in the history of psychological science, researchers attempted to understand mental categories by combining observations from all people in a sample of participants (Danziger, 1990, p. 81 also see Ch. 4). Aggregation was considered a quick and efficient solution to the tremendous variation that was routinely observed across people and situations. According to historian Kurt Danziger,

> the use of average data seemed to provide an acceptable way of coping with the utter lack of consistency that was characteristic of complex individual behavior … if data from many individuals were pooled, certain statistical regularities sometimes emerged. These regularities were to form the basis of psychological generalizations, even though the average pattern might not correspond to the actual behavior of a single individual member of the statistical group. (Danziger, 1990, p. 153; also see Bakan, 1967; Estes, 1956; Flournoy et al., 2024; Gallistel, 2012; Sidman, 1960; Westlin et al., 2023).

For example, if people in a sample scowl when angry on average 35% of the time (which is above what would be observed by chance), it does not follow that each person (or any person) scowled at above chance levels across "anger trials" within the experiment. Some people in the same may not have scowled at all. Nor does follow that people in a second group should scowl at above chance levels. Nonetheless, this is exactly what psychologists predict when they attempt to repeat the experiment with another sample. Similarly, when a unique pattern of brain activity is identified for a group of people experiencing anger, it does not follow that most people or even *any* single person in that group evidenced that brain pattern (e.g., Clark-Polner et al., 2017). Nor does it imply that this particular pattern of brain activity will replicate for different samples of ANGER (and, in fact, this is what the evidence suggests again and again (see Westlin et al., 2023). Likewise, if a group of people

receives training in an emotion regulation strategy called cognitive reappraisal, which is associated with increased support for policies that reduce political conflict (e.g., Halperin et al., 2013), then it follows that the average person in the sampled population would be similarly affected by the training. It does not follow, however, that any individual in the sample experienced a change in behavior (Danziger, 1990, Ch. 4; also see McManus et al., 2023).

In traditional scientific realism, the "average person" is real and variation around the average is error. This mindset was introduced earlier in the chapter as *typological thinking*. Statistical abstractions that meet some threshold of significance are taken as evidence about psychological reality. The hope is that these abstractions will eventually reveal clear evidence of psychological laws that govern human behavior (also see Danziger, 1990, p. 129). If aggregate data provide evidence of psychological laws that hold across individuals and situations, then it is unnecessary (and even ill-advised) to study individual people. A traditional scientific realist believes that the central tendency of a phenotypic distribution — the "type" — is real and variation is error.

The alternative view, in which variation across individuals is real, and the "type" or statistical abstraction (e.g., the "average animal"; Dunlap, 1935) is a simplification or fiction (Sidman, 1960, p. 53), was introduced earlier as *population thinking*. The evolutionary biologist Ernst Mayr attributed population thinking to Darwin (Mayr, 2004; for discussion see Barrett, 2017b, 2017c; Westlin et al., 2023 and references therein). Before Darwin wrote *On the Origin of Species*, each animal species was assumed to be a natural kind category, each unchanging and sharply distinguished from one another, containing identical individuals (with an essence). Phenotypic variation within a species was assumed to reflect error — accidents or imperfect manifestations of the category essence. Population thinking, on display in *Origin*, suggests that variation among individuals within a species (i.e., among instances of a category) is meaningful, arising from various sources (in modern science, sources of variation include mutation and degeneracy). Some of this phenotypic variation is a better fit to immediate environmental conditions, making it easier for certain animals to thrive and successfully reproduce, which continually shapes the distribution of phenotypes within a species over time.

Population thinking, with its attendant emphasis on complexity and structured variation, is more consistent with relational realism and the study of relational kinds. Variation is real and the type is a statistical idealization or stereotype.

Modeling

Say goodbye to the t-tests, ANOVAs, and regression models of linear, deterministic causation. Say hello to modeling that focuses on complex, dynamic causality. A full exposition of complexity theory and complex dynamic systems is beyond the scope of this chapter, but interesting and informative discussions can be found in virtually every scientific field, including psychology (Byrne & Callaghan, 2022; Gernigon et al., 2023; Favela, 2020; Tilak et al., 2022; for a quick tutorial, see Systems Innovation, 2017). Versions of complexity or complex dynamical systems have been discussed as concepts for understanding development, motor movement, perception, social relations, psychopathology, concepts, consciousness, and basic brain functions (for discussions and additional references, see, e.g., Bassett & Gazzaniga, 2011; Breakspear & McIntosh, 2011; Bressler & McIntosh, 2007; Buzsáki, 2019; Cisek, 2019; Eidelson, 1997; Fausto-Sterling, 2021; Gergen, 1978; Heft, 2001; Kelso, 2012; Kelso et al., 2013; Kirchhoff, 2018; Maturana & Varela, 1972; Rabinovich et al., 2015; Rubinov & Sporns, 2011; L. B. Smith et al., 2018; L. B. Smith & Thelen, 2003; Sporns, 2011; Vallacher et al., 2015; Wright & Woods, 2020; Zelazo, 2013). In these works, psychological

phenomena are events that emerge from dynamic relations among a multitude of influences that, in isolation, are only weakly causal.

Why is any of this important? Because the scientific study of relational kinds is best achieved using more holistic accounts of causation, such as constraint causality (Juarrero, 2023) or process causality (Gernigon et al., 2023). A complex system cannot be reduced to independent parts, each with its own inherent features. A complex system cannot enter into simple cause-effect sequences that are merely tweaked or otherwise moderated by context (i.e., linear determinism of the machine model of causation, also what Aristotle referred to as efficient causation). Epistemically, reductionism and simple mechanistic notions of causation are suitable for investigating natural kinds or empirical kinds, but not relational kinds.

Inferential Logic

Say goodbye to the hypothetico-deductive method. The rationality of the hypothetico-deductive method is premised on the assumption that reality is conventional, with unchanging, universal kinds governed by simple deterministic laws or at least highly probable regularities. Only in such circumstances are observations of low probability sufficient grounds to conclusively disconfirm a hypothesis. (Popper, for example, argued that if a hypothesis predicts specific observations at low probabilities but does not rule them out, then the hypothesis is not scientific.) But by now it should be clear that scientific laws are simplifications that require a secret ingredient to be useful: the contextual details that make the laws complex and relational.

Furthermore, in a relational reality of situated, complex systems of signals, where there are no laws but only contextually conditioned constraints, observations of low probability (i.e., exceptions or anomalies from what is predicted) are to be expected from time to time and therefore can never fully falsify a hypothesis. Any given event can be more or less likely in a given situation. "Acceptance" and "rejection" must be replaced with probability of occurrence. Science becomes more about quantifying confidence and doubt using, for example, Bayesian estimation procedures rather than accepting or rejecting hypotheses using frequentist statistics. Reality is more variable than natural and empirical kind concepts can deal with. For a traditional scientific realist, particularly one who behaves in accordance with Popper's view, hypotheses are no longer falsifiable. This assumption is not unscientific. It requires a different approach to science.

All Knowledge Is Relational

Knowledge claims always exist from a particular point of view. This philosophical approach was introduced earlier in the chapter as *scientific perspectivism* (Giere, 2010). To "know" is a momentary event, as signals in a brain interact with and constrain signals originating outside a brain. From a relational point of view, knowledge is not permanent but is a regularity that emerges from the categories that a brain constructs and the experiences (observations) that result. Not every perspective is equally valid and useful for every scientific question in any domain, and usually more than one perspective fits the bill, which is called *scientific pluralism*. Some philosophers use perspectivism to justify pluralistic realism (e.g., Massimi & McCoy, 2019; Putnam, 1981), which may be just another name for relational realism.

Options for different systems of categories imply choice. Choice is always linked to preference. Preference could be rooted in practicality, such as ease of measurement, or in the scale of inquiry, or

in more personal factors like interest, values, or whatever is easiest to publish. When scientists choose one ontology over another, they prioritize certain features and ignore others when creating groupings of equivalent events (this phenomenon was introduced as "boundary construction" earlier in the discussion of relational kinds). Such category construction reduces complexity, foregrounding some features as relevant while backgrounding others as noise, allowing prediction, action, and explanation in particular situations. Every categorization is a reduction of complexity, a simplification of some sort. Fail to simplify enough and the result is too much uncertainty to act in the world at large. Simplify too much and the result is a fiction that cannot perform in the world at large. Scientists must find the sweet spot for a given hypothesis. The granularity of a scientist's categories must be useful for whatever knowledge claims the scientist is trying to build.

As noted already, optimal meaning-making requires finding the most useful balance between complexity, which provides accuracy and precision, and simplicity, which provides generality but hides contextual variation (for a similar point, see Levins, 1966, 1968). If a simplification improves repeatability and predictability (outside a simplified experimental setting), then the category in question might be a natural kind. If simplification reduces repeatability and predictability, however, then it might be a relational kind.

Consider a popular simplification used in psychological research: the choice-from-array response method. When building an experiment, participants must have some way to respond to the stimuli that they are exposed to. A common method is to have participants choose a response from among an experimenter-selected array of options. This response method produces highly replicable results in controlled experiments but remove the simplification and the robust replication of observations breaks down (see Appendix, Section 4; also see Rice, 2013, 2016). Choice-from-array favors repeatable observations in an experimental setting (internal validity) over the meaning of those observations in the real world (external validity). This preference for optimizing one at the expense of the other is not unique in psychological science, but a useable science of psychology requires reversing that preference (Miller et al., 2019), even if the result is publishing fewer peer-reviewed empirical papers. This sounds like another philosophical approach called *pragmatism*, but it's not (see Appendix, Section 6). It's situated functionalism, which is relational realism.

Knowledge claims are also relative to the time and the culture in which they are made and embedded. When reading the works of famous scholars such as Darwin, for example, scientists must remember to view them as *historical texts*, not as definitive, contemporary scientific guides. For example, all of Darwin's books include concepts that are no longer current (this is more of an issue for *The Descent of Man* and *The Expression of Emotions in Man and Animals* than *On the Origin of Species*; Browne, 2007, 2010; for examples, see DeSilva, 2021). Furthermore, Darwin used different concepts than today's scientists do (even when the words are similar). Darwin's works contain important ideas that may speak to scientists today, but it is a mistake to infuse his work with 150 years of subsequent thought to conclude that Darwin was stating contemporary justified knowledge. In history, this mistake is called *presentism*. Darwin's books and papers, like the writings of many other influential scientists throughout the ages, are inspiring, important, and insightful as historical works.

The relevance of historicizing scientific work is no less important when considering the knowledge claims of modern scientists who publish over several decades. The theory of constructed emotion, for example, which is detailed in peer-reviewed papers and books from 2017 onwards, uses a broader multidisciplinary foundation, different conceptual tools, and hypotheses at different levels of analysis than its predecessor, the conceptual act theory, which was first published in 2006. Yet

the two are often presented as interchangeable rather than historically contingent, spanning almost 20 years in which concepts, empirical findings, and multidisciplinary context have changed (for discussion and examples, see Barrett & Lida, 2024).

Thomas Kuhn and the social construction of science

A key contribution to the philosophy of science as relationally constructed came from Thomas Kuhn. In his famous work, *The Structure of Scientific Revolutions* (1962/1970), Kuhn analyzed examples from the history of science to describe scientific practice as a series of cycles: a period of cumulative puzzle-solving or normal science, called a *scientific paradigm*, followed by a period of *scientific revolution* or *paradigm shift*, leading to a new normal. And then the cycle begins again. A scientific paradigm is a collectively endorsed way of doing science, a particular set of scientific lenses for experiencing and interacting with the world. It prescribes which phenomena exist (ontological claims) as well as the methods for building reliable knowledge about those phenomena (epistemological claims). Correspondingly, a paradigm shift involves more than just a change in scientific methods. It is also a revolution of ontology. Categories that were vitally important to the older paradigm may disappear and new ones may take center stage to be studied and understood. Kuhn wrote that the old and new conceptual systems were incommensurable (also see Feyerabend, 1975/2010) although he retreated from that strong view in later writing (Kuhn, 1979, 2002). Incommensurable paradigms not only have a wholly different view of the world, but also different philosophies about what counts as good scientific practice, as knowledge, and so forth.

One of the many important implications of Kuhn's *Structure* is that some knowledge claims are anointed as "Scientific Facts" while others are ignored, and the choice depends on the philosophical assumptions of the current scientific paradigm. His point, reinforced by others who expanded on the themes in *Structure,* was that epistemic strategies alone are insufficient to understand how science works because those strategies are carried out in a socially organized way and subject to incentive structures and other pressures (e.g., Bloor, 1991; Collins, 1992; Hacking, 1983; D. L. Hull, 1988; Latour & Woolgar, 1979/1986; Laudan, 1977, 1990; Longino, 1990, 2001; for a subtler view, see Kuhn, 1977; for a general review, see Godfrey-Smith, 2021; and see Appendix, Section 12). Facts are made, not found (Shapin & Schaffer, 1985/2011). Single investigators can create knowledge claims in the privacy of their own labs (assuming they can find funding), but their claims become knowledge only with the consent of other scientists. Simply put, science is a socially constructed practice — a set of activities to produce something that counts as scientific knowledge by a community of scientists under certain social conditions. The categories and scientific knowledge of reality depend on shared concepts, goals, values, and agreements about which phenomena exist (i.e., which ontology to use), which questions are admissible, and which methods count as acceptable tools of inquiry.

Kuhn was not offering a broad critique of science as a way of knowing, per se. He endorsed some of the customary virtues of science, like hypotheses that predict well, organize observations coherently, are internally consistent, broad in scope, rich in consequences, and generate new discoveries (Godfrey-Smith, 2021). *Structure* was more of a critique of traditional scientific realism. Scientific truth — the knowledge claims that scientists treat as justified knowledge — is a social product arising from the consensus of scientists at a particular time and place and conditioned on a whole set of philosophical commitments. That is, what counts as scientific truth is always relative. But again, not infinitely so. The physical world necessarily constrains knowledge claims in important ways but is alone insufficient to determine the truth value of those claims.

Kuhn was not claiming that science is biased or irrational, but that the rationality of science derives from its social nature. Every way of scientifically knowing the world depends on the actions of a group of humans in agreement. The social structure of science and the social practices of scientists impact how knowledge is generated and what counts as knowledge in the first place. "Scientists are not some special breed of pure, disinterested thinkers who receive private bulletins from Nature," writes the philosopher Peter Godfrey-Smith. Instead, scientific facts are collective accomplishments — or mistakes — of entire scientific communities.

Scientific knowledge is therefore fundamentally relational. It exists as collective organized experience (Bogdanov, 2020, p. 113), mutual and relative information between the brains of like-minded individuals who temporarily reify shared knowledge claims that are perspective dependent. Relative information allows scientists with shared concepts to experience a freezing animal as fearful. Scientists who do not share these same concepts cannot claim that this knowledge is wrong, but only that the knowledge exists in relation to a perspective that is not universal and therefore may have little utility outside the laboratory settings in which that knowledge was cultivated.

Objectivity is relational

Does the relationality of knowledge mean that scientists must throw objectivity out the window? Well, that depends on their philosophy of science. In the relational view of reality, objectivity is the minimization of bias, which is achievable and can be found in the social nature of scientific activities. Relative information between brains and the world becomes knowledge and stands in for objectivity under certain conditions. Relational realism is seen as unbridled relativism only in an individualistic account of science. "Values are not incompatible with objectivity, but objectivity [is] a function of community practices rather than as an attitude of individual researchers" (Longino, 1990, p. 216). The enterprise of science can be objective even if individual scientists cannot be. Only vetted knowledge claims become knowledge. The social practices of science do not corrupt rationality — they are sources of rationality in science (see Appendix, Section 7).

There are three important implications here. First, when science becomes too individualistic, indulging too much in the "great man" theory of discovery (rewarding individuals with grant money, awards, etc.), it creates pressures that can compromise the credibility and integrity of the enterprise and its products.

Second, trustworthy scientific knowledge requires more than replicability. It flourishes when a scientific community is sufficiently diverse (in people and in methods) and open to criticism. Is a field dominated by a small clique or a few aggressive individuals (Oreskes, 2019)? Are potential critiques being suppressed by asking about critics' personal motives? These are signals that certain knowledge claims are suspect, even when observations easily reproduce and are replicated (again, see Appendix, Section 4).

Third, when a field fails to come to consensus, the entire community must rectify the situation. In the science of emotion, much scientific work proceeds in theoretical silos supported by incommensurable philosophies (Barrett, 2015; Barrett & Lida, 2024). Researchers from different camps consistently misunderstand one another and mischaracterize each other's hypotheses in fundamental ways. At times, scientists can't agree on the specific experiments that might resolve their debates, in part because they disagree on what they are actually debating about. Attempts to summarize relevant research often do so selectively, sometimes failing to grapple with evidence

that calls their preferred hypotheses into question. There are notable exceptions, of course, but overall, the science of emotion remains mired in a scientific stalemate: its research silos compromise the accumulation of knowledge, and an emerging consensus is nowhere in sight. The linguist George Lakoff called emotion an *essentially contested concept* (Lakoff, 2016): Scientists agree that phenomena known as emotions should be studied, but after about 150 years of scholarly activity, that is about all they agree on.

A social view of objectivity suggests that when a field fails to be self-critical, then the entire community is culpable. An example, again from the science of emotion, is presented in Appendix, Section 4. Unfortunately, other forces can limit self-criticism, like competition for limited resources, a dependence on expensive equipment, and a reward system that is rooted in reputation and recognition (also see D. L. Hull, 1988), which is selective by design and bestowed upon a chosen few (usually) after years of toiling. As a result, scientists may become attuned to a special kind of ambition and self-interest, inviting more skepticism of others' knowledge claims than of their own and vigilance as to whether others use their claims with credit. The result is an obsession with publication (Merton, 1979) or what the philosopher Peter Godfrey-Smith refers to as a "mania to publish." Oreskes puts it beautifully,

> objectivity is likely to be maximized when there are recognized and robust avenues for criticism… when the community is open, non-defensive, and responsive to criticism, and when the community is sufficiently diverse that a broad range of views can be developed, heard and appropriately considered (Oreskes, 2019, p. 53).

Critiques Of Relational Realism

Relational realism runs counter to commonsense. The common criticisms of relational realism reflect its counterintuitive nature:

1. Relational realism denies the existence of reality and true knowledge about reality. *Response*: This is a caricature. Reality exists in relational to some perspective, as does true knowledge. This is not the same as claiming that reality and truth are a simple matter of belief.

2. Relational realism is irrational, and rationality is the best guide to objectivity truth. *Response*: Rationality is relational. It is always relative to a goal and there are competing goals in science (see Appendix, Section 12).

3. Relational realism is anti-scientific. It reduces science to merely one way of justifying knowledge among many. *Response*: Reality and knowledge about reality are perspectival. The existence of more than one perspective does not imply that all perspectives are equally justified. Research guided by relational realism is subject to the same epistemic standards as traditional scientific realism (i.e., coherence with other domains of science, predictive power, and explanatory power; Kuhn, 1977).

4. Relational realism is complex. Simplicity is a virtue. *Response*: Simplicity is only one epistemic standard among many, and it's a virtue only when the phenomena of interest are simple. Other epistemic standards are incompatible with simplicity (Magnus, 2000).

5. The study of relational kinds is impractical. It would require a major shift in the practice of psychological science. *Response*: Yes and no. Sure, experiments might be more complex and more

expensive to conduct. On the other hand, even the results of a tightly controlled lab experiment can be understood using relational realism. The subsequent inferences will be more incremental however, which could slow the publication of peer-reviewed papers. Either way, relational realism requires a shift in the incentive structure that governs psychological science (Barrett, 2019a, 2019b), in how psychological research is funded (see Appendix, Section 12), and in how students are trained. What might psychological science look like today if decades earlier James's radical empiricism or Lewin's field theory had been understood as relational realism?

A Final Philosophical Note: Natural Kinds As Theoretical Kinds

In the relational view of reality, scientific categories emerge as interactions between the signals in a scientist's brain and the signals in the periphery, such as those cultivated in an experiment. The intension of every relational kind category is partly determined by the scientist who is doing the categorizing, and so are its extensions. Intensions are also contextually variable and flexible rather than fixed and universal. Relational kinds are not infinitely malleable, of course, but neither are they fixed and universal. Natural kind and even empirical kind categories, from the perspective of relational realism, are idealizations or simplifications that do not exist in reality and arise when contextual relations are ignored or hidden and therefore taken for granted. They are theoretical kinds. In fact, many of the categories that have long been extolled as the best examples of natural kinds — chemical elements of the periodic table and molecules like H_2O (water) — are theoretical kinds (for discussion, see Appendix, Section 5).

