This is a draft of a chapter that has been accepted for publication by Oxford University Press in the forthcoming book *The Pragmatist Challenge: Pragmatist Metaphysics for Philosophy of Science* edited by H. K. Andersen and Sandra D. Mitchell due for publication in 2023.

4: The Bearable Thinness of Being: A Pragmatist Metaphysics of Affordances

Sandra D. Mitchell
History and Philosophy of Science, University of Pittsburgh

1. Introduction

Taking a pragmatist stance toward the practices and products of science shapes our answers to central philosophical questions[53]. In this paper, I will explicate how scientists' conceptual and representational practices work in concert with their observational and experimental ones to stabilize acceptance of scientific realism.

A pragmatist approach identifies the roles that agency and judgment play in the human reasoning and experimental practices that support claims about nature. Uncontroversially, particular goals, lofty or mundane, direct scientists to explore certain problems rather than others, focusing both the mind and the eye on specific regions for scientific investigation. As a consequence, the scope of what is deemed real is shaped by our aims and questions. I will argue that within a perspective consisting of goals, actions and questions, what we say there *is* and what we say it *does*, is justified by the ongoing interactions among representative models, causal experience and experiment, and conceptual frameworks in reaching a fallible convergence to what is real (see also Hacking 1983 and Chang 2022). I will defend an interactionist, pragmatist account to replace fundamentalist representationalist approaches to what constitutes realism of scientific theories and models. Fundamentalism about realism has been attached to two strategies, a top-down view that what is real is read off the *structures* described by the best scientific models and a bottom-up view that what is real are the *entities* presumed by experimental practice (Chakravartty 2021). I offer a non-dichotomous alternative. What we are justified in claiming as

---

[53] See Andersen and Mitchell (this volume). In this paper I treat "pragmatic" and "pragmatist" as distinct. While "pragmatic" references means-ends reasoning, or judgments in light of a goal, "pragmatist" reflects impacts on actions or practices resulting from a wide range of influences on scientific inquiry including differences in goals, features of experiential design and implementations, views on the nature of causal relationships in different contexts, etc.

real phenomena are the affordances[54] constructed from the integration of top-down and bottom-up strategies.[55] I will illustrate how this pragmatist realism works, by appeal to a case of a complex, robust phenomenon.

2. Representational Theories and Experimental Data

Much progress has been made since the early logical empiricists' simple logical rendering of the relationship between scientific theories and experimental observations in terms of the conditional relationships between H-statements and O-statements. A series of elaborations of what goes into the inferential and causal interactions in experience and experiment exposed other assumptions and finer structure to describe the warrant for accepting the claims of scientific theories and models. Explicating the additional assumptions necessary for theory testing is part of the legacy of Duhem's underdetermination arguments found in *The Aim and Structure of Physical Theory*, originally published in 1914 (1991) that focused on how to connect the practical activities of an experiment with the typically mathematical descriptions comprising scientific, specifically physical, theory. Judgments, more than just logic and raw perception, are required to relate theoretical predictions to the upshot of experimental interventions. "Judgement", I argue, involves appeals to interdependent semantic, epistemological and metaphysical assumptions. The language describing particular experiments and measurement depend on practical and instrumental features of detection. This is different from the language predicting what should be

---

[54] 'Affordance' is a term introduced by J. J. Gibson (1979), an ecological psychologist, in developing his non-representationalist account of perception. For Gibson, an affordance invokes what things external to an agent provide to the agent for action. For example, relative to the abilities of a given animal, a given surface may be 'walk-on-able and run-over-able' (Gibson 1979, p. 127). Affordances are the product of both the agent and the environment in which the agent acts.

[55] While my appeal to Gibsonian affordances is in support of a version of actualist, integrated realism, see Vetter 2020 for an appeal to affordances to support a non-Humean metaphysics. See also Rydenfelt (2021) for a defense of a Peircean pragmatist hypothetical realism that embeds the judgments of what is representationally real within the goals of scientific practice. I share with his view an acceptance of the contingency of claims of realism . My view also follows Price 2007 in "dimming the light" to reduce the sharp line between realism and irrealism that is required by some forms of representationalism. What I hope to do is to show "how the light gets in" (thank you, Leonard Cohen).

observed according to a theoretical model. That semantic gap must be closed for the results of an experiment to serve as evidence for or against a theory (Darling 2002).