An example of a theoretical kind is an ideal gas in chemistry. To understand the internal structure of a star, which is basically a big ball of gas, astronomers must predict the behavior of moving gas molecules with complex behavior. To simplify their task and make it tractable, astronomers strip away some complexity and model a fictional substance called an "ideal gas" composed of randomly moving molecules that bounce off one another with no friction or other meaningful interaction. This simplification makes scientific inquiry more tractable. The consequence, of course, is that observations do not generalize beyond the experiments to a reality in which gases are not "ideal." Is the psychological concept of a "basic emotion" like the concept of an "ideal gas," both of which strip away real-world complexity in favor of simplicity? Both ideal gases and so-called basic emotions are the result of a delicate balancing act between simplification and utility. How to best perform this balancing act — what is considered useful, how much error to tolerate — is a manifestation of a person's philosophy of science.

Elements of the periodic table are often presented as the best example of natural kind categories in science because the categories have firm boundaries in nature, defined by the number of protons in each element's nucleus, corresponding to its atomic number. The number of protons is insufficient to predict and explain an element's behavior in the physical world, however, and therefore cannot be considered a category's essential feature (see Appendix, Section 5). A scientist also needs additional contextual features to precisely predict or explain the physical properties of any sample of any element in the real world. Some of these contextual features are internal to the element, such as the state of excitement of its electrons, protons, or neutrons. Others are external to the element, such as the surrounding temperature and pressure. Chemical elements, therefore, are not natural kinds but theoretical kinds. If chemists designed experiments based on the idealizations in the periodic table, treating contextual features as mere moderators of universal phenomena, they'd experience *failures of replication on a massive scale*.

Strip away context from reality, reduce too much of the complexity, and the residue that remains is not a natural kind but a theoretical kind: a simplification of relations, an idealized fiction. To mistake a theoretical kind for a natural kind by ignoring context and complexity is known as *naïve realism*. If a scientist enacts naïve realism and designs experiments to cultivate idealizations, mistaking them for something real, they control variance and complexity instead of modeling it. They ignore all but one or two contextual factors, treating them as mere moderators of universal phenomena. The result will be experimental observations that are unlikely to be reproduced or replicated in any other experiment.

Summary

Relational realism and its attendant assumption of relational kinds have deep roots in philosophy and are practiced in every modern science, including psychological science, but often go by another name or remain nameless. Relational views of reality and science are based on a belief that the objects, processes, and events of reality emerge from interacting signals that on their own have no inherent meaning. Signals take on meaning as they interact and constrain one another, forming momentary complex systems that are inherently contextual. In science, relational kind categories always include the signals from a perceiver and are constructed in the moment. The intension of every category is relational and situated, partly determined by the goals of the perceiver who is doing the categorizing. Extensions are contextually variable and flexible. These qualities distinguish relational kinds from natural and empirical kinds, which are assumed to be fixed, enduring, and mind independent. Experiments that are designed to search for natural kinds or empirical kinds are unlikely to discover relational kinds because the experiments strip away context and complexity as a first principle, whereas experiments that assume relational kinds in all their complexity can conceivably discover natural or empirical kinds if they exist.

Relational realism assumes that scientific knowledge is perspective dependent. It acknowledges the social factors that influence what scientists choose to study (the phenomena that are admissible as topics of investigation), the questions they ask (about those topics), the methods they use to answer those questions, the sorts of observations that count as evidence, and so on. If a scientist believes that socially agreed-upon human inferences (as signal ensembles in human brains) transform systems of interacting energy signals into the reality of ELECTRONs, bits of interacting nucleic acid into the reality of DNA, certain rocks in orbit around the sun into the reality of PLANETs, certain changes in physical movements into the reality of FEAR, and so on, then that scientist is thinking like Kuhn and other scholars who view scientific knowledge as social reality. They acknowledge that human consensus transforms certain knowledge claims into scientific knowledge and anoints certain scientists as authorities under certain conditions. They understand that these practices function like norms, and they are often implicit habits of behavior rather than explicit rules (Godfrey-Smith, 2021). And they understand that when enough scientists withdraw from that consensus, reality changes.

CONCLUSIONS

Here are the main takeaways of this chapter:

1. Every scientist practices a philosophy of science with a web of concepts, call it a mindset, that includes beliefs about the nature of reality (i.e., ontological assumptions) and beliefs about the

nature of justified scientific knowledge (i.e., epistemic assumptions). These concepts are tools for doing science. A scientist's actions, not their stated beliefs, reveal their philosophy of science. These actions include who they read, who they cite, and who they ignore. Scientists publish knowledge claims. Other scientists make them knowledge. Or not.

2. A quick review of the practices and products in psychology suggests that many scientists are practicing a philosophy called traditional scientific realism, even though they may not realize it. Traditional scientific realism is the belief that reality exists independently of human concepts and is structured as natural kind categories, either as immutable, universal groupings with essences (classical categories) or highly probable groupings with recurring property clusters (prototype categories). Natural kinds are assumed to be largely independent of context and mostly governed by simple, mechanistic causes.

3. The incentive structure of psychological science is optimized for traditional scientific realism and the study of natural kinds. Scientific questions, experimental designs, publication attempts, and funding from granting agencies all go more smoothly for those who practice traditional scientific realism. A focus on natural kinds is better suited to the "mania to publish" that currently grips the field (Barrett, 2019b).

4. This focus on natural kinds makes psychological science unique because natural kinds do not seem to exist in other scientific fields, even physics. Chemical elements, molecules like H20, animal species, and even electrons do not meet the criteria for natural kind categories. What is the likelihood that any psychological category, including those discussed in the rest of this Handbook, are natural kinds?

5. There are philosophical and scientific reasons for questioning the utility of the hypothetico-deductive method and its weaker cousin, the inductive-statistical method.

6. Relational realism was introduced as a philosophy for practicing science. It is fundamentally at odds with the idea that reality is structured as fixed, eternal, and simple natural and empirical kinds. Instead, a relational view understands reality as ensembles of emergent, situated relational kind categories — dynamic, situated events rather than fixed natural kind or empirical kind entities with unchanging intensions and extensions. ELECTRONS, HYDROGEN, and FEAR, or even concrete categories such as BIRD or CAT, are understood not as single categories but as populations of situated categories which are created in the moment to serve a particular function or goal. The questions, methods and products of science are as relational and perspective dependent as the reality being studied.

7. To assume that reality is inherently relational is not anti-realism. The assumption that humans participate in creating reality is a form of realism called relational realism that is distinct from traditional realism and from versions of anti-realism such as empiricism and idealism.

8. Research guided by relational realism contradicts deeply embedded assumptions about the fundamental nature of reality, so it is harder for reviewers to understand (and may require a broader background to do so), is more expensive and slower to conduct, requires deeper sampling of individual participants across contexts, and therefore will result in fewer publications. All of which implies: A change in the incentive structure of psychological science is necessary to improve its scientific credibility.

9. Pluralism is possible with more complexity-based scientific methods. While more challenging to implement, such methods are likely to bear more fruit. And there is a bonus: Experiments that are designed to build knowledge about relational kinds make it possible to discover the existence of natural kinds if they exist. The reverse is not true.

10. Knowledge always requires simplification. Every scientific law is a simplification. So, too, is every model, every hypothesis, every experiment, and every category. Even abstract features of instances are simplifications of a sort. Find a sweet spot of granularity between complexity and simplicity — between uncertainty and absolutism — that is optimal to provide justified knowledge for acting in and on the world. There might be more than one sweet spot, depending on the goal at hand.

11. Scientists should avoid believing in the reality of simplifications they create. If reality is complex, then to believe in one's own simplifications is naïve realism.

12. Science is a set of collective human activities that operate in nested social contexts, whose meaning and validity is based on social consensus and therefore is subject to social (non-epistemic) forces. This does not weaken the rationality of science but may be its ultimate source.

13. The trustworthiness of scientific knowledge emerges from more than replicability. Trustworthy knowledge flourishes when a scientific community is sufficiently diverse (in people, ideas and methods) and open to criticism. Trustworthiness suffers with uniformity or individualism. When a field mistakenly comes to consensus and is unable to come to consensus for decades, the entire community is responsible.

14. Don't read Darwin or any other scholar as if their work is contemporary. This holds for all scientific works, even psychological science published in the last several decades. What counts as knowledge is determined by more than the work of individual scholars in their individual labs. For better and for worse.

15. Those who question the existence of the "replication crisis" or "credibility crisis" are not necessarily bad scientists. To question the existence of the "replication crisis" or "credibility crisis" in psychological science is philosophically sound. What appear to be scientific errors are often philosophical errors in disguise. Everyone agrees that questionable research practices and sloppiness are unacceptable. After these bad methodological habits are curtailed, however, psychological science must still grapple with modest effect sizes and highly variable observations from study to study. What appears at first to be a crisis of replication or credibility in psychology may be better understood as a much more serious crisis of an altogether different sort: naïve realism about the nature of reality, including the fundamental nature of a human mind.

AUTHOR NOTE

Many thanks to Karen Quigley, Tsiona Lida, Dana Brooks, Sarah Brown, Misha Pavel, Yorgos Stratis, David Melnikoff, Kay Shepherd, and Van Yang for comments on earlier drafts of this chapter. This work was supported by grants from the National Science Foundation (BCS 1947972), the National Institutes of Health (R01 AG071173, R01 MH109464, R01 MH113234, and R21 MH129902), the U.S. Army Research Institute for the Behavioral and Social Sciences (W911NF-16-1-019), and the Unlikely Collaborators Foundation. The views, opinions, and/or findings contained in this

manuscript are those of the authors and shall not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documents, nor do they necessarily reflect the views of the Unlikely Collaborators Foundation.

REFERENCES

Adolphs, R., Mlodinow, L., & Barrett, L. F. (2019). What is an emotion? *Current Biology*, *29*(20), R1060–R1064. https://doi.org/10.1016/j.cub.2019.09.008

Allen, V. L. (1965). Situational factors in conformity. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 133–175). Academic Press. https://doi.org/10.1016/S0065-2601(08)60105-7

Allport, F. H. (1919). Behavior and experiment in social psychology. *The Journal of Abnormal Psychology*, *14*(5), 297–306. https://doi.org/10.1037/h0073020

Allport, F. H. (1924). *Social psychology*. Houghton Mifflin.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 798–844), Clark University Press.

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Anderson, C. A. (1989). Temperature and aggression: Ubiquitous effects of heat on occurrence of human violence. *Psychological Bulletin*, *106*(1), 74–96. https://doi.org/10.1037/0033-2909.106.1.74

Anderson, D. J. (2022). *The nature of the beast: How emotions guide us*. Basic Books.

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell*, *157*(1), 187–200. https://doi.org/10.1016/j.cell.2014.03.003

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. MIT Press.

Arnold, M. B. (1960a). *Emotion and personality* (Vol. 1). Columbia University Press.

Arnold, M. B. (1960b). *Emotion and personality* (Vol. 2). Columbia University Press.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Asch, S. E. (1962). *Social psychology* (7th ed.). Prentice-Hall. (Original work published 1952)

Ashton, N. A. (2020). Scientific perspectives, feminist standpoints, and non-silly relativism. In A.-M. Crețu & M. Massimi (Eds.), *Knowledge from a human point of view* (pp. 71–85). Springer International Publishing. https://doi.org/10.1007/978-3-030-270414_5

Atzil, S., Gao, W., Fradkin, I., & Barrett, L. F. (2018). Growing a social brain. *Nature Human Behaviour*, *2*(9), 624–636. https://doi.org/10.1038/s41562-018-0384-6

Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Theories of emotion* (pp. 305–339). Academic Press. https://doi.org/10.1016/B978-012-558701-3.50018-1

Ayer, A. J. (1952). *Language, truth and logic* (2nd ed.). Dover Publications. (Original work published 1936)

Azari, B., Westlin, C., Satpute, A. B., Hutchinson, J. B., Kragel, P. A., Hoemann, K., Khan, Z., Wormwood, J. B., Quigley, K. S., Erdogmus, D., Dy, J., Brooks, D. H., & Barrett, L. F. (2020). Comparing supervised and unsupervised approaches to emotion categorization in the human brain, body, and subjective experience. *Scientific Reports*, *10*(1), 20284. https://doi.org/10.1038/s41598-020-77117-8

Azzalini, D., Rebollo, I., & Tallon-Baudry, C. (2019). Visceral signals shape brain dynamics and cognition. *Trends in Cognitive Sciences*, *23*(6), 488–509. https://doi.org/10.1016/j.tics.2019.03.007

Baghramian, M., & Carter, J. A. (2022). Relativism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2022/entries/relativism/

Bakan, D. (1967). *On Method: Toward a Reconstruction of Psychological Investigation*. Jossey-Bass.

Ball, P. (2024). *How life works*. Picador.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Barnes, B. (1981). On the conventional character of knowledge and cognition. *Philosophy of the Social Sciences*, *11*(3), 303–333. https://doi.org/10.1177/004839318101100303

Barrett, L. F. (2006a). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1),28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2009). The future of psychology: Connecting mind to brain. *Perspectives on Psychological Science*, *4*(4), 326–339. https://doi.org/10.1111/j.1745-6924.2009.01134.x

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*(3), 413–429. https://doi.org/10.1037/a0027555

Barrett, L. F. (2015). Ten common misconceptions about psychological construction theories of emotion. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 45–79). Guilford Press.

Barrett, L. F. (2017a). Categories and Their Role in the Science of Emotion. *Psychological Inquiry*, *28*(1), 20–26. https://doi.org/10.1080/1047840X.2017.1261581

Barrett, L. F. (2017b). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F. (2017c). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F. (2018). Seeing Fear: It's All in the Eyes? *Trends in Neurosciences*, *41*(9), 559–563. https://doi.org/10.1016/j.tins.2018.06.009

Barrett, L. F. (2019a). Looking at Psychology Through the Lens of Metascience. *APS Observer*, *32*. https://www.psychologicalscience.org/observer/looking-atpsychology-through-the-lens-of-metascience

Barrett, L. F. (2019b). The Publication Arms Race. *APS Observer*, *32*. https://www.psychologicalscience.org/observer/the-publications-arms-race

Barrett, L. F. (2020). *Seven and a Half Lessons About the Brain*. Houghton Mifflin Harcourt.

Barrett, L. F. (2021a). AI weighs in on debate about universal facial expressions. *Nature*, *589*(7841), 202–203. https://doi.org/10.1038/d41586-020-03509-5

Barrett, L. F. (2021b). Debate about universal facial expressions goes big. *Nature*, *589*(7841), 202–203. https://doi.org/10.1038/d41586-020-03509-5

Barrett, L. F. (2022). Context reconsidered: Complex signal ensembles, relational meaning, and population thinking in psychological science. *American Psychologist*, *77*(8), 894. https://doi.org/10.1037/amp0001054

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., & Lida, T. (2024). Constructionist approaches to emotion in psychology and related fields. In A. Scarantino (Ed.), *Emotion theory: The comprehensive Routledge guide,* Routledge.

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*(8), 327–332. https://doi.org/10.1016/j.tics.2007.06.003

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology*, *23*(3), 361–372. https://doi.org/10.1016/j.conb.2012.12.012

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950

Barrett, L. F., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). The conceptual act theory: A roadmap. In L. F. Barrett & J.A. Russell (Eds.), *The psychological construction of emotion* (pp. 83–110). Guilford Press.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*(1), 82–93. https://doi.org/10.3758/BF03197629

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227. https://doi.org/10.3758/BF03196968

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629-654. https://doi.org/10.1037//0278-7393.11.1-4.629

Barsalou, L. W. (1986). Are there static category representations in long-term memory? *Behavioral and Brain Sciences*, *9*(4), 651–652. https://doi.org/10.1017/S0140525X00051591

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge University Press.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76–121). Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.006

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 27, pp. 1–64). Academic Press. https://www.sciencedirect.com/science/article/pii/S0079742108601206

Barsalou, L. W. (1993). Concepts and meaning. In L. W. Barsalou, W. Yeh, B. Luka, K. Olseth, K. Mix, & L. Wu (Eds.), *Chicago linguistic society 29: Papers from the parasession on conceptual representations* (pp. 23–61). University of Chicago.

Barsalou, L. W. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, *28*(1), 61–80. https://doi.org/10.1080/01638539909545069

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281–1289. https://doi.org/10.1098/rstb.2008.0319

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, *2*(4), 716–724. https://doi.org/10.1111/j.1756-8765.2010.01115.x

Barsalou, L. W. (2017). Cognitively plausible theories of concept composition. In J. Hampton & Y. Winter (Eds.), *Compositionality and concepts in linguistics and psychology (Vol. 3*, pp. 9–30). Springer. https://doi.org/10.1007/978-3-319-45977-6_2

Barsalou, L. W. (2021). Categories at the interface of cognition and action. In C. Mauri, I. Fiorentini, & E. Goria (Eds.), *Building categories in interaction: Linguistic resources at work* (pp. 35–72). John Benjamins Publishing Company.

Barsalou, L. W., & Hale, C. R. (1993). Components of conceptual representation: From feature lists to recursive frames. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 97–144). Academic Press.

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social embodiment. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 43, pp. 43-92). Elsevier Science.

Barsalou, L. W., & Sewell, D. R. (1984). *Constructing representations of categories from different points of view*. Emory University.

Barsalou, L. W., Wilson, C. D., & Hasenkamp, W. (2010). On the vices of nominalization and the virtues of contextualizing. In B. Mesquita, L. F. Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 334–360). Guilford Press.

Bassett, D. S., & Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in Cognitive Sciences*, *15*(5), 200–209. https://doi.org/10.1016/j.tics.2011.03.006

Bateson, G., & Mead, M. (1942). *Balinese character; a photographic analysis*. New York Academy of Sciences.

Berger, P. L., & Luckmann, T. (2011). *The social construction of reality: A Treatise in the sociology of knowledge*. Open Road Media. (Original work published 1966)

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting and the incentive-sensitization theory of addiction. *The American Psychologist*, *71*(8), 670–679. https://doi.org/10.1037/amp0000059

Bhaskar, R. (2008). *A realist theory of science*. Verso.

Bird, A., & Tobin, E. (2024). Natural kinds. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2024/entries/natural-kinds/

Bloor, D. (1991). *Knowledge and social imagery* (2nd ed.). University of Chicago Press.

Bloor, D. (2011). Relativism and the sociology of scientific knowledge. In S. D. Hales (Ed.), *A companion to relativism* (pp. 431–455). Wiley-Blackwell.

Bogdanov, A. (2020). *Empiriomonism: Essays in philosophy, Books 1–3* (D. Rowley, Trans.). Haymarket Books.

Boiger, M., & Mesquita, B. (2012). The construction of emotion in interactions, relationships, and cultures. *Emotion Review*, *4*(3), 221–229. https://doi.org/10.1177/1754073912439765

Bonanno, G. A., Chen, S., Bagrodia, R., & Galatzer-Levy, I. R. (2024). Resilience and disaster: Flexible adaptation in the face of uncertain threat. *Annual Review of Psychology*, *75*, 573–599. https://doi.org/10.1146/annurev-psych-011123024224

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*(6), 948–957. https://doi.org/10.1037//0012-1649.38.6.948

Borkenau, P. (1990). Traits as ideal-based and goal-derived social categories. *Journal of Personality and Social Psychology*, *58*(3), 381–396. https://doi.org/10.1037/00223514.58.3.381

Boyd, R. (1979). Metaphor and theory change. In A. Ortony (Ed.), *Metaphor and thought* (pp. 481–532). Cambridge University Press.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, *61*(1), 127–148. https://doi.org/10.1007/bf00385837

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141–185). MIT Press.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(supplement_2), 10918–10925. https://doi.org/10.1073/pnas.1100290108

Braithwaite, R. B. (1953). *Scientific explanation; a study of the function of theory, probability and law in science*. Cambridge University Press.

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing Revealed by Human fMRI and MEG Decoding. *Journal of Neuroscience*, *37*(32), 7700–7710. https://doi.org/10.1523/JNEUROSCI.0582-17.2017

Breakspear, M., & McIntosh, A. R. (2011). Networks, noise and models: Reconceptualizing the brain as a complex, distributed system. *NeuroImage*, *58*(2), 293–295. https://doi.org/10.1016/j.neuroimage.2011.03.056

Bressler, S. L., & McIntosh, A. R. (2007). The role of neural context in large-scale neurocognitive network operations. In V. K. Jirsa & A. McIntosh (Eds.), *Handbook of brain connectivity* (pp. 403–419). Springer. https://doi.org/10.1007/978-3-54071512-2_14

Brewer, W. F. (2012). The theory ladenness of the mental processes used in the scientific enterprise: Evidence from cognitive psychology and the history of science. In R. W. Proctor & E. J. Capaldi (Eds.), *Psychology of science: Implicit and explicit processes* (pp. 289–334). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199753628.003.0013

Brick, C., Hood, B., Ekroll, V., & de-Wit, L. (2022). Illusory essences: A bias holding back theorizing in psychological science. *Perspectives on Psychological Science*, *17*(2), 491–506. https://doi.org/10.1177/1745691621991838

Bridgman, P. W. (1927). *The logic of modern physics*. Macmillan.

Brown, J. R. (2017). Mathematics, role in science. In W. H. Newton-Smith (Ed.), *A companion to the philosophy of science* (pp. 257–264). Blackwell.

Browne, J. (2007). *Darwin's origin of species: A biography*. Atlantic Books.

Browne, J. (2010). Making Darwin: Biography and the changing representations of Charles Darwin. *The Journal of Interdisciplinary History*, *40*(3), 347–373. https://doi.org/10.1162/jinh.2010.40.3.347

Bruner, J. S. (1990). *Acts of meaning*. Harvard University Press.

Bruner, J. S. (1973). *Beyond the information given: Studies in the psychology of knowing*. W. W. Norton.

Bull, N. (1945). Towards a clarification of the concept of emotion. *Psychosomatic Medicine*, *7*, 210–214. https://doi.org/10.1097/00006842-194507000-00002

Bursten, J. R. (2018). Smaller than a breadbox: Scale and natural kinds. *The British Journal for the Philosophy of Science*, *69*(1), 1–23. https://doi.org/10.1093/bjps/axw022

Buzsáki, G. (2019). *The brain from inside out*. Oxford University Press. https://doi.org/10.1093/oso/9780190905385.001.0001

Buzsáki, G., & Tingley, D. (2023). Cognition from the body-brain partnership: Exaptation of memory. *Annual Review of Neuroscience*, *46*, 191–210. https://doi.org/10.1146/annurev-neuro-101222-110632

Byrne, D., & Callaghan, G. (2022). *Complexity theory and the social sciences* (2nd ed.). Routledge.

Cadenasso, M. L., Pickett, S. T. A., Weathers, K. C., & Jones, C. G. (2003). A framework for a theory of ecological boundaries. *BioScience*, *53*(8), 750–758. https://doi.org/10.1641/0006-3568(2003)053[0750:AFFATO]2.0.CO;2

Candiotto, L. (2017). The reality of relations. *Giornale Di Metafisica*, *2*, 537-551. https://philsci-archive.pitt.edu/14165/

Carnap, R. (1950). Empiricism, semantics and ontology. *Revue Internationale de Philosophie*, *4*(11), 20–40. https://www.jstor.org/stable/23932367

Carnap, R. (2002). *The logical syntax of language*. Open Court Publishing. (Original work published 1937)

Carruthers, P. (1992). *Human knowledge and human nature: A new introduction to an ancient debate*. Oxford University Press.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford University Press.

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566). MIT Press.

https://direct.mit.edu/books/editedvolume/3430/chapter/115706/All-Concepts-Are-Ad-Hoc-Concepts

Case, J. (2001). The heart of Putnam's pluralistic realism. *Revue internationale de philosophie*, *218*(4), 417–430. https://doi.org/10.3917/rip.218.0417

Cassirer, E. (1923). *Einstein's theory of relativity*. Open Court. (Original work published 1921)

Cassirer, E. (1923). *Substance and function*. Open Court. (Original work published 1910)

Cassirer, E. (1956). *Determinism and indeterminism in modern physics*. Yale University Press. (Original work published 1936)

Castelhano, M. S., & Pereira, E. J. (2018). The influence of scene context on parafoveal processing of objects. *The Quarterly Journal of Experimental Psychology*, *71*(1), 229–240. https://doi.org/10.1080/17470218.2017.1310263

Ceulemans, E., Kuppens, P., & Van Mechelen, I. (2012). Capturing the structure of distinct types of individual differences in the situation-specific experience of emotions: The case of anger. *European Journal of Personality*, *26*(5), 484–495. https://doi.org/10.1002/per.847

Chakravartty, A. (2007). *A metaphysics for scientific realism: Knowing the unobservable*. Cambridge University Press. https://doi.org/10.1017/CBO9780511487354

Chakravartty, A. (2017). Scientific realism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/

Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences*, *20*(2), 96–106. https://doi.org/10.1016/j.tics.2015.11.005

Chang, H. (2007). Inventing temperature: Measurement and scientific progress. Oxford University Press.

Chang, H. (2012). *Is Water H2O?: Evidence, realism and pluralism* (Vol. 293). Springer Netherlands. https://doi.org/10.1007/978-94-007-3932-1

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, *54*(4), 541–557. https://doi.org/10.1037/0022-3514.54.4.541

Chomsky, N. (1980). *Rules and representations*. Columbia University Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, *78*(2), 67–90. https://doi.org/10.2307/2025900

Cisek, P. (2019). Resynthesizing behavior through phylogenetic refinement. *Attention, Perception, & Psychophysics*, *81*(7), 2265–2287. https://doi.org/10.3758/s13414-019-01760-1

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*, *27*(3), 1944–1948. https://doi.org/10.1093/cercor/bhw028

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Cohnitz, D., & Rossberg, M. (2024). Nelson Goodman. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2024/entries/goodman/

Collins, H. (1992). *Changing order: Replication and induction in scientific practice*. University of Chicago Press.

Corbett, M. (2015). From law to folklore: Work stress and the Yerkes-Dodson Law. *Journal of Managerial Psychology*, *30*(6), 741–752. https://doi.org/10.1108/JMP-03-20130085

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion (Washington, D.C.)*, *18*(1), 75–93. https://doi.org/10.1037/emo0000302

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). Oxford University Press.

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, *24*(2), 299–349. https://doi.org/10.1207/s15516709cog2402_4

Cowen, A. S., & Keltner, D. (2021). Semantic space theory: A Computational approach to emotion. *Trends in Cognitive Sciences*, *25*(2), 124–136. https://doi.org/10.1016/j.tics.2020.11.004

Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, *589*(7841), 251–257. https://doi.org/10.1038/s41586-020-3037-7

Crivelli, C., & Fridlund, A. J. (2018). Facial displays are tools for social influence. *Trends in Cognitive Sciences*, *22*(5), 388–399. https://doi.org/10.1016/j.tics.2018.02.006

Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J.-M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, *145*(7), 830–843. https://doi.org/10.1037/xge0000172

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J.-M. (2016). The fear gasping face as a threat display in a Melanesian society. *Proceedings of the National Academy of Sciences*, *113*(44), 12403–12407. https://doi.org/10.1073/pnas.1611622113

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*(2), 116–127. https://doi.org/10.1037/h0076829

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. John Wiley.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, *16*(2), 137–163. https://doi.org/10.1111/j.2044-8317.1963.tb00206.x

Cunningham, W. A., Dunfield, K. A., & Stillman, P. E. (2013). Emotional states from affective dynamics. *Emotion Review*, *5*(4), 344–355. https://doi.org/10.1177/1754073913489749

Danziger, K. (1988). On theory and method in psychology. In W. J. Baker, L. P. Mos, H. V. Rappard, & H. J. Stam (Eds.), *Recent trends in theoretical psychology* (pp. 87–94). Springer.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological rResearch*. Cambridge University Press. https://doi.org/10.1017/CBO9780511524059

Danziger, K. (1997). *Naming the mind: How psychology found its language*. SAGE.

Danziger, K. (2000). Making social psychology experimental: A conceptual history, 1920–1970. *Journal of the History of the Behavioral Sciences*, *36*(4), 329–347. https://doi.org/10.1002/1520-6696(200023)36:4<329::AID-JHBS3>3.0.CO;2-5

Dashiell, J. F. (1927). A new method of measuring reactions to facial expression of emotion. *Psychological Bulletin*, *24(3)*, 174–175. https://doi.org/10.1037/h0065883

Dashiell, J. F. (1928). Are there any native emotions? *Psychological Review*, *35*(4), 319–327. https://doi.org/10.1037/h0075969

Dashiell, J. F. (1935). Experimental studies of the influence of social situations on the behavior of individual human adults. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 1097–1158). Clark University Press.

Dawkins, R. (2006). *The selfish gene*. (30th anniversary ed.). OUP Oxford. (Original work published 1976)

DeJesus, J. M., Callanan, M. A., Solis, G., & Gelman, S. A. (2019). Generic language in scientific communication. *Proceedings of the National Academy of Sciences*, *116*(37), 18370–18377. https://doi.org/10.1073/pnas.1817706116

Dennett, D. C. (1987). *The intentional stance*. MIT Press.

DeSilva, J. (Ed.). (2021). *A most interesting problem: What Darwin's descent of man got right and wrong about human evolution*. Princeton University Press.

Dewey, J. (1895). The theory of emotion. *Psychological Review*, *2*(1), 13–32. https://doi.org/10.1037/h0070927

Dewey, J. (1900). Some stages of logical thought. *The Philosophical Review*, *9*(5), 465–489. https://doi.org/10.2307/2176692

Dewey, J. (1903). *Studies in logical theory*. AMS Press.

Di Biagio, A., & Rovelli, C. (2021). Stable facts, relative facts. *Foundations of Physics*, *51*(1), 30. https://doi.org/10.1007/s10701-021-00429-w

DiGirolamo, M. A., & Russell, J. A. (2017). The emotion seen in a face can be a methodological artifact: The process of elimination hypothesis. *Emotion*, *17*(3), 538–546. https://doi.org/10.1037/emo0000247

Donati, P., & Archer, M. S. (2015). *The relational subject*. Cambridge University Press. https://doi.org/10.1017/CBO9781316226780

Doyle, C. M., Lane, S. T., Brooks, J. A., Wilkins, R. W., Gates, K. M., & Lindquist, K. A. (2022). Unsupervised classification reveals consistency and degeneracy in neural network patterns of emotion. *Social Cognitive and Affective Neuroscience*, *17*(11), 995–1006. https://doi.org/10.1093/scan/nsac028

Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General*, *147*(1), 62–73. https://doi.org/10.1037/xge0000361

Drew, T., Võ, M. L.-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological Science*, *24*(9), 1848–1853. https://doi.org/10.1177/0956797613479386

Dubova, M., & Goldstone, R. L. (2023). Carving joints into nature: Reengineering scientific concepts in light of concept-laden evidence. *Trends in Cognitive Sciences*, *27*(7), 656–670. https://doi.org/10.1016/j.tics.2023.04.006

Duffy, E. (1934a). Emotion: An example of the need for reorientation in psychology. *Psychological Review*, *41*(2), 184–198. https://doi.org/10.1037/h0074603

Duffy, E. (1934b). Is emotion a mere term of convenience? *Psychological Review*, *41*(1), 103–104. https://doi.org/10.1037/h0075951

Duffy, E. (1941). An explanation of "emotional" phenomena without the use of the concept "emotion." *Journal of General Psychology*, *25*, 283–293. https://doi.org/10.1080/00221309.1941.10544400

Duhem, P. M. M. (1991). *The aim and structure of physical theory* (P. P. Wiener, Trans.; 2nd ed.). Princeton University Press. (Original work published 1954)

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, *22*(9), 780–793. https://doi.org/10.1016/j.tics.2018.06.004

Dunlap, K. (1932). Are emotions teleological constructs? *The American Journal of Psychology*, *44*, 572–576. https://doi.org/10.2307/1415359

Dunlap, K. (1935). The average animal. *Journal of Comparative Psychology*, *19*(1), 1–3. https://doi.org/10.1037/h0054792

Dupré, J. (1981). Natural kinds and biological taxa. *The Philosophical Review*, *90*(1), 66–90. https://doi.org/10.2307/2184373

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Harvard University Press.

Dupré, J. (1999). On the impossibility of a monistic account of species. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 3–22). Bradford Books.

Durán, J. I., & Fernández-Dols, J.-M. (2021). Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion. *Emotion (Washington, D.C.)*, *21*(7), 1550–1569. https://doi.org/10.1037/emo0001015

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(24), 13763–13768. https://doi.org/10.1073/pnas.231499798

Eidelson, R. J. (1997). Complex adaptive systems in the behavioral and social sciences. *Review of General Psychology*, *1*(1), 42–71. https://doi.org/10.1037/1089-2680.1.1.42

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*(1), 31–34. https://doi.org/10.1177/1745691615596992

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*, *5*(1), 90–96. https://doi.org/10.1177/1754073912451332

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Ellis, B. (2001). *Scientific essentialism*. Cambridge University Press.

Ellis, B. (2002). *The philosophy of nature: A guide to the new essentialism*. Routledge.

Elms, A. C. (1975). The crisis of confidence in social psychology. *American Psychologist*, *30*, 967–976. https://doi.org/10.1037/0003-066X.30.10.967

Engelen, T., Solcà, M., & Tallon-Baudry, C. (2023). Interoceptive rhythms in the brain. *Nature Neuroscience*, *26*(10), 1670–1684. https://doi.org/10.1038/s41593-023-01425-1

Epperson, M., & Zafiris, E. (2013). *Foundations of relational realism: A topological approach to quantum mechanics and the philosophy of nature*. Lexington Books.

Ereshefsky, M. (2018). Natural kinds, mind independence, and defeasibility. *Philosophy of Science*, *85*(5), 845–856. https://doi.org/10.1086/699676

Esfeld, M. (2009). Explanation (Deductive-Nomological, Mereological, Reductive). In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of neuroscience* (pp. 1495–1498). Springer. https://doi.org/10.1007/978-3-540-29678-2_3214

Esserman, L., & Eggener, S. (2023, August 30). Opinion | Not Everything We Call Cancer Should Be Called Cancer. *The New York Times*. https://www.nytimes.com/2023/08/30/opinion/cancer-breast-prostatetreatment.html

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*(2), 134–140. https://doi.org/10.1037/h0045156

Estes, W. K. (1994). *Classification and cognition*. Oxford University Press.

Fanselow, M. S. (2018). Emotion, motivation and function. *Current Opinion in Behavioral Sciences*, *19*, 105–109. https://doi.org/10.1016/j.cobeha.2017.12.013

Fausto-Sterling, A. (2021). A dynamic systems framework for gender/sex development: From sensory input in infancy to subjective certainty in toddlerhood. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.613789

Favela, L. H. (2020). Cognitive science as complexity science. *WIREs Cognitive Science*, *11*(4), e1525. https://doi.org/10.1002/wcs.1525

Feest, U. (2010). Concepts as tools in the experimental generation of knowledge in cognitive neuropsychology. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, *4*(1), 173–190. https://doi.org/10.4245/sponge.v4i1.11938

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*(3), 464-486. https://doi.org/10.1037/0096-3445.113.3.464

Festinger, L. (1953). Laboratory experiments. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences* (pp. 137–172). Dryden Press.

Festinger, L. (1980). *Retrospections in social psychology* (L. Festinger, Ed.). Oxford University Press.

Festinger, L., Back, K., Schachter, S., Kelley, H. H., & Thibaut, J. (1950). *Theory and experiment in social communication*. Research Center for Dynamics Institute for Social Research.

Feyerabend, P. K. (1962). Explanation, reduction, and empiricism. In H. Feigl & G. Maxwell (Eds.), *Scientific explanation, space, and time* (Vol. 3, pp. 28-97). University of Minnesota Press.

Feyerabend, P. K. (2010). *Against method* (4th ed.). Verso. (Original work published 1975)

Feynman, R. (1965). *The character of physical law*. MIT Press.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *The Behavioral and Brain Sciences*, *39*, e229. https://doi.org/10.1017/S0140525X15000965

Flournoy, J. C., Bryce, N. V., Dennison, M. J., Rodman, A. M., McNeilly, E. A., Lurie, L. A., Bitran, D., Reid-Russell, A., Vidal Bustamante, C. M., Madhyastha, T., & McLaughlin, K. A. (2024). A precision neuroscience approach to estimating reliability of neural responses during emotion processing: Implications for task-fMRI. *NeuroImage*, *285*, 120503. https://doi.org/10.1016/j.neuroimage.2023.120503

Fraley, R. C., & Marks, M. J. (2007). The null hypothesis significance-testing debate and its implications for personality research. In R. W. Robins, R. C. Fraley, & R. F. Kruger (Eds.), *Handbook of research methods in personality psychology* (pp. 149–169). The Guilford Press.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist*, *43*(5), 349–358. https://doi.org/10.1037/0003-066X.43.5.349

Frijda, N. H. (2006). *The laws of emotion*. Psychology Press.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Futch, M. (2008). *Leibniz's metaphysics of time and space*. Springer Science & Business Media.

Gallistel, C. R. (2012). On the evils of group averaging: Commentary on Nevin's "Resistance to extinction and behavioral momentum." *Behavioural Processes*, *90*(1), 98–99. https://doi.org/10.1016/j.beproc.2012.02.013

Gardiner, P. L. (1959). *Theories of history: Readings from classical and contemporary sources*. Free Press.

Gelman, A. (2015). The connection between varying treatment effects and the crisis of unreplicable research: A Bayesian perspective. *Journal of Management*, *41*(2), 632–643. https://doi.org/10.1177/0149206314525208

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195154061.001.0001

Gelman, S. A. (2023). Looking beyond the obvious. *American Psychologist*, *78*(5), 667–677. https://doi.org/10.1037/amp0001152

Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7900–7907. https://doi.org/10.1073/pnas.1621073114

Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality reconsidered: Diversity in making meaning of facial expressions. *Current Directions in Psychological Science*, *27*(4), 211–219. https://doi.org/10.1177/0963721417746794

Gendron, M., & Barrett, L. F. (2009). Reconstructing the Past: A Century of Ideas About Emotion in Psychology. *Emotion Review*, *1*(4), 316–339. https://doi.org/10.1177/1754073909338877

Gendron, M., Hoemann, K., Crittenden, A. N., Mangola, S. M., Ruark, G. A., & Barrett, L. F. (2020). Emotion perception in Hadza hunter-gatherers. *Scientific Reports*, *10*(1), 3867. https://doi.org/10.1038/s41598-020-60257-2

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion (Washington, D.C.)*, *12*(2), 314–325. https://doi.org/10.1037/a0026007

Gendron, M., Mesquita, B., & Barrett, L. F. (2020). The brain as a cultural artifact: Concepts, actions, and experiences within the human affective niche. In C. M. Worthman, C. A. Cummings, L. J. Kirmayer, R. Lemelson, & S. Kitayama (Eds.), *Culture, mind, and brain: Emerging concepts, models, and applications* (pp. 188–222). Cambridge University Press. https://doi.org/10.1017/9781108695374.010

Gendron, M., Roberson, D., & Barrett, L. F. (2015). Cultural variation in emotion perception is real: A response to Sauter et al. *Psychological Science*, *26*(3), 357–359. https://doi.org/10.1177/0956797614566659

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, *25*(4), 911–920. https://doi.org/10.1177/0956797613517239

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion (Washington, D.C.)*, *14*(2), 251–262. https://doi.org/10.1037/a0036052

Gergen, K. J. (1978). Experimentation in social psychology: A reappraisal. *European Journal of Social Psychology*, *8*(4), 507–527. https://doi.org/10.1002/ejsp.2420080407

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, *40*(3), 266–275. https://doi.org/10.1037/0003-066X.40.3.266

Gernigon, C., Den Hartigh, R. J. R., Vallacher, R. R., & van Geert, P. L. C. (2023). How the complexity of psychological processes reframes the issue of reproducibility in psychological science. *Perspectives on Psychological Science*, 17456916231187324. https://doi.org/10.1177/17456916231187324

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*(1), 197–209. https://doi.org/10.1037/a0017808

Giere, R. N. (2010). *Scientific perspectivism*. University of Chicago Press.

Giffin, C., Wilkenfeld, D., & Lombrozo, T. (2017). The explanatory effect of a label: Explanations with named categories are more satisfying. *Cognition*, *168*, 357–369. https://doi.org/10.1016/j.cognition.2017.07.011

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 89-150). McGraw-Hill

Gleiser, M. (2015). *The island of knowledge: The limits of science and the search for meaning*. Basic Books.