Additionally, what can be known from experimentation rests on two legs, one on the causal process of intervening on a presumed natural phenomena and the other on the measurement or description of the results of that process. It is not enough to just "twist the lion's tail"[56] to empirically justify a claim about nature. Decisions about what is being twisted, what features are causally engaging with the detecting devise also are required. Most experiments operate on phenomenon-adjacent materials. What is in nature is not what directly interacts in the laboratory. For example, in protein structure prediction, the target protein is no longer culled directly from natural materials, as when cow's blood was used in early studies of the protein hemoglobin, but rather it is made through gene cloning, purification and isotope-enrichment. The object in laboratory experiments is a form of a protein tuned to be detectible by a specific type of causal interaction.[57] The "same" protein destined for X-ray crystallography detection is prepared very differently than the one for nuclear magnetic resonance spectroscopy experiments. Judgments must be made about how the version of the protein in the causal interactions of an experiment is related to the protein "in the wild"[58]. In addition, what counts as causation itself, the criteria for claiming a causal relationship between the phenomenon and instrument yielding data, sits behind the kinds of decisions made about which parts of the initial data output of an experiment are kept and used and which are thrown out and ignored.

The well-known logical problem of underdetermination of theory by data is usually attributed to Duhem, who exposed the role of auxiliary assumptions involved in the designs, practices, and measurements in experiments, all additional to the theory under test, that are needed to derive experimental predictions. Such assumptions are required in the inference from abstract theory to concrete data prediction to specify the experimental set up, the conditions for blocking

---

[56] This is attributed to Francis Bacon (perhaps apocryphally) by Thomas Kuhn (1976), to distinguish interventionist experimentation from passive observation.

[57] Acton, T. B., Xiao, R., et al (2011).

[58] Related judgments are made in the use of model organisms, or representative samples in statistical analysis.

confounders, etc. Given the logical role of auxiliary assumptions, when an experimental result does not match the theoretically predicted result, then, Duhem argues, there is no way to isolate the theory under test as responsible for that failure. This is the standard logical account of underdetermination of theory by evidence in the context of theory testing. But Duhem also argued for a form of semantic underdetermination (Duhem 1991, Darling 2002). He identified two components to experimentation. The first he called practical facts, namely what is observed where no physical theory, or at least not the abstract theory to be tested, is required. Practical facts represent the observed results of the causal interaction of the target phenomenon and the detecting device. The second component, Duhem called theoretical facts, namely the predicted observations derived from the theory.

Duhem posed the problem of semantic underdetermination as follows: "The same theoretical fact may correspond to an infinity of practical facts….The same practical fact may correspond to an infinity of mathematically incompatible theoretical facts." (1991 p 152). There is often a mismatch of the precision in a theoretical prediction, especially if it is of a numerical value, and the imprecision accompanying the measurements of the outcome of an experiment. The range of values resulting from repeated experiments, sometimes reported as the average, sometimes "cleaned" by various methods, may be less discriminating than what is deducible from a theoretical model. For example, a theory might predict a precise value, say the melting point of lead as 621.43 °F. The experimental protocol may permit a degree of precision only to the whole number. What is the inferential import of the practical measurement, for example, of 621 °F to the predicted value of 621.43 °F? The practical measurement does not discriminate between 621.00°F, 621.40°F, 621.99 °F and so on. Mathematically these are not equivalent, experimentally they are indistinguishable, but the theoretical interpretation has to make a judgment of how to treat them, that is, decide whether they support or refute the prediction of 621.43 °F.[59] Duhem argues that requires translation.

"But translation is treacherous: *traduttore, traditore* (to translate

---

[59] See Tal (2017) for an extended discussion of measurement standards and accuracy. On his account, epistemic measurement accuracy is influenced by more than precision. He argues, as I also will, for a pragmatist understanding of accuracy and robustness in which both theoretical and experimental factors are jointly incorporated.