Godfrey-Smith, P. (2021). *Theory and reality: An introduction to the philosophy of science* (2nd ed.). University of Chicago Press.

Goldstone, R. L., Kersten, A., & Carvalho, P. F. (2017). Categorization and concepts. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience (4th ed., Vol. 3,* pp. 275–317). Wiley. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119170174.epcn308

Goodman, N. (1978). *Ways of worldmaking*. Harvester Press.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205–1222. https://doi.org/10.1111/1467-8624.00224

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*(2), 409–427. https://doi.org/10.1111/j.1467-8624.2004.00683.x

Gray, J. S. (1935). An objective theory of emotion. *Psychological Review*, *42*(1), 108–116. https://doi.org/10.1037/h0061472

Greenwald, A. G., & Ronis, D. L. (1981). On the conceptual disconfirmation of theories. *Personality and Social Psychology Bulletin*, *7*(1), 131–137. https://doi.org/10.1177/014616728171020

Greenwood, J. D. (2000). Individualism and the social in early American social psychology. *Journal of the History of the Behavioral Sciences*, *36*(4), 443–455. https://doi.org/10.1002/1520-6696(200023)36:4<443::AID-JHBS9>3.0.CO;2-N

Greenwood, J. D. (2004). *The disappearance of the social in American social psychology*. Cambridge University Press.

Gross, C. T., & Canteras, N. S. (2012). The many paths to fear. *Nature Reviews. Neuroscience*, *13*(9), 651–658. https://doi.org/10.1038/nrn3301

Guillory, S. A., & Bujarski, K. A. (2014). Exploring emotions using invasive methods: Review of 60 years of human intracranial electrophysiology. *Social Cognitive and Affective Neuroscience*, *9*(12), 1880–1889. https://doi.org/10.1093/scan/nsu002

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.

Hacking, I. (1991). A tradition of natural kinds. *Philosophical Studies*, *61*(1/2), 109–126. https://doi.org/10.1007/BF00385836

Hacking, I. (1999). *The social construction of what?* Harvard University Press.

Hahn, C., Cowell, J. M., Wiprzycka, U. J., Goldstein, D., Ralph, M., Hasher, L., & Zelazo, P. D. (2012). Circadian rhythms in executive function during the transition to adolescence: The effect of synchrony between chronotype and time of day. *Developmental Science*, *15*(3), 408–416. https://doi.org/10.1111/j.14677687.2012.01137.x

Halgren, E., Walter, R. D., Cherlow, D. G., & Crandall, P. H. (1978). Mental phenomena evoked by electrical stimulation of the human hippocampal formation and amygdala. *Brain: A Journal of Neurology*, *101*(1), 83–117. https://doi.org/10.1093/brain/101.1.83

Halperin, E., Porat, R., Tamir, M., & Gross, J. J. (2013). Can emotion regulation change political attitudes in intractable conflicts? From the laboratory to the field. *Psychological Science*, *24*(1), 106–111. https://doi.org/10.1177/0956797612452572

Hampton, J. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441–461. https://doi.org/10.1016/S0022-5371(79)90246-9

Hampton, J. A. (1991). The combination of prototype concepts. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 91–116). Lawrence Erlbaum Associates, Inc.

Hampton, J. A., & Winter, Y. (Eds.). (2017). *Compositionality and concepts in linguistics and psychology*. Springer Cham.

Hanson, N. R. (1958). *Patterns of discovery: an inquiry into the conceptual foundations of science*. Cambridge University Press.

Harding, S. (1986). *The science question in feminism*. Cornell University Press.

Harding, S. (1992). Rethinking standpoint epistemology: What is "strong objectivity?" *The Centennial Review*, *36*(3), 437–470. https://www.jstor.org/stable/23739232

Harding, S. (2015). *Objectivity and diversity: another logic of scientific research*. University of Chicago Press.

Harlow, H. F., & Stagner, R. (1932). Psychology of feelings and emotions: I. Theory of feelings. *Psychological Review*, *39*(6), 570–589. https://doi.org/10.1037/h0072961

Harlow, H. F., & Stagner, R. (1933). Psychology of feelings and emotions. II. Theory of emotions. *Psychological Review*, *40*(2), 184–195. https://doi.org/10.1037/h0075696

Harmon-Jones, E., & Peterson, C. K. (2009). Supine body position reduces neural response to anger evocation. *Psychological Science*, *20*(10), 1209–1210. https://doi.org/10.1111/j.1467-9280.2009.02416.x

Harre, R. (1986). *The social construction of emotion*. Blackwell.

Haslam, N., & Whelan, J. (2008). Human natures: Psychological essentialism in thinking about differences between people. *Social and Personality Psychology Compass*, *2*(3), 1297–1312. https://doi.org/10.1111/j.1751-9004.2008.00112.x

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Lawrence Erlbaum Associates Publishers.

Hempel, C. G. (1942). The function of general laws in history. *Journal of Philosophy*, *39*(2), 35–48. https://doi.org/10.2307/2017635

Hempel, C. G. (1958). The theoretician's dilemma: A study in the logic of theory construction. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Concepts, theories, and the mind-body problem, (Minnesota studies in the philosophy of science, Vol. 2*, pp. 37–98). University of Minnesota Press. http://conservancy.umn.edu/handle/11299/184621

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. The Free Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*(2), 135–175. https://www.jstor.org/stable/185169

Herry, C., & Johansen, J. P. (2014). Encoding of fear learning and memory in distributed neuronal circuits. *Nature Neuroscience*, *17*(12), 1644–1654. https://doi.org/10.1038/nn.3869

Hildebrand, D. (2024). John Dewey. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2024/entries/dewey/

Hinzen, W., Machery, E., & Werning, M. (Eds.). (2012). *The Oxford handbook of compositionality*. Oxford University Press.

Hoemann, K., Crittenden, A. N., Msafiri, S., Liu, Q., Li, C., Roberson, D., Ruark, G. A., Gendron, M., & Feldman Barrett, L. (2019). Context facilitates performance on a classic cross-cultural emotion perception task. *Emotion*, *19*(7), 1292–1313. https://doi.org/10.1037/emo0000501

Hoemann, K., Gendron, M., Crittenden, A. N., Mangola, S. M., Endeko, E. S., Dussault, È., Barrett, L. F., & Mesquita, B. (2024). What we can learn about emotion by talking with the Hadza.

Perspectives on Psychological Science, *19*(1), 173–200. https://doi.org/10.1177/17456916231178555

Hoemann, K., Khan, Z., Feldman, M. J., Nielson, C., Devlin, M., Dy, J., Barrett, L. F., Wormwood, J. B., & Quigley, K. S. (2020). Context-aware experience sampling reveals the scale of variation in affective experience. *Scientific Reports*, *10*(1), 12459. https://doi.org/10.1038/s41598-020-69180-y

Hoemann, K., Wu, R., LoBue, V., Oakes, L. M., Xu, F., & Barrett, L. F. (2020). Developing an understanding of emotion categories: Lessons from objects. *Trends in Cognitive Sciences*, *24*(1), 39–51. https://doi.org/10.1016/j.tics.2019.10.010

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*, *55*(9), 1830–1849. https://doi.org/10.1037/dev0000686

Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, *125*(3), 293-328. https://doi.org/10.1037/rev0000094

Höhn, P. A. (2017). Toolbox for reconstructing quantum theory from rules on information acquisition. *Quantum*, *1*, 38. https://doi.org/10.22331/q-2017-12-14-38

Hoyningen-Huene, P. (1989). Epistemological reductionism in biology: intuitions, explications, and objections. In P. Hoyningen-Huene & F. M. Wuketits (Eds.), *Reductionism and systems theory in the life sciences: Some problems and perspectives* (pp. 29–44). Springer Netherlands. https://doi.org/10.1007/978-94-0091003-4_3

Huggett, N., Hoefer, C., & Read, J. (2023). Absolute and relational space and motion: PostNewtonian theories. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/spacetime-theories/

Hull, C. L. (1931). Goal attraction and directing ideas conceived as habit phenomena. *Psychological Review*, *38*(6), 487–506. https://doi.org/10.1037/h0071442

Hull, D. L. (1988). *Science as a process: an evolutionary account of the social and conceptual development of science*. University of Chicago Press.

Hume, D. (2004). *A treatise of human nature*. Penguin UK. (Original work published 1739)

Hunt, W. A. (1941). Recent developments in the field of emotion. *Psychological Bulletin*, *38*(5), 249–276. https://doi.org/10.1037/h0054615

Hutchinson, B., & Barrett, L. F. (2019). The power of predictions: An emerging paradigm for psychological research. *Current Directions in Psychological Science*, *28*(3), 280–291. https://doi.org/10.1177/0963721419831992

Huxley, J. (1942). *Evolution, the modern synthesis*. Harper & Brothers.

Izard, C. E. (1977). *Human emotions*. Plenum Press.

Jablonka, E., & Lamb, M. J. (2014). *Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life* (Rev. ed.). MIT Press.

Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B., & Schyns, P. G. (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General*, *145*(6), 708–730. https://doi.org/10.1037/xge0000162

Jackman, H. (2020). Meaning holism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/meaning-holism/

James, W. (1884). What is an emotion? *Mind*, *9*(34), 188–205. https://doi.org/10.1093/mind/os-IX.34.188

James, W. (1905). The experience of activity. *Psychological Review*, *12*(1), 1–17. https://doi.org/10.1037/h0070340

James, W. (1907). Pragmatism and humanism. In *Pragmatism: A new name for some old ways of thinking* (pp. 92–104). Longmans, Green and Co. https://doi.org/10.1037/10851-007

James, W. (1931). *The principles of psychology* (Vol. 1). Henry Holt and Co. (Original work published 1890)

James, W. (1994). The physical basis of emotion. *Psychological Review*, *101*(2), 205–210. https://doi.org/10.1037/0033-295X.101.2.205 (Original work published 1894)

James, W. (2009). *The will to believe and other essays in popular philosophy*. Longmans, Green, and Co; Project Gutenberg. https://www.gutenberg.org/files/26659/26659-h/26659-h.htm (Original work published 1897)

Juarrero, A. (2023). *Context changes everything: How constraints create coherence*. The MIT Press.

Kaipayil, J. (2009). *Relationalism: A theory of being*. JIP Publications.

Katsumi, Y., Theriault, J. E., Quigley, K. S., & Barrett, L. F. (2022). Allostasis as a core feature of hierarchical gradients in the human brain. *Network Neuroscience*, 1–22. https://doi.org/10.1162/netn_a_00240

Katsumi, Y., Zhang, J., Chen, D., Kamona, N., Bunce, J. G., Hutchinson, J. B., Yarossi, M., Tunik, E., Dickerson, B. C., & Quigley, K. S. (2023). Correspondence of functional connectivity gradients across human isocortex, cerebellum, and hippocampus. *Communications Biology*, *6*(1), 401. https://doi.org/10.1038/s42003-023-04796-0

Kelso, J. A. S. (2012). Multistability and metastability: Understanding dynamic coordination in the brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1591), 906–918. https://doi.org/10.1098/rstb.2011.0351

Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Networks: The Official Journal of the International Neural Network Society*, *37*, 120–131. https://doi.org/10.1016/j.neunet.2012.09.003

Khalaf, A., Nabian, M., Fan, M., Yin, Y., Wormwood, J., Siegel, E., Quigley, K. S., Barrett, L. F., Akcakaya, M., Chou, C.-A., & Ostadabbas, S. (2020). Analysis of multimodal physiological signals within and between individuals to predict psychological challenge vs. Threat. *Expert Systems with Applications*, *140*, 112890. https://doi.org/10.1016/j.eswa.2019.112890

Kilaj, A., Gao, H., Rösch, D., Rivero, U., Küpper, J., & Willitsch, S. (2018). Observation of different reactivities of para and ortho-water towards trapped diazenylium ions. *Nature Communications*, *9*(1), 2096. https://doi.org/10.1038/s41467-018-04483-3

Kirchhoff, M. (2018). Predictive brains and embodied, enactive cognition: An introduction to the special issue. *Synthese*, *195*(6), 2355–2366. https://doi.org/10.1007/s11229-0171534-5

Kitcher, P. (1984). Species. *Philosophy of Science*, *51*(2), 308–333. https://doi.org/10.1086/289182

Klein, D. B. (1970). *A history of scientific psychology: Its origins and philosophical backgrounds*. Basic Books.

Kolada, E., Bielski, K., Wilk, M., Rymarczyk, K., Bogorodzki, P., Kazulo, P., Kossowski, B., Wypych, M., Marchewka, A., Kaczmarek, L., Knapska, E., & Szatkowska, I. (2023). The human centromedial amygdala contributes to negative prediction error signaling during appetitive and aversive pavlovian gustatory learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *43*(17), 3176–3185. https://doi.org/10.1523/JNEUROSCI.0926-22.2023

Koslicki, K. (2008). Natural kinds and natural kind terms. *Philosophy Compass*, *3*(4), 789–802. https://doi.org/10.1111/j.1747-9991.2008.00157.x

Kraut, R. (2022). Plato. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2022/entries/plato/

Kripke, S. A. (1980). *Naming and necessity*. Harvard University Press. (Original work published 1972)

Kuhn, T. S. (1970). *The structure of scientific revolutions*. University of Chicago Press. (Original work published 1962)

Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension* (pp. 320–339). University of Chicago Press. https://doi.org/10.7208/9780226217239014

Kuhn, T. S. (1979). *The essential tension: Selected studies in scientific tradition and change*. University of Chicago Press.

Kuhn, T. S. (1992). *The Copernican revolution: Planetary astronomy in the development of western thought*. Harvard University Press. (Original work published 1957)

Kuhn, T. S. (2002). *The road since structure: Philosophical Essays, 1970-1993, with an Autobiographical Interview* (J. Conant & J. Haugeland, Eds.). University of Chicago Press.

Kuppens, P., Van Mechelen, I., Smits, D. J., & De Boeck, P. (2003). The appraisal basis of anger: Specificity, necessity and sufficiency of components. *Emotion*, *3*(3), 254-269. https://doi.org/10.1037/1528-3542.3.3.254

Lakatos, I. (Ed.). (1976). *Proofs and refutations: the logic of mathematical discovery*. Cambridge University Press.

Lakatos, I. (1978). Falsification and the methodology of scientific research programmes. In J. Worrall & G. Currie (Eds.), *The methodology of scientific research programmes* (pp. 8–101). Cambridge University Press. (Original work published 1970)

Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge: Proceedings of the international colloquium in the philosophy of science, London, 1965* (Vol. 4). Cambridge University Press. https://doi.org/10.1017/CBO9781139171434

Lakoff, G. (2016). Language and emotion. *Emotion Review*, *8*(3), 269–273. https://doi.org/10.1177/1754073915595097

Lakoff, G., & Nuñez, R. (2001). *Where mathematics come from*. Basic Books.

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology*, *30*, 191–202. https://doi.org/10.1007/s10682-016-9821z

Laland, K. N., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G. A., Hoekstra, H. E., Futuyma, D. J., Lenski, R. E., Mackay, T. F. C., Schluter, D., & Strassmann, J. E. (2014). Does evolutionary theory need a rethink? *Nature*, *514*(7521), 161–164. https://doi.org/10.1038/514161a

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., & Odling-Smee, J. (2015). The extended evolutionary synthesis: Its structure, assumptions and predictions. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1813), 20151019. https://doi.org/10.1098/rspb.2015.1019

Landis, C. (1924). Studies of emotional reactions. II. General behavior and facial expression. *Journal of Comparative Psychology*, *4*(5), 447–510. https://doi.org/10.1037/h0073039

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton University Press. https://www.jstor.org/stable/j.ctt32bbxc (Original work published 1979)

Laudan, L. (1977). *Progress and its problems: Towards a theory of scientific growth*. University of California Press.

Laudan, L. (1990). *Science and relativism: Some key controversies in the philosophy of science*. University of Chicago Press.

Lavoisier, A. L. (2009). *Elements of chemistry, in a new systematic order, containing all the modern discoveries* (R. Kerr, Trans.). G.G. and J. J. Robinsons; Project Gutenberg. https://www.gutenberg.org/ebooks/30775 (Original work published 1789)

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lazarus, R. S. (2001). Relational meaning and discrete emotions. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 37–67). Oxford University Press.

Le Mau, T., Hoemann, K., Lyons, S. H., Fugate, J. M. B., Brown, E. N., Gendron, M., & Barrett, L. F. (2021). Professional actors demonstrate variability, not stereotypical expressions, when portraying emotional states in photographs. *Nature Communications*, *12*(1), 5037. https://doi.org/10.1038/s41467-021-25352-6

Leary, D. (1987). From act psychology to problematic functionalism: The place of Egon Brunswik in the history of psychology. In M. G. Ash & W. R. Woodward (Eds.), *Psychology in twentieth-century thought and society* (pp. 115–142). Cambridge University Press. https://scholarship.richmond.edu/psychology-facultypublications/18

Lebois, L. A., Wilson-Mendenhall, C. D., Simmons, W. K., Barrett, L. F., & Barsalou, L. W. (2020). Learning situated emotions. *Neuropsychologia*, *145*, 106637. https://doi.org/10.1016/j.neuropsychologia.2018.01.008

Ledoux, J. (2015). *Anxious: Using the brain to understand and treat fear and anxiety*. Penguin Books.

Leek, J. T., & Peng, R. D. (2015). What is the question? *Science*, *347*(6228), 1314–1315. https://doi.org/10.1126/science.aaa6146

Leleu, A., Godard, O., Dollion, N., Durand, K., Schaal, B., & Baudouin, J.-Y. (2015). Contextual odors modulate the visual processing of emotional facial expressions: An ERP study. *Neuropsychologia*, *77*, 366–379. https://doi.org/10.1016/j.neuropsychologia.2015.09.014

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review*, *3*(4), 379–386. https://doi.org/10.1177/1754073911410743

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*(4), 421–431. https://www.jstor.org/stable/27836590

Levins, R. (1968). *Evolution in changing environments: Some theoretical explorations*. Princeton University Press.

Lewin, K. (1931). The conflict between Aristotelian and Galileian modes of thought in contemporary psychology. *The Journal of General Psychology*, *5*(2), 141-177. https://www.tandfonline.com/doi/abs/10.1080/00221309.1931.9918387

Lewin, K. (1935). *Dynamic theory of personality*. McGraw Hill.

Lewin, K. (1936). *Principles of topological psychology*. McGraw-Hill.

Lewin, K. (1938). *The conceptual representation and the measurement of psychological forces*. Duke University Press.

Lewin, K. (1952). Field theory and experiment in social psychology. In K. Lewin (Ed.), *Field theory in social science* (pp. 130–154). Tavistock. https://www.jstor.org/stable/2769418 (Original work published 1939)

Lewin, K. (1992). Law and experiment in psychology. *Science in Context*, *5*(2), 385–416. https://doi.org/10.1017/S0269889700001241 (Original work published 1927)

Lewin, K. (1999). Cassirer's philosophy of science and the social sciences. In M. Gold (Ed.), *The complete social scientist: A Kurt Lewin reader* (pp. 23–36). American Psychological Association. (Original work published 1949)

Lewontin, R. C. (2000). *The triple helix: Gene, organism, and environment*. Harvard University Press.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 199–249). Academic Press Inc. https://doi.org/10.1016/s0065-2601(02)80006-5

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review*, *5*(4), 356–368. https://doi.org/10.1177/1754073913489750

Lindquist, K. A. (2017). The role of language in emotion: Existing evidence and future directions. *Current Opinion in Psychology*, *17*, 135–139. https://doi.org/10.1016/j.copsyc.2017.07.006

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, *5*(1), 66–71. https://doi.org/10.1177/1754073912451351

Lindquist, K. A., Jackson, J. C., Leshin, J., Satpute, A. B., & Gendron, M. (2022). The cultural evolution of emotion. *Nature Reviews Psychology*, *1*(11), 669–681. https://doi.org/10.1038/s44159-022-00105-4

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology*, *6*, 444. https://doi.org/10.3389/fpsyg.2015.00444

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion? *Current Directions in Psychological Science*, *24*(2), 99–108. https://doi.org/10.1177/0963721414553440

Livio, M. (2009). *Is God a mathematician?* Simon & Schuster.