> is to betray). There is never a complete equivalence between two texts when one is a translated version of the other. Between the concrete facts, as the physicist observes them, and the numerical symbols by which these facts are represented in the calculations of the theorist, there is an extremely great difference." (1991, p 133)

The semantic source of underdetermination (Darling 2002) exposes the role of judgments which rely on reasons not contained in either the abstract theoretical model, nor in the concrete causal experiment. These include judgments about what counts as an outlier data point, what counts as noise in the data, what counts as data skewed by systemic "error", or more generally, when variation in the data is taken as the signal and when it is taken as deviation from the signal. Duhem identifies the problem, but what is the solution?[60] What goes into the decisions that permit the empirical results of experimentation to arbitrate between alternative theoretical models? As I will describe below, those judgments appeal to the theory under test, as well as theories of the instruments being used, assumptions about what the bounds are of a phenomenon, and to philosophical frameworks about what counts as, or serves as criteria for, causation generally.

Another difference between abstract theory and concrete data is recognized by Bogen and Woodward's (1988) influential distinction between data and phenomena. According to them, scientific theories predict and explain facts about stable phenomena, like the melting point of lead, or the 3-dimensional structures of a particular protein, or the chemotaxis behavior of e. coli. On their view, data acquired observationally or experimentally can serve as evidence for the existence of such phenomena, but data is not predicted or systematically explained by theories. This is because data are "idiosyncratic to particular experimental contexts, and typically cannot occur outside of those contexts." (p 305-306). Data provide a record of what is locally observed, whereas the phenomenon, which is the causal target of a particular experiment generating specific data, goes beyond any single experiment. Phenomena have stable repeatable

---

[60] See Darling, K. M. (2003) for different type of realist account of Duhem's own solution. See Rydenfelt (2021) for a pragmatist defense of realism

characteristics that may be detectable by means of a variety of different experiments yielding different data (see also Teller 2010 for a model-theoretic interpretation of their view.)

I interpret their distinction in the following terms: natural phenomena are what we posit as the source of the signals entering into causal interactions with detecting devices in experimental, observational contexts (see also Woodward 2000, 2010). The results of experimental causal interactions are represented as data. If the experimental detection devices and procedures are *reliable*, i.e. the produced effects represented as data reflect the properties of the source phenomenon, then, from measurements of the produced signals, we can construct a model of the data that can be used as evidence for the existence of the (stable, non-idiosyncratic) phenomenon. Theories explain and predict features of the phenomenon, experimental data reflect only what is causally detected in an experiment. This is a version of Duhemian semantic underdetermination, cashed out in terms of inferential role.

Bogen and Woodward (1988) introduced phenomena as a triangulation point between data (which stand in a causal relationship to phenomena) and theories (which inferentially predict and explain phenomena). However, the inferential separation of theories from data (theories don't explain data, they explain phenomena) cannot be strict. Some have argued that theory-ladenness of observation defeats a claim that theory and data are independent. (Schindler 2011, see also Boyd and Bogen, 2021). Others, like Suppes and Giere, provide more fine grained distinctions among theories that are involved in prediction and experiments that support a more nuanced architecture for the inferential roles of theory and data and thus new spaces to consider the judgments required for translating the results of a particular experiment so that it can speak to a particular abstract theoretical model.

Suppes (1962) articulated a hierarchy of theoretical models[61] involved in assessing the adequacy of causal processes in generating reliable experimental data. On his picture an analysis of a

---

[61] Both Suppes and Giere promoted the semantic view of theories, and the content of scientific theories was interpreted in terms of theoretical models consisting of structures or sets of relations, typically mathematical.

relationship between theory and experiment is iterated at every level of the hierarchy, from non-formalized considerations of confounding factors in the experimental set up, to experimental design choices, statistical models of the data, models of the experiment, and experimental models of the phenomena. Different criteria are applied to determine if an experiment is run correctly, than those invoked in deciding whether the experimental data count as a realization of the theoretical model. In short, Suppes claims that "a whole hierarchy of models stands between the model of the basic theory and the complete experimental experience. Moreover, for each level of the hierarchy there is a theory in its own right." (260).