Locke, J. (1964). *An essay concerning human understanding*. Meridian Books. (Original work published 1690)

Loken, B., & Ward, J. (1990). Alternative approaches to understanding the determinants of typicality. *Journal of Consumer Research*, *17*(2), 111–126. https://doi.org/10.1086/208542

Lombrozo, T. (2009). Explanation and categorization: How "why?" informs 'what?". *Cognition*, *110*(2), 248–253. https://doi.org/10.1016/j.cognition.2008.10.007

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

Longino, H. E. (2001). *The fate of knowledge*. Princeton University Press.

Longino, H. E. (2019). The social dimensions of scientific knowledge. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2019/entries/scientific-knowledge-social/

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, *28*(1), 41–50. https://doi.org/10.3758/BF03211575

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*(2), 95–107. https://doi.org/10.1037/h0056029

MacEvoy, S. P., Hanks, T. D., & Paradiso, M. A. (2008). Macaque V1 activity during natural vision: Effects of natural scenes and saccades. *Journal of Neurophysiology*, *99*(2), 460–472. https://doi.org/10.1152/jn.00612.2007

Magnus, D. (2000). Down the primrose path: competing epistemologies in early twentieth-century biology. In R. Creath & J. Maienschein (Eds.), *Biology and epistemology* (pp. 91–121). Cambridge University Press.

Mahon, B. Z., & Hickok, G. (2016). Arguments about the nature of concepts: Symbols, embodiment, and beyond. *Psychonomic Bulletin & Review*, *23*(4), 941–958. https://doi.org/10.3758/s13423-016-1045-2

Mallon, R. (2019). Naturalistic approaches to social construction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/socialconstruction-naturalistic/

Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. W.W. Norton.

Mandler, G. (1990). William James and the construction of emotion. *Psychological Science*, *1*(3), 179–180. https://doi.org/10.1111/j.1467-9280.1990.tb00193.x

Marder, E., & Taylor, A. L. (2011). Multiple models to capture the variability in biological neurons and networks. *Nature Neuroscience*, *14*(2), 133–138. https://doi.org/10.1038/nn.2735

Markus, H. R. (2017). American = independent? *Perspectives on Psychological Science, 12(5),* 855–866. https://doi.org/10.1177/1745691617718799

Markus, H. R. (2021). Gender inequality and well-being: Concepts and their measures are cultural products—A reflection on Li et al. (2021). *Psychological Science, 32(6),* 952–954. https://doi.org/10.1177/095679762110182

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Markus, H. R., & Kitayama, S. (1994). The cultural shaping of emotion: A conceptual framework. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 339–351). American Psychological Association. https://doi.org/10.1037/10152-020

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press. (Original work published 1982)

Massimi, M., & McCoy, C. D. (2019). *Understanding perspectivism: Scientific challenges and methodological prospects*. Routledge.

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion*, *14*(3), 195–214. https://doi.org/10.1007/BF00995569

Maturana, H. R., & Varela, F. J. (1972). *Autopoiesis and cognition: The realization of the living*. Springer Science & Business Media.

May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age differences in memory. *Psychological Science*, *4*(5), 326–330. https://doi.org/10.1111/j.1467-9280.1993.tb00573.x

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.

Mayr, E. (1992). A local flora and the biological species concept. *American Journal of Botany*, *79*(2), 222–238. https://doi.org/10.2307/2445111

Mayr, E. (2004). *What makes biology unique? Considerations on the autonomy of a scientific discipline*. Cambridge University Press.

McDougall, W. (1921). *An introduction to social psychology* (14th ed.). John W. Luce & Company. (Original work published 1908)

McDougall, W. (1923). *An outline of psychology*. Methuen & Co. Ltd.

McGuire, W. J. (1973). The yin and yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology*, *26*(3), 446–456. https://doi.org/10.1037/h0034345

McIntosh, A. R. (2004). Contexts and catalysts: A resolution of the localization and integration of function in the brain. *Neuroinformatics*, *2*(2), 175–182. https://doi.org/10.1385/NI:2:2:175

McManus, R. M., Young, L., & Sweetman, J. (2023). Psychology Is a property of persons, not averages or distributions: Confronting the group-to-person generalizability problem in experimental

psychology. *Advances in Methods and Practices in Psychological Science*, *6*(3), 25152459231186615. https://doi.org/10.1177/25152459231186615

McNally, G. P., Johansen, J. P., & Blair, H. T. (2011). Placing prediction into the fear circuit. *Trends in Neurosciences*, *34*(6), 283–292. https://doi.org/10.1016/j.tins.2011.03.005

McVeigh, K., Kleckner, I., Quigley, K., & Satpute, A. (2023). Fear-related psychophysiological patterns are situation and individual dependent: A Bayesian model comparison approach. *Emotion*, *24*(2). 506-521. https://doi.org/10.1037/emo0001265

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*(2), 158–190. https://doi.org/10.1016/0010-0285(88)90018-7

Mellor, D. H. (1977). Natural kinds. *British Journal for the Philosophy of Science*, *28*(4), 299–312. https://doi.org/10.1093/bjps/28.4.299

Melnikoff, D. E., & Bailey, A. H. (2018). Preferences for moral vs. immoral traits in others are conditional. *Proceedings of the National Academy of Sciences*, *115*(4). https://doi.org/10.1073/pnas.1714945115

Melnikoff, D. E., & Strohminger, N. (2020). The automatic influence of advocacy on lawyers and novices. *Nature Human Behaviour*, *4*(12), 1258–1264. https://doi.org/10.1038/s41562-020-00943-3

Merton, R. K. (1979). *The sociology of science: Theoretical and empirical investigations* (N. W. Storer, Ed.). University of Chicago Press.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513

Mesquita, B. (2022). *Between us: How cultures create emotions*. W. W. Norton & Company.

Mesquita, B., Barrett, L. F., & Smith, E. R. (2010). *The mind in context*. Guilford Press.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, *112*(2), 179–204. https://doi.org/10.1037/0033-2909.112.2.179

Miller, L. C., Shaikh, S. J., Jeong, D. C., Wang, L., Gillig, T. K., Godoy, C. G., Appleby, P. R., Corsbie-Massay, C. L., Marsella, S., Christensen, J. L., & Read, S. J. (2019). Causal inference in generalizable environments: Systematic representative design. *Psychological Inquiry*, *30*(4), 173–202. https://doi.org/10.1080/1047840X.2019.1693866

Millikan, R. G. (1999). Historical kinds and the "special sciences." *Philosophical Studies*, *95*(1/2), 45–65. https://doi.org/10.1023/A:1004532016219

Mischel, W. (1968). *Personality and assessment*. John Wiley & Sons Inc.

Mobbs, D., Headley, D. B., Ding, W., & Dayan, P. (2020). Space, time, and fear: Survival computations along defensive circuits. *Trends in Cognitive Sciences*, *24*(3), 228–241. https://doi.org/10.1016/j.tics.2019.12.016

Müller, G. B. (2017). Why an extended evolutionary synthesis is necessary. *Interface Focus*, *7*(5), 20170015. https://doi.org/10.1098/rsfs.2017.0015

Murphy, G. (2004). *The big book of concepts*. MIT Press.

Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. Harcourt, Brace & World.

Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition*, *80*(3), B11-20. https://doi.org/10.1016/s0010-0277(01)00112-3

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, *69*(1), 511–534. https://doi.org/10.1146/annurev-psych122216-011836

Nelson, N. L., & Russell, J. A. (2016). A facial expression of pax: Assessing children's "recognition" of emotion from faces. *Journal of Experimental Child Psychology*, *141*, 49–64. https://doi.org/10.1016/j.jecp.2015.07.016

Newton, I. (2009). *Philosophiae naturalis principia mathematica*. University of Michigan Library. (Original work published 1687)

Newton-Smith, W. (1981). *The rationality of science*. Routledge.

Nickles, T. (2024). Scientific Revolutions. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2024/entries/scientificrevolutions/

Nisbett, R. E. (1980). The trait construct in lay and professional psychology. In L. Festinger (Ed.), *Retrospections on social psychology* (pp. 109–130). Oxford University Press.

Noble, R., & Noble, D. (2023). *Understanding living systems*. Cambridge University Press. https://doi.org/10.1017/9781009277396

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104–114. https://doi.org/10.1037/0278-7393.10.1.104

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

Oberheim, E., & Hoyningen-Huene, P. (2018). The Incommensurability of Scientific Theories. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/incommensurability/

Oreskes, N. (2003). The role of quantitative models in science. In C. D. Canham, J. J. Cole, & W. K. Lauenroth (Eds.), *Models in ecosystem science* (pp. 13–31). Princeton University Press.

Oreskes, N. (2019). *Why trust science?* Princeton University Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331. https://doi.org/10.1037/0033-295X.97.3.315

Panksepp, J. (1988). Brain emotional circuits and psychopathologies. In M. Clynes & J.

Panksepp (Eds.), *Emotions and psychopathology* (pp. 37–76). Plenum Press. https://doi.org/10.1080/026999399379195

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Pavlov, I. (2025). *Physiology of digestion*. Nobel Prize Outreach 2025. https://www.nobelprize.org/prizes/medicine/1904/pavlov/lecture/ (Original work published 1904)

Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, *12*(1), 1–15. https://www.peirce.org/writings/p107.html

Peirce, C. S. (1878). How to make our ideas clear. *Popular Science Monthly*, *12*(Jan.), 286–302. https://www.peirce.org/writings/p119.html

Peirce, C. S. (1998). Pragmatism as the logic of abduction. In N. Houser (Ed.), *The essential Peirce: Selected philosophical writings (1893-1913)* (pp. 226–241). Indiana University Press. https://www.jstor.org/stable/j.ctt16gz4vr (Original work published 1903)

Pepitone, A. (1976). Toward a normative and comparative biocultural social psychology. *Journal of Personality and Social Psychology*, *34*, 641–653. https://doi.org/10.1037/0022-3514.34.4.641

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001

Pezzulo, G., Zorzi, M., & Corbetta, M. (2021). The secret life of predictive brains: What's spontaneous activity for? *Trends in Cognitive Sciences*, *25*(9), 730–743. https://doi.org/10.1016/j.tics.2021.05.007

Pinker, S. (1995). *The language instinct*. Harper Perennial.

Platt, J. R. (1964). Strong inference. *Science*, *146*(3642), 347–353. https://doi.org/10.1126/science.146.3642.347

Pocheville, A. (2015). The ecological niche: History and recent controversies. In T. Heams, P. Huneman, G. Lecointre, & M. Silberstein (Eds.), *Handbook of evolutionary thinking in the sciences* (pp. 547–586). Springer Netherlands. https://doi.org/10.1007/978-94-017-9014-7_26

Popper, K. (2005). *The logic of scientific discovery*. Routledge. (Original work published 1959)

Popper, K. R. (1974). Scientific reduction and the essential incompleteness of all science. In F. J. Ayala & T. Dobzhansky (Eds.), *Studies in the philosophy of biology: Reduction and related problems* (pp. 259–284). Macmillan Education UK. https://doi.org/10.1007/978-1-349-01892-5_16

Popper, K. R. (2002). *Conjectures and refutations: The growth of scientific knowledge*. Psychology Press. (Original work published 1963)

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363. https://doi.org/10.1037/h0025953

Prather, A. A., Bogdan, R., & Hariri, A. R. (2013). Impact of sleep quality on amygdala reactivity, negative affect, and perceived stress. *Psychosomatic Medicine*, *75*(4), 350–358. https://doi.org/10.1097/PSY.0b013e31828ef15b

Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, *16*(4), 202–206. https://doi.org/10.1111/j.1467-8721.2007.00504.x

Putnam, H. (1975a). *Philosophical papers (Vol. 1: Mathematics, matter and method)*. Cambridge University Press.

Putnam, H. (1975b). *The meaning of "meaning."* In H. Putnam (Ed.), *Philosophical papers* (Vol. 2: Mind, language and reality, pp. 215-271). Cambridge University Press.

Putnam, H. (1981). *Reason, truth, and history*. Cambridge University Press.

Quigley, K. S., Gianaros, P. J., Norman, G. J., Jennings, R., J., Berntson, G. G., & de Geus, E. J. C. (2024). Publication guidelines for human heart rate and heart rate variability. Studies in psychophysiology—Part 1: Physiological underpinnings and foundations of measurement. *Psychophysiology*, *61*(9), e14604. https://doi.org/10.1111/psyp.14604

Quine, W. V. (1951a). On Carnap's views on ontology. *Philosophical Studies*, *2*(5), 65–72. https://www.jstor.org/stable/4318118

Quine, W. V. (1951b). Two dogmas of empiricism. *Philosophical Review*, *60*(1), 20–43. https://doi.org/10.2307/2266637

Quine, W. V. (1969). Epistemology naturalized. In W. V. Quine (Ed.), *Ontological relativity and other essays* (pp. 69–90). Columbia University Press.

Quine, W. V. (1977). Intensions revisited. *Midwest Studies In Philosophy*, *2*(1), 5–11. https://doi.org/10.1111/j.1475-4975.1977.tb00024.x

Quine, W. V., & Ullian, J. S. (1978). *The web of belief* (2nd ed.). McGraw-Hill.

Rabinovich, M. I., Tristan, I., & Varona, P. (2015). Hierarchical nonlinear dynamics of human attention. *Neuroscience & Biobehavioral Reviews*, *55*, 18–35. https://doi.org/10.1016/j.neubiorev.2015.04.001

Radulescu, A., Shin, Y. S., & Niv, Y. (2021). Human representation learning. *Annual Review of Neuroscience*, *44*(1), 253–273. https://doi.org/10.1146/annurev-neuro-092920120559

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Raykov, T., Dimitrov, D. M., Marcoulides, G. A., & Harrison, M. (2019). On true score evaluation using item response theory modeling. *Educational and Psychological Measurement*, *79*(4), 796–807. https://doi.org/10.1177/0013164417741711

Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., Jacob, Y., Atzil, S., Admon, R., Bleich-Cohen, M., Maron-Katz, A., Hendler, T., & Barrett, L. F. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(4), 709–723. https://doi.org/10.3758/s13415-016-0425-4

Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology*, *58*(6), 1048–1061. https://doi.org/10.1037/0022-3514.58.6.1048

Reardon, S. (2016). A mouse's house may ruin experiments. *Nature*, *530*(7590), 264–264. https://doi.org/10.1038/nature.2016.19335

Reddy, W. M. (1997). Against constructionism: The historical ethnography of emotions. *Current Anthropology*, *38*(3), 327–351. https://doi.org/10.1086/204622

Reddy, W. M. (2007, September 16). [Personal communication].

Reichenbach, H. (1938). *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. University of Chicago Press.

Reydon, T. A. C. (2016). From a zooming-in model to a co-creation model: Towards a more dynamic account of classification and kinds. In C. E. Kendig (Ed.), *Natural kinds and classification in scientific practice* (pp. 59–73). Routledge.

Reynolds, S. M., & Berridge, K. C. (2008). Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. *Nature Neuroscience*, *11*(4), 423–425. https://doi.org/10.1038/nn2061

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531. https://doi.org/10.1073/pnas.1208951109

Rice, C. (2013). Moving beyond causes: Optimality models and scientific explanation. *Noûs*, *49*(3), 589–615. https://doi.org/10.1111/nous.12042

Rice, C. (2016). Factive scientific understanding without accurate representation. *Biology and Philosophy*, *31*(1), 81–102. https://doi.org/10.1007/s10539-015-9510-2

Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331–363. https://doi.org/10.1037/1089-2680.7.4.331

Richerson, P. J., & Boyd, R. (2006). *Not by genes alone: How culture transformed human evolution*. University of Chicago Press.

Robbins, J., & Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly*, *81*(2), 407–420. http://www.jstor.org/stable/30052755

Roediger, H. L. (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology*, *59*(1), 225–254. https://doi.org/10.1146/annurev.psych.57.102904.190139

Rooney, P. (1992). On values in science: Is the epistemic/non-epistemic distinction useful? *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1992*(1), 13–22. https://www.jstor.org/stable/192740

Rosaldo, R. (1980). *Ilongot headhunting, 1883-1974: A study in society and history*. Stanford University Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/00100285(75)90024-9

Roseman, I. J. (2011). Emotional behaviors, emotivational goals, emotion strategies: Multiple levels of organization integrate variable and consistent responses. *Emotion Review*, *3*(4), 434–443. https://doi.org/10.1177/1754073911410744

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. McGraw-Hill.

Rovelli, C. (2021). *Helgoland: Making sense of the quantum revolution* (E. Segre & S. Carnell, Trans.). Riverhead Books.

Ruba, A. L., & Pollak, S. D. (2020). The development of emotion reasoning in infancy and early childhood. *Annual Review of Developmental Psychology*, *2*, 503–531. https://doi.org/10.1146/annurev-devpsych-060320-102556

Rubinov, M., & Sporns, O. (2011). Weight-conserving characterization of complex functional brain networks. *NeuroImage*, *56*(4), 2068–2079. https://doi.org/10.1016/j.neuroimage.2011.03.069

Ruphy, S. (2010). Are stellar kinds natural kinds? A challenging newcomer in the monism/pluralism and realism/antirealism debates. *Philosophy of Science*, *77*(5), 1109–1120. https://doi.org/10.1086/656544

Russell, J. A. (1991a). Culture and the categorization of emotions. *Psychological Bulletin*, *110*(3), 426–450. https://doi.org/10.1037/0033-2909.110.3.426

Russell, J. A. (1991b). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology*, *60*(1), 37–47. https://doi.org/10.1037/0022-3514.60.1.37

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141. https://doi.org/10.1037/0033-2909.115.1.102

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Salmon, W. C. (1981). Rational prediction. *British Journal for the Philosophy of Science*, *32*(2), 115–125. https://doi.org/10.1093/bjps/32.2.115

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61–71. https://doi.org/10.2307/2548836

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, *15*(60), 243–253. https://doi.org/10.2307/2549561

Satpute, A. B., & Lindquist, K. A. (2021). At the neural intersection between language and emotion. *Affective Science*, *2*(2), 207–220. https://doi.org/10.1007/s42761-02100032-2

Sattler, J. M. (1970). Racial "experimenter effects" in experimentation, testing, interviewing, and psychotherapy,. *Psychological Bulletin*, *73*(2), 137–160. https://doi.org/10.1037/h0028511

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review*, *3*(4), 444–454. https://doi.org/10.1177/1754073911410745

Schachter, S. (1959). *The psychology of affiliation: experimental studies of the sources of gregariousness*. Stanford University Press.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*(5), 379–399. https://doi.org/10.1037/h0046234

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). Oxford University Press.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*(7), 1307–1351. https://doi.org/10.1080/02699930902928969

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Schlick, M. (1985). *General theory of knowledge*. Open Court. (Original work published 1918)

Schwartz, S. P. (Ed.). (1977). *Naming, necessity, and natural kinds*. Cornell University Press.

Schwartz, S. P. (1980). Natural kinds and nominal kinds. *Mind*, *89*(354), 182–195. https://doi.org/10.1093/mind/LXXXIX.354.182

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *The Behavioral and Brain Sciences*, *21*(1), 1–17; discussion 17-54. https://doi.org/10.1017/s0140525x98000107

Scully, R. R., Basner, M., Nasrini, J., Lam, C., Hermosillo, E., Gur, R. C., Moore, T., Alexander, D. J., Satish, U., & Ryder, V. E. (2019). Effects of acute exposures to carbon dioxide on decision making and cognition in astronaut-like subjects. *Npj Microgravity*, *5*(1), 1–15. https://doi.org/10.1038/s41526-019-0071-6

Searle, J. R. (1995). *The construction of social reality*. Free Press.