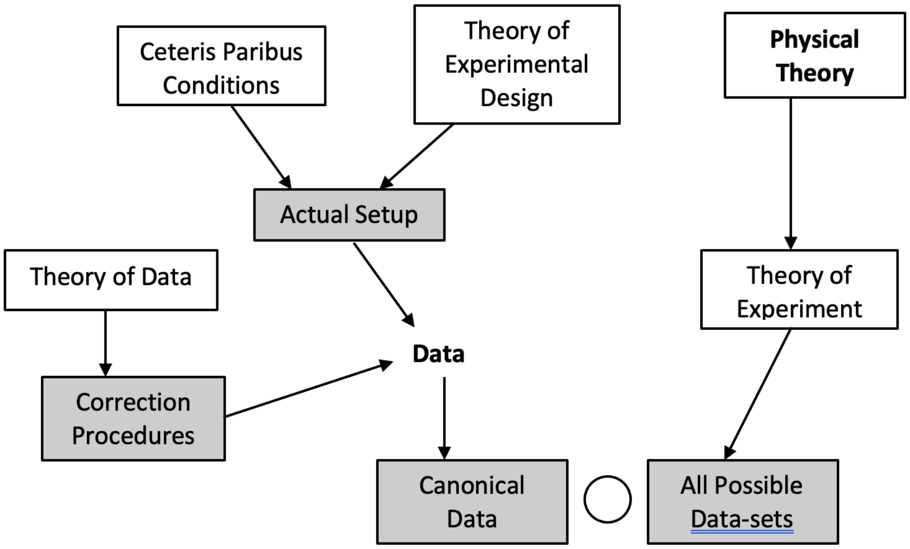

Figure 1: Recreation of Darling's representation of Suppes account (2002, 524).

What are the relationships among the multiple theories involved in testing the abstract physical theory that represents stable, non-idiosyncratic phenomena? Suppes separates the theory of the phenomenon from the theory of the data, in two distinct hierarchical tracks (Figure 1). Giere (2010) proposes a hierarchy of levels (Figure 2) to locate the two tracks by opposite directions of the (not necessarily strictly inferential) arrows, up from data and down from theory. Yet even with these elaborations, the semantic gap persists. The point of contact between theory and data occurs in the confrontation of measured data from experiment with predicted data from theory. Bogen and Woodward's explanatory difference between phenomena and data, Suppes's and Giere's hierarchies of theoretical and experimental steps point to important differences.

However, the character of the relationship between the distinct components is still in question. We might ask in what ways the theory and the data independent of each other, in what ways they are not (see Franklin and Perovich 2021, Epstein and Forber 2013).

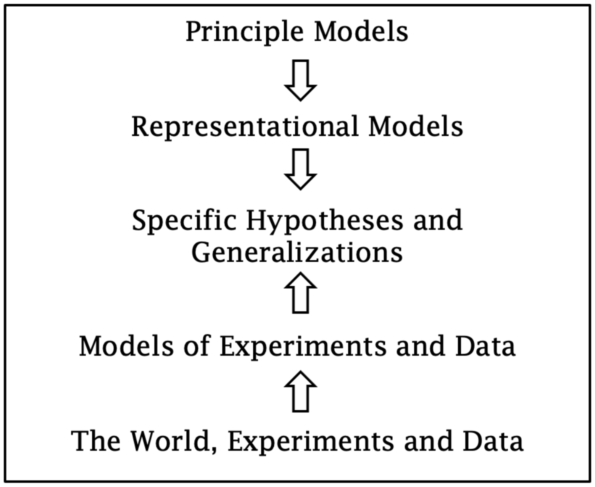

Figure 2. Recreated from Giere (2010, p 270)