Seeley, W. W. (2019). The salience network: A neural system for perceiving and responding to homeostatic demands. *The Journal of Neuroscience*, *39*(50), 9878–9882. https://doi.org/10.1523/JNEUROSCI.1138-17.2019

Seward, J. P. (1956). Drive, incentive, and reinforcement. *Psychological Review*, *63*(3), 195–203. https://doi.org/10.1037/h0048229

Shaffer, C., Barrett, L. F., & Quigley, K. S. (2023). Signal processing in the vagus nerve: Hypotheses based on new genetic and anatomical evidence. *Biological Psychology*, *182*, 108626. https://doi.org/10.1016/j.biopsycho.2023.108626

Shaffer, C., Westlin, C., Quigley, K. S., Whitfield-Gabrieli, S., & Barrett, L. F. (2022). Allostasis, action, and affect in depression: Insights from the theory of constructed emotion. *Annual Review of Clinical Psychology*, *18*(1), 553–580. https://doi.org/10.1146/annurev-clinpsy-081219-115627

Shannon, C., & Weaver, W. (1964). *The mathematical theory of communication* (10th ed.). The University of Illinois Press. (Original work published 1949)

Shapin, S., & Schaffer, S. (2011). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life* (Rev. ed.). Princeton University Press. (Original work published 1985)

Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? *Current Directions in Psychological Science*, *20*(6), 395–399. https://doi.org/10.1177/0963721411424739

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–1086. https://doi.org/10.1037/0022-3514.52.6.1061

Sherif, M. (1977). Crisis in social psychology: Some remarks towards breaking through the crisis. *Personality and Social Psychology Bulletin*, *3*(3), 368–382. https://doi.org/10.1177/014616727700300305

Sherman, M. (1927). The differentiation of emotional responses in infants. I. Judgments of emotional responses from motion picture views and from actual observation. *Journal of Comparative Psychology*, *7*(3), 265–284. https://doi.org/10.1037/h0073204

Sherman, M. (1928). The differentiation of emotional responses in infants. *Journal of Comparative Psychology*, *8*(5), 385–394. https://doi.org/10.1037/h0069981

Shweder, R. A., & Bourne, E. J. (1982). Does the concept of the person vary cross-culturally? In A. J. Marsella & G. M. White (Eds.), *Cultural conceptions of mental health and therapy* (pp. 97–137). Springer Netherlands. https://doi.org/10.1007/978-94-0109220-3_4

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Basic Books.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Authors Cooperative.

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A metaanalytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, *144*(4), 343–393. https://doi.org/10.1037/bul0000128

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. https://doi.org/10.1068/p281059

Sinaceur, M., & Tiedens, L. Z. (2006). Get mad and get more than even: When and why anger expression is effective in negotiations. *Journal of Experimental Social Psychology*, *42*(3), 314–322. https://doi.org/10.1016/j.jesp.2005.05.002

Singh, A., Westlin, C., Eisenbarth, H., Losin, E. A. R., Andrews-Hanna, J. R., Wager, T. D., Satpute, A. B., Barrett, L. F., Brooks, D. H., & Erdogmus, D. (2021). Variation is the norm: Brain state dynamics evoked by emotional video clips. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 6003–6007. https://ieeexplore.ieee.org/abstract/document/9630852/

Sismondo, S., & Chrisman, N. (2001). Deflationary metaphysics and the natures of maps. *Philosophy of Science*, *68*(3), S38–S49. https://www.jstor.org/stable/3080933

Skinner, B. F. (1990). *Lifetime scientific contribution remarks*. American Psychological Association, Boston. https://www.youtube.com/watch?v=Bf-GKbcSFNo

Slater, M. H. (2009). Macromolecular Pluralism. *Philosophy of Science*, *76*(5), 851–863. https://doi.org/10.1086/605817

Slater, M. H. (2015). Natural kindness. *British Journal for the Philosophy of Science*, *66*(2), 375–411. https://doi.org/10.1093/bjps/axt033

Smart, J. J. C. (1963). *Philosophy and scientific realism*. Routledge.

Smith, E. E. (1989). Concepts and induction. In M. Posner (Ed.), *Foundations of cognitive science* (pp. 501–526). MIT Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108–131. https://doi.org/10.1207/S15327957PSPR0402_01

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. https://doi.org/10.1016/S1364-6613(03)00156-6

Sober, E. (1980). Evolution, population thinking, and essentialism. *Philosophy of Science*, *47*(3), 350–383. https://doi.org/10.1086/288942

Sober, E. (1994). *From a biological point of view: Essays in evolutionary philosophy*. Cambridge University Press.

Sorge, R. E., Martin, L. J., Isbester, K. A., Sotocinal, S. G., Rosen, S., Tuttle, A. H., Wieskopf, J. S., Acland, E. L., Dokova, A., Kadoura, B., Leger, P., Mapplebeck, J. C. S., McPhail, M., Delaney, A., Wigerblad, G., Schumann, A. P., Quinn, T., Frasnelli, J., Svensson, C. I., … Mogil, J. S. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nature Methods*, *11*(6), 629–632. https://doi.org/10.1038/nmeth.2935

Spencer, H. (1894). *The principles of psychology*. D. Appleton & Company.

Spivey, M. (2007). *The continuity of mind*. Oxford University Press.

Sporns, O. (2011). The human connectome: A complex network. *Annals of the New York Academy of Sciences*, *1224*, 109–125. https://doi.org/10.1111/j.17496632.2010.05888.x

Srinivasan, R., & Martinez, A. M. (2021). Cross-cultural and cultural-specific production and perception of facial expressions of emotion in the Wild. *IEEE Transactions on Affective Computing*, *12*(3), 707–721. https://doi.org/10.1109/TAFFC.2018.2887267

Stanford, K. (2023). Underdetermination of scientific theory. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/scientificunderdetermination/

Steiner, I. D. (1966). Personality and the resolution of interpersonal disagreements. In B. A. Maher (Ed.), *Progress in experimental personality research* (Vol. 3, pp. 195–239). Academic Press

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, *106*(1), 5–15. https://doi.org/10.1016/j.physbeh.2011.06.004

Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. MIT Press.

Sullivan, J. A. (2016). Neuroscientific kinds through the lens of scientific practice. In C. Kendig (Ed.), *Natural kinds and classification in scientific practice* (pp. 47–56). Routledge. https://philarchive.org/rec/SULNKT

Systems Innovation (Director). (2017, May 17). *Complexity Theory Overview*. https://www.youtube.com/watch?v=i-ladOjo1QA

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–102. https://doi.org/10.1038/scientificamerican1170-96

Teigen, K. H. (2002). One hundred years of laws in psychology. *The American Journal of Psychology*, *115*(1), 103. https://doi.org/10.2307/1423676

Teller, P. (2018). Referential and perspectival realism. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, *9*(1), 151. https://doi.org/10.4245/sponge.v9i1.26990

Teller, P. (2019). What is perspectivism, and does it count as realism? In M. Massimi & C. D. McCoy (Eds.), *Understanding perspectivism: Scientific challenges and methodological prospects* (pp. 49–64). Routledge.

Theriault, J. E., Shaffer, C., Dienel, G. A., Sander, C. Y., Hooker, J. M., Dickerson, B. C., Barrett, L. F., & Quigley, K. S. (2023). A functional account of stimulation-based aerobic glycolysis and its role in interpreting bold signal intensity increases in neuroimaging experiments. *Neuroscience & Biobehavioral Reviews*, *153*, 105373. https://doi.org/10.1016/j.neubiorev.2023.105373

Tilak, S., Glassman, M., Kuznetcova, I., & Pelfrey, G. L. (2022). Applications of cybernetics to psychological theory: Historical and conceptual explorations. *Theory & Psychology*, *32*(2), 298–325. https://doi.org/10.1177/09593543211053804

Titchener, E. B. (1896). *An outline of psychology*. Macmillan.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Tomkins, S. S., & McCarter, R. (1964). What and Where are the Primary Affects? Some Evidence for a Theory. *Perceptual and Motor Skills*, *18*(1), 119–158. https://doi.org/10.2466/pms.1964.18.1.119

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). Oxford University Press.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*(4), 397–405. https://doi.org/10.1177/1754073911410747

Tronson, N. C. (2020). Uncertainty versus prediction error in Pavlovian fear conditioning: Commentary on Walker et al. (2019). *European Journal of Neuroscience*, *52*(5), 3485–3486. https://doi.org/10.1111/ejn.14578

Vallacher, R. R., Van Geert, P., & Nowak, A. (2015). The intrinsic dynamics of psychological process. *Current Directions in Psychological Science*, *24*(1), 58–64. https://doi.org/10.1177/0963721414551571

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*(1), 3–15. https://doi.org/10.1037/0033-295X.94.1.3

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences*, *113*(23), 6454–6459. https://doi.org/10.1073/pnas.1521897113

van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press. https://doi.org/10.1093/0198244274.001.0001

van Fraassen, B. C. (2001). Constructive empiricism now. *Philosophical Studies*, *106*(1–2), 151–170. https://doi.org/10.1023/a:1013126824473

van Fraassen, B. C. (2004). *The Empirical Stance* (Rev. ed.). Yale University Press.

Van Fraassen, B. C. (2010). Rovelli's world. *Foundations of physics*, *40*(4), 390–417. https://doi.org/10.1007/s10701-009-9326-5

Van Kleef, G. A., & Côté, S. (2007). Expressing anger in conflict: When it helps and when it hurts. *The Journal of Applied Psychology*, *92*(6), 1557–1569. https://doi.org/10.1037/0021-9010.92.6.1557

van Zomeren, M., Spears, R., Fischer, A. H., & Leach, C. W. (2004). Put your money where your mouth is! Explaining collective action tendencies through group-based anger and group efficacy. *Journal of Personality and Social Psychology*, *87*(5), 649–664. https://doi.org/10.1037/0022-3514.87.5.649

Varela, F. J., Thompson, E. T., & Rosch, E. (1992). *The embodied mind: Cognitive science and human experience*. MIT Press.

Vasilyeva, N., & Lombrozo, T. (2020). Structural thinking about social categories: Evidence from formal explanations, generics, and generalization. *Cognition*, *204*, 104383. https://doi.org/10.1016/j.cognition.2020.104383

Voorspoels, W., Storms, G., & Vanpaemel, W. (2013). Idealness and similarity in goal-derived categories: A computational examination. *Memory & Cognition*, *41*(2), 312–327. https://doi.org/10.3758/s13421-012-0252-y

Voorspoels, W., Vanpaemel, W., & Storms, G. (2011). A formal ideal-based account of typicality. *Psychonomic Bulletin & Review*, *18*(5), 1006–1014. https://doi.org/10.3758/s13423-011-0122-9

Vouloumanos, A., & Waxman, S. R. (2014). Listen up! Speech is for thinking during infancy. *Trends in Cognitive Sciences*, *18*(12), 642–646. https://doi.org/10.1016/j.tics.2014.10.001

Wang, Y., Kragel, P. A., & Satpute, A. B. (2022). *Neural predictors of subjective fear depend on the situation*. https://doi.org/10.1101/2022.10.20.513114

Waxman, S. R., & Gelman, S. A. (2010). Different kinds of concepts and different kinds of words: What words do for human cognition. In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts* (pp. 99–129). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199549221.003.06

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302. https://doi.org/10.1006/cogp.1995.1016

Weisberg, M. (2006). Water is Not H2O. In D. Baird, E. Scerri, & L. McIntyre (Eds.), *Philosophy of chemistry: Synthesis of a new discipline* (pp. 337–345). Springer Netherlands. https://doi.org/10.1007/1-4020-3261-7_18

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199933662.001.0001

Weisberg, M., Needham, P., & Hendry, R. (2019). Philosophy of chemistry. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/chemistry/

Welder, A. N., & Graham, S. A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development*, *72*(6), 1653–1673. https://doi.org/10.1111/1467-8624.00371

Westlin, C., Theriault, J. E., Katsumi, Y., Nieto-Castanon, A., Kucyi, A., Ruf, S. F., Brown, S. M., Pavel, M., Erdogmus, D., Brooks, D. H., Quigley, K. S., Whitfield-Gabrieli, S., & Barrett, L. F. (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in Cognitive Sciences, 27*(3), 246-257. https://doi.org/10.1016/j.tics.2022.12.015

Wikan, U. (1990). *Managing turbulent hearts: A Balinese formula for living*. University of Chicago Press.

Wilkins, J. S. (2009). *Species: A history of the idea*. University of California Press.

Wilson, R. A., Barker, M. J., & Brigandt, I. (2007). When traditional essentialism fails: Biological natural kinds. *Philosophical Topics*, *35*(1–2), 189–215.

https://doi.org/10.5840/philtopics2007351/29

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2015). Variety in emotional life: Within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social Cognitive and Affective Neuroscience*, *10*(1), 62–71. https://doi.org/10.1093/scan/nsu037

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, *49*(5), 1105–1127. https://doi.org/10.1016/j.neuropsychologia.2010.12.032

Winston, A. S. (1990). Robert Sessions Woodworth and the "Columbia Bible": How the psychological experiment was redefined. *The American Journal of Psychology*, *103*(3), 391–401. https://doi.org/10.2307/1423217

Winston, A. S., & Blais, D. J. (1996). What counts as an experiment? A transdisciplinary analysis of textbooks, 1930–1970. *The American Journal of Psychology*, *109*(4), 599–616. https://doi.org/10.2307/1423397

Witkower, Z., Rule, N. O., & Tracy, J. L. (2023). Emotions do reliably co-occur with predicted facial signals: Comment on Durán and Fernández-Dols (2021). *Emotion*, *23*(3), 903–907. https://doi.org/10.1037/emo0001162

Wittgenstein, L. (1973). *Philosophical investigations (3rd ed.)*. Macmillan.

Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, *16*, 49–74. https://doi.org/10.1146/annurev-clinpsy102419-125032

Wu, L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*(2), 173–189. https://doi.org/10.1016/j.actpsy.2009.02.002

Wundt, W. (1998). *Outlines of psychology*. (3rd ed. from 7th rev. German ed., C. H. Judd, Trans.) Wm. Englemann. (Original work published 1897)

Xu, F. (2002). Language and conceptual development: Words as essence placeholders. *Behavioral and Brain Sciences*, *25*(6), 704–705. https://doi.org/10.1017/s0140525x02540127

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Young, P. T. (1943). *Emotion in man and animal; Its nature and relation to attitude and motive*. Wiley.

Zelazo, P. D. (2013). Developmental psychology: A new synthesis. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology* (Vol. 1, pp. 3–12). Oxford University Press.

Zhou, F., Zhao, W., Qi, Z., Geng, Y., Yao, S., Kendrick, K. M., Wager, T. D., & Becker, B. (2021). A distributed fMRI-based signature for the subjective experience of fear. *Nature Communications*, *12*(1), 6643. https://doi.org/10.1038/s41467-021-26977-3

Zimmerman, C. A., Lin, Y.-C., Leib, D. E., Guo, L., Huey, E. L., Daly, G. E., Chen, Y., & Knight, Z. A. (2016). Thirst neurons anticipate the homeostatic consequences of eating and drinking. *Nature*, *537*(7622), 680–684. https://doi.org/10.1038/nature18950

Zirnsak, M., & Moore, T. (2014). Saccades and shifting receptive fields: Anticipating consequences or selecting targets? *Trends in Cognitive Sciences*, *18*(12), 621–628. https://doi.org/10.1016/j.tics.2014.10.002

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. https://doi.org/10.1017/S0140525X17001972

ENDNOTES

APPENDIX

Section 1: "Truth" Without Empirical Success

The most common example of "truth without empirical success" comes from the history of the science: the heliocentric view of the solar system as attributed to Copernicus (it was actually first proposed by ancient Greek astronomer and mathematician Aristarchus of Samos in the 3rd century BC). Compared to Aristotle's and Ptolemy's geocentric view, the Copernican view is closer to the current understanding of our solar system, but was nonetheless an empirical disaster for many years, in part because it contained aspects that were not observable at the time (Kuhn, 1957/1992). If Earth orbits the sun, as in the heliocentric view, then stars that are closer to Earth should appear to move in relation to distant stars whose positions appear fixed, making it seem as if the nearby stars change their position during different seasons of the year. This phenomenon is called stellar parallax. If Earth remained stationary, as in the geocentric view, then no stellar parallax would be observed. For many years, stellar parallax went unobserved because of a faulty assumption: Stars are much further away than was believed. As a result, the stellar parallax was much, much smaller than expected, and 16th century telescopes were not up to the task of detecting it. In addition, the heliocentric view required both mathematical and technical advances to improve its predictive and explanatory power, when compared to Ptolemaic theory which was worked out in great mathematical detail. A traditional scientific realist would have stuck with the Aristotelian view rather than take a chance on Copernicus, because to entertain a geocentric view is to ignore your own senses (Feyerabend, 1975/2010).

Section 2: Empirical Success Without "Truth"

Scientific theories can enjoy empirical success for some time yet turn out to be wrong in the end. Historical examples include miasmas (the so-called "bad air" that appeared at night, or vapors arising from rotting organic matter that were thought to cause disease), caloric (the imagined fluid (a compressible fluid or weightless gas) that passed from one body to heat another (Lavoisier, 1789/2009; for review, see Weisberg et al., 2019), luminiferous aether (the medium through which light waves supposedly traveled), and phlogiston (an imaginary element released into the air by burning objects). Scientists believed in these imaginary categories (and many others) for what they considered good, rational reasons — these concepts did some heavy lifting in major scientific theories of the time, allowing scientists to successfully explain and manipulate the world to some extent. Even today, Ptolemaic equations and Newtonian mechanics are still used, even though they contain incorrect descriptions of time, space, and mass.

For an example closer to home, classical conditioning is taught in every introductory psychology textbook, yet Pavlov's explanation for the phenomenon was most likely incorrect. In the typical telling, the 19th century physiologist Ivan Pavlov famously won the 1904 Nobel Prize in physiology for conditioning increased salivation and gut secretions in dogs when they heard a neutral sound (a ticking metronome) that had been previously paired with food. His explanation of the effect: The salivary and gut responses to food are unconditioned reflexes that become associated over time with the neutral sound, in effect creating a conditioned reflex (Pavlov, 1904/2025). In this explanation, the sound becomes a stimulus that elicits a response of increased secretions. The modern explanation: Pavlov's dogs did not learn to react to a sound by salivating. Their brains learned to expect food upon hearing the sound, and this expectation manifested as a preparation to consume and digest. There is growing evidence that a brain's most important job is anticipating metabolic events in the body and preparing to deal with those events before they arise (Sterling & Laughlin, 2015). This predictive process is known as allostasis (Sterling, 2012). Recent evidence suggests that Pavlovian conditioning is actually the result of predictive processing in the brain rather than the associative transfer of a reactive reflex to a new stimulus (e.g., Herry & Johansen, 2014; Kolada et al., 2023; McNally et al., 2011; Tronson, 2020).

Section 3: What Psychological Scientists Get Wrong About The Philosopher Karl Popper

The philosopher Karl Popper was strongly committed to the rationality of science, but in his view, only deductive inferences were rock-solid enough to be considered rational. Popper argued that good scientific practice had nothing to do with the steady accumulation of knowledge, which relied on inductive inferences. Instead, rationality was a combination of *conjecture* (a hypothesis that describes or explains something about a category that is part of the world) and *attempted refutation* (selection by testing through observation of, or interaction with, the environment). The conjecture should be bold and put the hypothesis at *significant risk*. Any attempt at refutation should be *sufficiently critical* that it can falsify the hypothesis. If observations are inconsistent with a hypothesis, then it's "game over" for that part of scientific concept and time for an improved conjecture. If a hypothesis is not refuted, a scientist should try harder.

Popper's logic on conjecture and refutation went like this: Observations alone cannot support inductive generalizations because verifications are always provisional. Categories can have infinite

instances, and the next one might disconfirm the hypothesis of interest. Therefore, observations can never prove a hypothesis true, no matter how many times observations are replicated. Observations can only *corroborate* a hypothesis (can be *consistent with* a hypothesis) but never *confirm* them. Popper also believed that only the fittest scientific concepts survive refutation (Godfrey-Smith, 2021, p. 85). But corroboration and fitness are not synonyms of confirmation. That a hypothesis repeatedly survives observational tests cannot increase confidence that the hypothesis is true. For example, Popper did not believe in *strong inference*: the notion that it is possible to pit a scientific concept (or part of it) against all the other competing concepts (or their relevant parts), eliminating alternatives to identify the last one standing (Platt, 1964). This is an inductive strategy that, as Popper pointed out, is doomed to fail because it is impossible to eliminate all viable alternatives. It was the philosopher Emil Lakatos, not Popper, who suggested that each time a theory successfully fends off a possible refutation, its predictive power is increased and progressively extends to cover new phenomena (Lakatos, 1976, 1970/1978). Popper never claimed that a hypothesis can be *supported* by observations. Falsifications, on the other hand, are definitive, as those as they are premised on something irrefutable, such as a scientific law.

All this to say, according to Popper, psychological science would not be considered rational. (Ironically, Popper received his PhD in psychology, not philosophy.) Hypotheses in psychology are never premised on scientific laws because there are no psychological laws. Moreover, psychological phenomena are complex: Some category instances might be infrequent and might be observed only rarely (the evolutionary biologist Ernst Mayr made a similar point about biology; Mayr, 2004). From a Popperian point of view, hypotheses that deal with such categories cannot be falsified, and therefore they cannot be tested scientifically. Psychological inferences must always trade in inductive logic. Psychological scientists must always work out just how improbable observations must be to falsify a hypothesis. Inductive inferences, even those guided by powerful statistical tools, are not grounded in the rock-solid deductive logic and therefore weaken (or destroy, depending on one's point of view) the main criterion for the rationality of science from Popper's point of view (Godfrey-Smith, 2021, p. 90).

Popper argued that what distinguishes science from other forms of human meaning-making is not its method, per se, but its attitude of skepticism. In Popper's view, scientific practice is rational if and only if a hypothesis can be refuted by observations that authentically put to the test. In science these days, the phrase "high risk, high reward" is a popular way to describe research that is highly innovative, but Popper believed that all scientific efforts should be high risk; otherwise, they are not rational (at least not with respect to building knowledge).

Section 4: Epistemic Problems In A Scientific Community: The Example Of Choice-From-Array

Consider the following puzzle from the scientific study of emotion. In the 1920's, a social psychologist named John Dashiell introduced a method for assessing whether young children recognized certain facial movements as expressions of particular emotion categories such as ANGER, SADNESS, FEAR, etc. (Dashiell, 1927, 1928). The goal was to ascertain whether those categories are innate. This method and its variations came to be the most frequently used developmental studies of emotion (Ruba & Pollak, 2020). Across a series of trials, children are presented with an emotion word (e.g., "sad"), a description of a behavior (e.g., "Suzie frowned and was crying"), or a photo of a person posing an exaggerated configuration of facial movements (e.g., a

pouting face), and they select a matching target (e.g., a photograph or emotion word) from a small array of choices. This method, more formally known as "choice-from-array," became the standard throughout the field, and a particular favorite for studies involving participants from small, non-industrial cultures who had relatively less exposure to Western cultural practices and norms (Gendron et al., 2018).

In a 2002 meta-analysis, 95% of published studies that were included used a choice-from-array method (Elfenbein & Ambady, 2002). Most presented participants with a stimulus (e.g., a photograph of a posed, caricatured configuration of facial movements, a photo of the eye region only, a computer generated facial configuration, a vocalization, a brief scenario, an emotion label) and across a series of trials, the participants' task was to select a matching target from a small array of choices (e.g., a set of words, a pair of photos, etc.). Studies using this method reliably produced respectable effect sizes across age groups and most cultural contexts. So, what's the problem?

Research conducted and published in the early 20th century generally observed that participants were highly variable in the emotions they perceived when viewing photographs of posed expressions alone, even when these expressions were Western stereotypes (e.g., smiling in happiness, frowning in sadness, scowling in anger, etc.). The expressions associated with each category of emotion appeared to have little reliability and specificity. This larger body of research gave rise to the hypothesis that emotions and their expressions were socially constructed and culturally variable, and studies of this hypothesis dominated the published research for several decades.

In the early 1960's, choice-from-array was resurrected, expressly for the purpose of reducing the variation observed in the prior decades of published research (Ekman & Friesen, 1971, p. 125). And so it has. More than 50 years of published evidence now indicates that choice-from-array constrains and harmonizes participants' responses, making them more consistent than they otherwise would be (for comprehensive reviews, see Barrett, 2022; Barrett et al., 2019; Gendron et al., 2018; Russell, 1994). Various choice-from-array, in combination with other common experimental design features such as a small sample of stimuli repeated across trials (DiGirolamo & Russell, 2017; Gendron et al., 2015; Gendron, Mesquita, et al., 2020; N. L. Nelson & Russell, 2016; for discussion, see Yarkoni, 2022), a small array of words to choose from (Crivelli, Jarillo, et al., 2016; Crivelli, Russell, et al., 2016; Crivelli & Fridlund, 2018; e.g., Doyle & Lindquist, 2018; Gendron et al., 2012; Lindquist, 2017; Lindquist, MacCormack, et al., 2015; Lindquist & Gendron, 2013; Satpute & Lindquist, 2021), and elaborate manipulation checks (Gendron et al., 2015) are sufficiently potent to make completely contrived expressions appear universal for novel emotion categories (Hoemann, Crittenden, et al., 2019). Yet choice-from-array, in combination with these other design features, continues to dominate in peer-reviewed, published papers, even in those studies that now use larger samples and sophisticated machine learning algorithms (e.g., Cowen et al., 2021).

What's the upshot? An entire field of study continues to be dominated by peer-reviewed, published papers that largely depend on versions of a specific experimental setup. The method appears to *test* the hypothesis that certain emotion categories are natural kinds but may instead *manufacture* that evidence. Remove any of the design elements from the standard choice-from-array setup and the results change. Empirical support for natural kinds of emotion categories pretty much disappears. Yet, scientists continue to design, conduct, and publish experiments that use this experimental setup. Borrowing the words of philosopher Peter Godfrey-Smith (2021, p. 95), ideas about the universality of emotions may "appear to have lots of contact with observations when in fact they only have kind of 'pseudo-contact' with observation, because they avoid all risks."

This is a serious problem that the entire scientific community must solve. Research from the anthropologist Bruno Latour provides food for thought. Latour embedded himself in working laboratories to observe the internal dynamics of scientific work, including the decisions that are required to transform certain knowledge claims into scientific knowledge (Latour & Woolgar, 1979/1986). Latour argued that socially agreed-upon scientific practices create scientific facts (see also, Shapin & Schaffer, 1985/2011) that then appear as if given by nature because the work of creating them is mostly hidden from plain sight. Not because scientists are dishonest, but because science is a set of cultural practices within a community of actors, performed within context with a system of culturally specific beliefs, values, traditions, and behavioral practices. Scientists, wrote Latour, through their practices and decisions, routinely focus on certain observations more than others to create knowledge claims. Experimental designs have a profound influence on what can (and can't) be observed. Habits of data analysis (e.g., focusing on group averages instead of individuals, removing outliers, etc.) actively shape how these observations become evidence to create knowledge claims. Scientists then build networks of support around certain knowledge claims, making them look "given" by nature, after which those claims become more difficult to question and easier to accept as facts.

An interesting aspect of Latour's view deals with the resolution of scientific disputes and controversies, if they are resolved at all. He described how scientists use concepts like "truth" and "nature" (and we would add "evolution") as rhetoric to "persuade others and mark out the terrain of reasonableness" (Godfrey-Smith, 2021, p. 169), building confidence in certain knowledge claims over others and effectively constructing networks of support to insulate them from critique, making them easier to accept and propagate. The resolution of scientific disputes does not determine what is real, but it does settle what scientists take to be real, which can delay beneficial course corrections for decades or longer (if history is any guide).

Section 5: Theoretical Kinds

Relational kinds are not natural kinds, but scientists sometimes treat them as if they were. Astronomers, for example, need to predict the complex behavior of moving, colliding gas molecules in a star. To make their task tractable, astronomers strip away some complexity and model a hypothetical natural kind called an "ideal gas" composed of randomly moving molecules that bounce off one another with no friction or other meaningful interaction. An ideal gas does not exist in nature. It's an example of the delicate balancing act between simplification and utility. It's also an example of a type of relational kind called a *theoretical kind* (Giere, 2010).

Theoretical kinds are always inherently relational because to make them useful, scientists must add details of their choosing; they must decide which features are important and which can be safely ignored. Such choices necessarily depend on scientists' concepts and goals, and even their values, in addition to the incentive structure and the larger political, cultural, and historical contexts in which scientists work.

A theoretical kind is the residue that remains when a relational kind is stripped of its varying contextual factors. This ontological maneuver effectively neuters the relational kind's inferential powers. A theoretical kind is not usable in real life or in science. It's a simplification or an abstraction — a fiction, if you will — waiting for the additional contextual detail that is required to give it any inferential power whatsoever.

Scientists routinely use theoretical kinds, including scientific models, experiments, and laws. For example, Newton's "universal" law of gravity (quotes added), which he derived inductively from empirical observation (Newton, 1687/2009), is very much an idealized statement. Here is succinct summary taken from the Wikipedia article on Newton's law of universal gravitation: "Every particle attracts every other particle with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between their centers." Or, in more modern language, from the same article: "Every point mass attracts every other point mass by a force acting along the line intersecting the two points. The force is proportional to the product of the two masses, and inversely proportional to the square of the distance between them." Like all scientific laws, Newton's is a simplification by design to make it as general and applicable to as many different situations as possible. Such laws are templates for building more precise models of the world (Cartwright, 1983) from which it is possible to derive further hypotheses, predictions, and experiments.

It's also surprising to learn that many presumed natural kind categories are in fact better understood as theoretical kinds. Two notable examples are water (i.e., H_2O) and the chemical elements of the periodic table.

Water As A Theoretical Kind

Most people know the category WATER as collections of molecules composed to two hydrogen atoms and one oxygen atom or H_2O. If WATER is a natural kind category, then H_2O is the intension of the concept for WATER and should be responsible for the descriptive features of WATER that we experience, as well as its causal impacts, such as its transparency, its low viscosity, its ability to quench thirst and freeze into solid form at 0^o C, etc. But here's the thing; water is not H_2O in reality. H_2O is an idealization (Chang, 2012; Weisberg, 2006) of the stuff we drink, we bathe in, and that fills lakes and oceans. And the molecule H_2O does not alone cause the observable properties and behaviors of water as science currently understands them.

A glass of water certainly contains H_2O molecules, but it is more than H_2O. Water's observable behaviors, the features that make it drinkable, freezable, suitable for bathing, and so on, depend on an ongoing, complex molecular drama. A sample of water also contains ions, which are atoms or molecules that have an electric charge because they have lost or gained electrons. There are also many hydrogen-bonded polymers of H_2O molecules and of its ionized products. Let's not forget that natural samples of water also contain some concentration of bacteria and other microorganisms, as well as sodium, which determines whether or not it is drinkable by humans. All chemical elements have two or more versions called isotopes, which contain the same number of protons as their variants in the periodic table but different numbers of neutrons in their nuclei. Most isotopes of the same element differ in their physical features (their weight differs), but their behavior remains the same for the most part. And finally, it is not just the identity and quantity of atoms that compose a molecule (the chemical formula) that is necessary to predict its behavior, but also their arrangement in relation to one another. Molecules with the same chemical formula but with distinct structures are called isomers, and H_2O has two isomers with different profiles of chemical reactivity (Kilaj et al., 2018).

The fact that there are important explanatory connections between the continually fluctuating, complex concoction of molecules, ions, plus other bits and bobs in water on the one hand, and the features and the causal behavior of water on the other, means that it is a simplification to define water as H_2O.

Chemical Elements As Theoretical Kinds

The chemical elements of the periodic table are often presented as the best example of natural kind categories in science because each category has an essential feature: the number of protons in the element's nucleus, corresponding to its atomic number. Element categories are organized according to this number. Each category is distinctive, seeming to cut nature at its joints: Adding or removing one proton to the instances of a category fundamentally alters their chemical properties, meaning the boundaries between elements are discrete and represent real discontinuities in the world.

Nevertheless, the number of protons in the nucleus of an atom is insufficient to predict and explain its behavior in the physical world. Additional contextual features are needed to precisely predict or explain the physical properties of any sample of any element in the real world. Some of these contextual features are internal to the element, such as the state of excitement of its electrons, protons, or neutrons. Some are external, such as temperature, pressure, and any other substances mixed in (i.e., impurities). (For an instructive example of the element gold, see Bursten, 2018.) In philosophy of science, this means the proton number is not functioning like an essential feature of the category.

The upshot is that elements of the periodic table are idealizations. When a scientist speaks of a property of an element (like the melting point of gold or the boiling point of water), there are always hidden conditionals ("the melting point of gold when x, y, and z conditions are in place"; "the boiling point of water when x, y, and z are true"). The moment scientists move away from these idealizations and consider contextual factors that allow generalization, they are dealing with a complex system. If chemists designed experiments based on the idealizations, treating contextual features as mere moderators of universal phenomena, they'd experience failures of replication on a massive scale.

Section 6: Pragmatism

Pragmatism is a philosophy of science that formally emerged in the late 19th and early 20th century. It redefined truth in practical terms rather than focusing on the representation of a mind-independent reality. Pragmatism is commonly misunderstood as simple utility — if something works, it counts as knowledge. But for pragmatists, not just any run-of-the-mill utility will do. Pragmatism is utility with a specific goal in mind. Whatever brings us closer to that goal counts as knowledge.

The pragmatist philosopher Charles Sanders Peirce (Peirce, 1877, 1878) believed in a mind-independent reality and defined "true knowledge" practically as that which is certain. Whose certainty is important? Everyone's. Every person who could ever investigate the concept under ideal circumstances, if scientific inquiry continued for infinite time given unlimited resources. When this community of scientists comes to consensus over a concept, then the concept is permanently settled and counts as justified knowledge. Truth is permanently settled consensus that removes doubt about natural kind categories. The scientific method is not necessary to achieve true knowledge, he thought, but it's *useful* for gradually converging on a stable, consensual set of concepts that is resistant to doubt.

Peirce seems to be endorsing a consensus theory of reality. A category is real when there exists a permanently settled (i.e., idealized) consensus for its concept. (This view foreshadows Longino and

others who argue that objective knowledge necessarily requires a consensus among diverse people, with varied backgrounds and experiences, who differ in their assumptions and beliefs; see Appendix, Section 7.) Simplistically: Truth, as permanently settled consensus, is reality. Reality is not mind-independent per se but depends on a broad consensus of minds in the long run. Reality can be independent of anyone's personal concepts; what matters here are concepts that are stable and coherent in consensus with others.

William James also contributed to the philosophy of pragmatism. "We carve out everything," James states, "just as we carve out constellations, to serve our human purposes" (James, 1907, p. 98). When James considered categories to be real, he was referring to the system of stable and coherent categories that we construct in our experience, as a combination of pure experience and concepts, that allow us to act in an efficient, goal-directed way.

John Dewey was perhaps the most prominent pragmatist philosophy of the early 20th century (for an overview, see Hildebrand, 2024). A cornerstone of his philosophical approach, called instrumentalism, suggests that a concept's scientific value is determined solely by its utility for making accurate empirical predictions (e.g., Dewey, 1900, 1903). Whether the concept corresponds to reality or not (i.e., whether or not it is "true") is beside the point. Instrumentalism has its roots in the 16th century, introduced by the clergyman Andreas Osiander in his preface to Copernicus's book: Theories are merely calculating tools (Godfrey-Smith, 2021, p. 20).

Section 7: A Social Account Of Objectivity

Consensus is not objectivity. Human agreement does not make something objectively real. But in the philosophy of science, there is a logical reason to redefine objectivity as reduction of bias rather than as mind-independence. Science is a human activity that operates in a social context. What counts as knowledge in any domain of science depends on agreements about which questions are admissible, which methods count as acceptable tools of inquiry, which observations count as evidence, and which inferences are deemed defensible.

To move a science in the direction of objectivity, one must take the social nature of science seriously. The notion of scientific consensus is much more powerful (and daresay, more objective) when it emerges from a diverse and self-critical scientific community of experts (Oreskes, 2019). Any scientist's beliefs are necessarily influenced by their background and experiences, so the only way that consensus begets objectivity is to increase the diversity of the scientific communities. This point of view, called "standpoint epistemology" (Harding, 1986, 1992; for review, see Godfrey-Smith, 2021, pp. 173–184) emphasizes the ways in which a scientist's background, culture, status and power in the world, and even their physical constitution makes certain experiences more or less likely, and in effect situates them to make discoveries that are not available to others whose lives are situated differently. It permits the sampling of perspectives that might not otherwise be considered. A more diverse community will share some commitments but not others, and de facto bring more perspectives to bear on an issue (Harding, 1986, 1992; Longino, 1990). A larger pool of perspectives makes for more valid scientific findings. Objectivity, in this case, is not truth per se, but something closer to "critically achieved consensus of the scientific community" (Oreskes, 2019, p. 55).

Consensus is far less valuable, and undermines claims of objectivity, when it arises among like-minded people who share an uncritical acceptance of the same goals, interests, values, norms, and beliefs. Such homogeneous consensus increases the likelihood that certain concepts become reified

and masquerade as knowledge (particularly those that benefit said scientists). It's also more consistent with the original Enlightenment notion of science that suggests a more dispassionate ideal for scientific practice, in which the personal attributes and social background of a scientist are irrelevant to the scientific value of their concepts and efforts (e.g., Merton, 1979; Godfrey-Smith, 2021, p. 153). Standpoint epistemology, in contrast, suggests that epistemic values regarding the best practices for evidence, inference, and so on, which contribute to reliable and trustworthy knowledge of reality, will be served by a non-epistemic value of *diversity of point of view* (for a discussion between the two categories of values, see Rooney, 1992).

Furthermore, "for science to be self-correcting, scientists with diverse views must question, challenge, and critique each other through the mechanisms and avenues of professional practices," the epistemic value from philosopher Helen Longino that is discussed in the main text, called "transformative interrogation" — the collective give-and-take of critical examination (Longino, 1990). The community must be open and responsive to inconsistencies, disagreements and anomalies. Ideally this process allows a scientific product, as a whole, to aim at objectivity even when individual scientists don't. Transformative interrogation depends on both the cooperation and competition of scientists — getting along to get ahead (D. L. Hull, 1988). To remain open to criticism and feel comfortable offering it publicly, community members must trust in the intentions that motivate others' caution and skepticism. Trust is not a matter of goodwill per se — it is a special kind of self-interest (D. L. Hull, 1988) that improves one's scientific work.

Transformative interrogation breaks down when trust erodes, such as when a goal for recognition (a non-epistemic value) comes to outweigh epistemic values. This situation is usually characterized as a conflict between a glory-seeking scientist versus a knowledge-seeking scientific community, but this is an oversimplification, since the observations of a single scientist cannot become justified scientific knowledge unless it's used and cited by other scientists. Simply put, when a field fails to be self-correcting, the cause may lie not only with individual members but also in the social character of science. "Social character" refers to the social practices and procedures of adjudication that transform knowledge claims into knowledge, or that identify anomalies and correct for them, e.g., whether the incentive structure of science rewards or even allows such openness, whether the recognized avenues for criticism and evaluation include enough diversity, etc. New ideas and evidence may threaten past scientific achievements or the reputations and career status of the associated scientists (Kuhn, 1962/1970). In the understated words of Oreskes (2019, p. 60), "When people's life work is being questioned, they may get testy." Of course, there is still a role for individual scientists to be self-critical, to do unto themselves as they would do unto others.

Ultimately, scientific truth — the claims that scientists treat as knowledge — is a social product arising from the consensus of scientists, and so greater diversity in scientific communities, at all levels, makes scientific products more enduring and trustworthy. In the words of Longino,

> the greater the number of different points of view included in a given community, the more likely it is that its scientific practice will be objective … [and] it will result in descriptions and explanations … that are more reliable …. than would otherwise be the case. (Longino, 1990; quoted in Oreskes, 2019, p. 54).

In discussing Longino's proposal the historian Naomi Oreskes writes, "diversity does not heal all epistemic ills, but [all things being equal] a diverse community that embraces criticism is more likely to detect and correct error than a homogenous and self-satisfied one" (Oreskes, 2019, p. 54).

This more social account of scientific objectivity contrasts sharply with the "great man" view of science, which emphasizes the contributions and foibles of individual scientists to the ultimate value of scientific products. Individual values and beliefs are only a threat to objectivity if scientific practices and scientific knowledge are thought to arise from individual contributions to begin with.

When a scientific community lacks diversity and self-criticism, consensus can be evidence of bias. This can be compounded by the training of students and by the institutions of science (conferences, workshops, societies, peer reviewed journals, books), which together create a shared scientific system of concepts, as well as observations that are interpreted in accordance with these concepts. Again, from Longino,

> When, for instance, background assumptions are shared by all members of a community, the acquire an invisibility that renders them unavailable for criticism. They do not become visible until individuals who do not share the community's assumptions can provide alternative explanation of the phenomena without those assumptions. (Longino, 1990; quoted in Oreskes, 2019, pp. 53–54).

The very same bias can result in a lack of curiosity and open-mindedness to critical evidence, resulting in something more like ideology than science.

Section 8: Ad Hoc Categories

Ad hoc categories are constructed on-line. They are not stable across time and situation and individuals, out there in the world to be learned and retrieved when needed but are variable groupings of similar instances (for a review, see Mesquita et al., 2010; Hoffman et al., 2018; Schyns et al., 1998; Spivey, 2007; Yee & Thompson-Schill, 2016). Ad hoc categories are constructed on the spot, dynamically, using conceptual combination, in which the brain remembers and generatively combines bits and pieces of past events. Ad hoc categories are built in a brain as needed — a brain constructs them to serve a purpose in a specific context. The most important features of equivalence in an ad hoc category pertain to the perceiver's goal for the category (the function that the category serves) in a specific situation. Instances have a similar function in a given situation rather than physical similarities.

To make the discussion of an ad hoc category less cumbersome, consider a category from everyday life, such as the category CAT. Cats are similar in various ways… they have whiskers and pointy ears, they purr, they are carnivores, they share a certain genome — this is a partial list of physical features that cats share. But cats also vary from one another in their physical characteristics. They come in various sizes, with hair that varies in thickness and length (and some cats are hairless). They vary in eye color, hair color (for those that have hair), whisker length, tail length, and so on. It's tempting to think that a category like CAT has stable, context-independent core features of similarity, but this is an illusion created by commonalities across the various instances of situated CAT categories (discussed in Casasanto & Lupyan, 2015). To assume that CAT is a natural or empirical kind requires summarizing over all this variation to create a single concept that would be stored in memory to be retrieved when needed. And as Darwin observed in *On the Origin of Species*, individuals in any species category have plenty of variation. One of Darwin's great insights was that any single list of features that attempts to summarize the entire category in a universal way is a theoretical abstraction (see Mayr, 2004) and works like a stereotype (Barrett, 2022). In this ontological mindset, variation is real and inextricably linked to situated function.

In everyday life, cats are encountered in situations, never in isolation. A cat might be a house pet, an animal to be admired in the zoo, a mouse-control device, or even a human with a reputation for being a smooth operator. In each situation, a human brain backgrounds certain features and foregrounds others to tailor the category CAT to a specific goal. A category of cats that make ideal pets might be soft, cuddly, and purr softly when stroked. Cats who are successful mousers will be agile and quick perhaps a bit aggressive. Cats who draw visitors to a zoo will large and majestic. Cool cats, regardless of their biological species, are attractive, aloof, and a touch dismissive. And so on. The category *cat* is therefore not something out there in some mind-independent world, referring to all possible past and future instances of the word "cat." A human brain creates a category for CAT in a specific context. The features of similarity are linked to the function the category serves in that specific instance. And the features of equivalence change depending on the requirements of the situation because the function of CAT changes, meaning the concept for CAT changes with and depends on the requirements of the situation. The most important similarities that form a category CAT are not physical but functional.

Notice, here, that the line between concepts and categories is dissolved. Relational kind categories do not exist in the world, separate from the concepts in a human brain. Functional features only exist in the brain of a perceiver. They are computed as multimodal compressed summaries of sensory and motor features, meaning that they can correspond to different patterns of sensory and motor properties. To make sense of the continuous flow of signals in the world, the brain reinstates past signal ensembles that are functionally similar to the present signals, in effect "placing a boundary" to create a category. This is how a brain generalizes from the past for the purposes of dealing with the present (Gershman et al., 2010). So, a category is not solely "in the world." It is not solely "in the mind" either. It is a grouping whose equivalence is created by the brain. So relational kind categories, like concepts, exist in the brain. They only exist when they are assembled for use.

If the category CAT is assumed to be a natural kind, then there should be exactly one universal intension for the concept, consisting of the core features of equivalence that makes something a cat. The word "cat" is assumed to pick out all instances of cats (the extension of cats) in all situations. But if CAT is an ad hoc category, then there is one intension for "cat" and one grouping of extensions for each *situation* in which CAT is constructed.

Correspondingly, a word that names a concept never refers to all possible members of the corresponding category that ever existed and will exist, all with some universal features of equivalence in a single, stable distribution of possibilities. The graded, probabilistic feature structure of the category is context dependent. In a sense, an ad hoc category is a momentary, situated prototype, which is the ideal instance that best meets the function in that context (Borkenau, 1990; Chaplin et al., 1988; Loken & Ward, 1990; Lynch et al., 2000; Read et al., 1990; Voorspoels et al., 2011, 2013); the category prototype can change for categories whose share the word that are constructed on different occasions (e.g., the best instance of a "cat" will depend on the function that the category serves in the relevant situation). The categories and their situated prototypes can be similar across instances, but they are always constructed, not retrieved, in a specific situation.

The implication is that every category is a situated event with no static, perceiver-independent prototype (for more on the view of situated conceptualization as a process, see Barsalou, 1987, 2010; Casasanto & Lupyan, 2015; Spivey, 2007). Its features of equivalence are always functional, constructed by a particular perceiver for a particular function in a particular situation. The summary of any ad hoc, situated category is analogous to a prototype that best suits the functional goal of the categorizer in that specific situation (Barsalou & Hale, 1993; Voorspoels et al., 2011).

An important aspect of ad hoc categories is that they can be invented on the fly as needed by conceptual combination, combining bits and pieces of past events into new categories (Barsalou, 2017; J. A. Hampton & Winter, 2017; Werning et al., 2012). Scientists consider conceptual combination to be one of the most powerful abilities of the human brain. It's not just for making novel concepts on the fly. It is the normal process by which a brain constructs concepts. There are different accounts of how this construction proceeds, including standard information processing accounts (e.g., Barsalou, 1991, also reviewed in 2021) and predictive processing accounts (Barrett, 2017c, 2022) that are compatible with grounded cognition and the use of simulations (Barsalou, 1999, 2009; Barsalou et al., 2003; Lebois et al., 2020; also reviewed in Barsalou, 2021).

Section 9: An Example Of Relational Kinds In Psychology

In *Naming the Mind*, the historian Kurt Danziger tells the story of when he visited a university in Indonesia with the intention to teach a course in psychology (Danziger, 1997). He was raised and educated in the West on a steady diet of folk psychology topics that organize introductory psychology textbooks. On his arrival, he learned that a course already existed with "roots in Hindu philosophy with Javanese additions and reinterpretations," a mindset that was new to him. Danziger suggested to the current instructor that they teach joint seminars to "explain our approach to the same set of psychological topics, followed by an analysis of our differences". He reasoned, "if both my Indonesian colleague and I were dealing with psychological reality, there ought to be some points of contact, even convergence, between our domains" (all quotes from Danziger, 1997, p. 1).

To Danziger's utter surprise, the two instructors could not agree on a single topic. "For instance, I wanted to discuss the topic of motivation," wrote Danziger. His counterpart, however,

> said that would be quite difficult… because from his point of view motivation was not really a topic. The phenomena that I quite spontaneously grouped together as 'motivational' seemed to him to be no more than a heterogeneous collection of things that had nothing interesting in common. (p. 1).

"I tried other topics," lamented Danziger,

> intelligence, learning, and so on [but] my colleague would not recognize any of them as domains clearly marked off from other domains. He granted that each of them had some common features, but he regarded these features as trivial or as artificial and arbitrary" (p. 2).

In the other direction, the psychology topics suggested by his Indonesian colleague "were not only unfamiliar to me, I found his description of them hard to follow. They did not seem to me to constitute natural domains" (p. 2). Ultimately, they abandoned the idea of joint seminars.

Danziger was describing the Western and Indonesian psychologies as *incommensurate* (Feyerabend, 1962; Kuhn, 1962/1970; Oberheim & Hoyningen-Huene, 2018). In philosophy of science, two frameworks are incommensurate when their categories are incomparable and there is no clear way to translate from one to the other. Scientists working within incommensurate frameworks are speaking different scientific languages.

If Danziger were a traditional scientific realist, he would have assumed that his Western categories were the correct ones, prescribed by genes as part of a universal human nature. He might have

argued that the Indonesian categories were mere cultural variants on a common core set of universal categories (for an example of this style of argument, see Cordaro et al., 2018) or combinations of the more basic and true Western categories, or simply that they were wrong. His Indonesian counterpart could have made analogous arguments about the Western categories.

But that's not what happened. Instead, Danziger realized that

> it was clearly possible to carve up the field of psychological phenomena in very different ways and still end up with a set of concepts that seemed quite natural, given the appropriate cultural context. Moreover, these different sets of concepts could each make perfect practical sense. (Danziger, 1997, p. 2–3).

In experimental settings, these concepts "could be employed to produce specific psychological phenomena as reliably — and perhaps more reliably — as many of our psychological experiments" (p. 2). In effect, Danziger began to think about psychological categories as relational kinds.

Section 10: The Problem Of Trivial Effect Sizes

A challenging question for psychological scientists is whether the range and degree of empirical successes achieved are sufficient to support the "no miracles" argument. Psychological experiments tend to produce modest effect sizes that are statistically better than random chance, but entirely inadequate to support real-world successes on the order of the Golden Gate Bridge and the COVID-19 vaccine. A constant tension throughout the history of psychological science is whether or not to be satisfied with this state of affairs. In psychology, at best, it's a "glass half empty vs. half full" sort of situation.

From the "glass half full" perspective, we find psychological scientists who are satisfied with modest effect sizes and vigorously defend them. In the science of emotion, Witkower, Rule & Tracy (2023) is one example of a tendency to equate deviations from chance with empirical success, which is exactly the point raised by those who fall on the "glass half empty" side of the discussion (e.g., Barrett, 2022; Cronbach, 1975; Funder & Ozer, 2019; Gergen, 1978; Heft, 2001; McGuire, 1973; Mischel, 1968; Nisbett, 1980; Zelazo, 2013).

Witkower et al. published a critique of several meta-analyses (Durán & Fernández-Dols, 2021) that examined how people spontaneously moved their faces when angry, sad, afraid, or experiencing instances of other emotion categories. The proposed prototypic expressions, in their full configuration, were observed at proportions greater than chance levels, but only just. A mere 13% of almost 4,000 participants across 69 experiments moved their facial muscles in a manner that was predicted for that category (scowling when angry, frowning when sad, etc.), and only 23% of those participants made bits of the expected expressions. Taking the intensity of the emotion into account did not improve the overall correspondence between instances of a given emotion category and the frequency with which the hypothesized facial expression was observed, $r = .30$, 95% CI [.18, .41]. Durán and Fernández-Dols interpreted these results as evidence that "basic emotions do not reliably co-occur with their predicted facial signal" (Durán & Fernández-Dols, 2021, p. 1550). Witkower et al., however, claimed that "this conclusion is at odds with the data presented." (Witkower et al., 2023, p. 903). Their logic? The correlation between emotion experience and expected expression, $r = .30$, was a medium effect by the standards of Cohen (1988), a classic text used to educate psychological scientists. A Pearson's correlation of $r = 0.30$ corresponds to a

coefficient of determination (or the relative information shared between two measured variables) of $r^2 = 0.09$.

Consider the implications: A moderate effect size is defined as the relationship between two variables, where knowing the values of one leaves more than 90% of the variance in the other unexplained. This meager standard is enshrined as institutional knowledge in the science of psychology. Witkower et al. went on to note that this effect size was "substantially larger (nearly 150%) than the average effect size across personality and social psychology (r = .21; Richard et al., 2003; also see Fraley & Marks, 2007; Gignac & Szodorai, 2016)" and "larger than roughly 76% of all published meta-analytic effects in social psychology (Gignac & Szodorai, 2016; Richard et al., 2003)" (both quotes from p. 904). Their interpretation: "we view the observed association as medium-to-strong in size" (p. 905). Talk about setting a low bar for empirical success.

(To be fair, Witkower et al. (2023) also included another pair of meta-analyses showing that, for example, 41% of the almost 4,000 participants moved their faces as expected when they were amused, meaning that knowing the intensity of people's amusement predicted just over 16% of the variance in their facial movements.)

Section 11: Justified Knowledge Is Under-Determined By Evidence

The choice of one scientific concept over another is almost always under-determined by the available evidence. This is just a fancy way of saying that any hypothesis can be supported by a variety of evidence depending on how the hypothesis is particularized with detail, and any collection of observations will always be consistent with more than one hypothesis. Concepts remain successful long after there is evidence to disconfirm them, and concepts that are more in line with observable evidence often take decades (or centuries) to be taken seriously. The takeaway: Preference for one hypothesis (or the concept from it was derived) versus another is influenced by factors other than its fit to observations.

Every hypothesis is an idealization. Before a scientist can test a hypothesis, they must contextualize it with various constraints and assumptions to create a prediction. In the discussion of the hypothetico-deductive method, the deductive transformation from hypothesis to prediction is conditioned on *additional details* of design decisions, such as the circumstances in which the observations (i.e., experiences) take place (usually referred to as "initial conditions"), the assumptions entailed in the measurements, and other features of the study. In other words, it is never possible to test a simple hypothesis (e.g., "Is the category fear a basic emotion?"). Instead, when a scientist compares a hypothesis to the experiences they create within a scientific experiment, they are actually testing a *larger web of relations* in which the hypothesis is nested (e.g., "Does fear give evidence of basicness in this particular experimental setup, when we sample in this particular way, with these particular subjects," etc.). This is called the Duhem-Quine thesis (Duhem, 1954/1991; Quine, 1951b). Observations emerge from the whole web, and therefore a hypothesis is always evaluated relationally in the context of this web, an idea known as *confirmation holism* (Quine, 1951b). The holistic nature of hypothesis testing makes the rationality of science hard to defend. A hypothesis, by itself, never has observable consequences that can be used to test it. This weakens the rationality of the scientific method.

When Quine articulated his notion of a web, he was thinking about how to verify the meanings of sentences, like a hypothesis formalized in the text of a published scientific article, and the meanings

of observations in relation to the hypothesis being tested. Holism means that a word's semantics are interdependent with those of all other words in a language and are intimately bound up with their use.

The holistic nature of hypothesis testing opens the door to a pernicious problem in the practice of science: The web of (often invisible) assumptions increase the likelihood of disconfirmation dilemmas (Greenwald & Ronis, 1981), because any hypothesis can be retained by simply altering the contextual assumptions ad hoc. Logically, a scientist can continue holding a particular hypothesis in the face of potentially falsifying observations, because it is always possible to blame the failure on something else and reject any disconfirming observations. Popper called this "patching" a hypothesis. But if one can make a hypothesis predict any set of observations merely by adding an assumption or two, then strictly speaking, this further compromises the rationality of the scientific method.

Or perhaps the method is still rational, but that rationality that is aimed at something other than truth. When a hypothesis is patched too many times, its corresponding concept becomes something else: an ideology. A hypothesis is only scientific if it can, in principle, be rejected when inconsistent with observations.

Other evidence from the history of science gives further credence to the claim that scientific decision-making is rarely guided by empirical evidence alone. Historians have questioned how often paradigm shifts in science are necessarily preceded by a crisis in the way that Kuhn described (Godfrey-Smith, 2021, pp. 117, 126) and instead point to other reasons for dramatic shifts in scientific focus. Newtonian physics was done in by the early 20th century revolution in physics, but the "chemical revolution" may not be described as a crisis followed by a revolution; instead, scientific concepts were rejected before conclusively disconfirming evidence was in (for examples from chemistry, see Chang, 2012; for seminal examples from physics, see Duhem, 1954/1991). Nor did Darwin's books cause a cataclysmic Kuhnian-style scientific revolution. Other concepts are retained despite a mountain of disconfirming evidence (e.g., MacLean's triune brain concept). Put simply: Sometimes evidence plays a much smaller role in the decision than it should. Scientists are instrumentally rational, meaning they make decisions in service of achieving certain goals, and there are always competing goals.

Theory selection in the scientific world, like natural selection in the biological world, does not aim itself at one true solution along some gradient of steady improvement. The biological world contains many selection pressures. This is also true in science. The success of a particular scientific theory always entails a particular goal, and scientists have many goals that serve as selection pressures to influence whether or not they favor a particular theory with a particular set of concepts and categories. Sometimes those goals can be epistemic. A theory might be favored because it is simple and elegant. Often, however, those goals are non-epistemic. A theory might be favored because it is normative and concerns how things ought to be, consistent with a worldview or other values (such as a commitment to evolution), because it gives a person an identity and sense of belonging, or because it is better for a scientist's own career. Any of these non-epistemic goals can guide the theory selection and, correspondingly, the system of concepts and categories chosen. Research on the automaticity of goal pursuit suggests that aligning theory choice with other goals is not irrational per se, but motivated rationality in the service of something other than the discovery of reliable and trustworthy knowledge (Melnikoff & Bailey, 2018; Melnikoff & Strohminger, 2020). It is a commitment to certain answers rather than to the path to the answers.

Section 12. Non-Epistemic Goals In Science

Rational scientific practice has been described as a selection process akin to biological evolution — a sort of scientific survival of the fittest (van Fraassen, 1980). Many scientific theories compete with one another, the argument goes, and only the ones that latch onto true regularities in nature survive refutation. A sort of scientific survival of the fittest. (Godfrey-Smith, 2021, p. 85). This was Popper's view: Good scientific practice involves a combination of conjecture (which varies across competing theories) and refutation (which is selection by testing through observation of, or interaction with, the environment). The philosopher Emil Lakatos had a similar idea, although he suggested that each time a theory successfully fends off a possible refutation, it increases its predictive power and progressively builds in scope and predictive power (Lakatos, 1970/1978). This is the means by which scientific knowledge is supposed to steadily build a picture of objective reality.

There are a variety of challenges the scientific evolution argument, however. Kuhn refuted this picture of science when he brought the history of science to bear in *The Structure of Scientific Revolutions* (Kuhn, 1962/1970). Kuhn's *Structure* laid bare the ways in which scientific knowledge is conditioned on these social agreements. Scientific truths shift over time in ways that are not cumulative and therefore forfeit the right to claim any sort of perspective-independent objectivity. With these ideas, Kuhn was not making normative claims, criticizing the way science is practiced (vs. how it should be practiced). He was merely describing how scientists *actually behave* using historical examples.

Another problem with Popper's scientific survival-of-the-fittest idea is that it gets biological evolution wrong. Selection in the natural world, like selection in the scientific world, does not aim itself at one true solution along some gradient of steady improvement. Natural selection is more variable than that. It produces convergent solutions to the same problem (e.g., there is more than one way for animals to see or fly).

Moreover, the natural world is full of competing selection pressures. Similarly, scientists have many competing goals that they select on, some of which have nothing to do with scientific accuracy. A scientist might choose one theory over another to maximize a goal of how things in the world *ought* to be. Or how things ought to be for them.

Kuhn's *Structure* gave impetus to a field of inquiry called *science studies*, which is the study of scientists and how they practice science (i.e., applying science to itself to explain how scientific knowledge arises; e.g., Bloor, 1991; Collins, 1992; Hacking, 1983; D. L. Hull, 1988; Latour & Woolgar, 1979/1986; Laudan, 1977, 1990; Longino, 1990, 2001; for a subtler view, see Kuhn, 1977; for a general review, see Godfrey-Smith, 2021; for a discussion of how facts are made, not found, see Shapin & Schaffer, 1985/2011). One important takeaway from this large body of work is that scientists are as self-interested as any other people working in a community and competing for limited resources. Scientific decisions involve non-epistemic values and commitments such as intellectual allegiances, values, financial interests, beliefs about gender, class, and race, the pursuit of power and influence, and other career considerations. These factors influence which scientific concepts are retained and which knowledge claims become knowledge (e.g., Collins, 1992; Longino, 1990, 2001; Shapin & Schaffer, 1985/2011). Matters of reputation, recognition and influence invite scientists to deviate from rational decision making in science, sometimes leading to the systematic belief in falsehoods or a defense of mediocrity. Another equally viable interpretation, however, is that any person can

hold multiple goals at once, and self-interest invites scientists to rationally pursue non-epistemic goals.

Scientists might also favor a theory if it is easier to understand. A simple theory has advantages over a complex theory in this regard. Scientists will also favor theories that are more consistent with the conceptual framework of their community or cultural context. A theory that assumes psychological categories are natural kinds will be favored over a theory that assumes relational kinds in a scientific community committed to traditional scientific realism. Similarly, the broader cultural conception of a person as independent, individual actor may constrain the plausibility of a relational view of psychological science (in the same way that general relativity and quantum mechanics might have seemed implausible or even incoherent in prior centuries).

All things being equal, research funding also favors simple theories (Stanford, 2023). As reviewers, scientists tend to be more conservative in their funding decisions, in part because there are too few resources to go around (particularly for behavioral science). This is particularly problematic for high risk, high reward projects, in which novel concepts require some protection so they can be fully developed (Kuhn, 1962/1970). When there is a pressure to produce usable results on a three-to-five-year timeline, simple, conventional theories win, making it harder for unintuitive concepts to properly develop.