I have argued elsewhere (Mitchell and Gronenborn 2017, Mitchell 2020b) that neither strict independence nor a strict hierarchy is adequate to account for the inferential relations among theories and data in scientific practice. Consider the prevalence of semi-empirical modeling where there is extensive feedback between data and theories in generating the point of contact between the two. In those cases, predictions are derived not solely from an interpretation of abstract, formal functions as in, e.g. testing the mathematical structures that provide the content of Newton's laws in different representational models of that structure in falling bodies, springs or pendula experiments (Giere 1999). In semi-empirical modeling, the formal functions that predict what should be observed in the data are derived from both the theoretical model and from relationships supplied directly by patterns in data. In these cases, the theory-data connection is more complicated. For example, consider the identification of protein structure, i.e. the three dimensional location of the atoms of a protein in is most stable conformation, which is important in understanding protein function. In homologous semi-empirical modeling for predicting protein structure from its amino acid sequence, data is used in a constructive manner. Predictive algorithms embed the results of prior experiments on sequence-similar targets to limit the space of protein conformations that need to be searched to locate the conformation with the least mean energy, and hence predicted to be the "real", functional atomic structure (Mitchell and Groneborn 2017). "The accuracy of a comparative model is related to the percentage (of) sequence identity on which it is based, correlating with the relationship between the structural and sequence similarity of two proteins." (Baker and Sali 2001, 93) There are a variety of automated search algorithms to determine sequence similarity and fold similarity among the protein structures obtained through experimentation that has been deposited in scientific data bases. The predictive algorithms that start with a comparative homology analysis have, to date, been much more accurate, tested against experimental results for the target proteins, than *ab initio* predictions based only on the sequence of the target protein and molecular dynamics of the physical interactions of the atoms in that protein (for a recent overview of the different approaches, see Deng, Jia and Zhang 2018; see also Humphreys 1991, Ramsey 1997). A static, hierarchical view of the relationships among theories, predictive models, and experiments obscures some of the processes and judgments required to construct both the empirical and theoretical bases for inferences about protein confirmation. The same is true of inferences in the many and diverse cases in which scientists use semi-empirical models to understand phenomena.

In crafting scientific knowledge, formal theories of the phenomenon and data from experiments are inextricably entwined. The theory at test, by its characterization of the phenomenon, is required both in designing an experiment and determining how to describe the experimental results that would count as a test of its accuracy. Theories of the experiment and the data are required to set standards of causal reliability so that the appropriate data are considered. The entanglements of theory and experimental observations generate new questions for how realism about phenomena can be warranted. So far, the discussion has been about theories and experiments, both products of human conceptual and causal practices. What else goes into the stronger commitment to the reality of the phenomena that are represented in the confirmed theory and the source of the experimental data?

3. Metaphysical implications of the complexities of theory-phenomena-data relationships in practice

In this section I will consider the sources of warrant for realism about phenomena. There are top-down and bottom-up alternatives to both the content of and the warrant for scientific realism. Anjan Chakravartty (2021) describes what he calls the “realist tightrope”, on which one aims to balance between the thinnest and thickest descriptions of what is real and associates these with bottom up and top down methods of warrant. Minimally, a realist stance seeks to anchor the reference to a phenomenon in the world in a way that is independent of individual detections or conceptualizations of it. Maximally, the realist commits to the phenomenon as having precisely the detailed features described in the theoretical model.

These degrees of “thinness” are related to two fundamentalist ways of warranting metaphysical claims about phenomena from scientific practice: a bottom-up approach (entity realism) and a top-down approach (structural realism).[62] The bottom-up approach takes causation as the foundation of positing what's real. Unobservable phenomena are taken to be the cause of experimental data, and thus the interactions and results of empirical practices provide the required metaphysical warrant. Defenders of this view include Ian Hacking (1983) and Nancy Cartwright (2007). Hacking, in discussing an experiment he observed at a Stanford laboratory where electrons and positrons were sprayed, one after the other, onto a superconducting metal sphere, famously claimed that "if you can spray them, then they are real" (1963, p 24). Cartwright suggests that what are real are causal capacities in the world, identifying capacities as the source of stable causal laws that we can infer from causal interactions. When a particular experiment yields specific measurements, for Cartwright, the explanation for these measurements posits entities in the world that have the capacity to generate those signals, similar to Bogen and Woodward’s stable, non-idiosyncratic phenomena. Entity realism is inferred bottom-up from causal manipulations. Data are evidence for the existence of phenomena.

A fundamentalist top-down view recognizes that unobservable phenomena are the referents of abstract explanatory theories. Structural realists read what is real off of the formal relations represented in scientific models of such theories. Using this strategy, the best confirmed theories are the source of warranting claims about what is real. On this top-down approach what science

---

[62] See Galison (1988) for a historical analysis of the two sources of fundamentalism.

discovers about nature are not the entities which have causal capacities but rather the structural relations that explain patterns in our observations. The structures described by the mathematical relations in our best theories are taken to be isomorphic or otherwise similar to what is real. John Worrall (1989) defends this type of structural realism when he claims that "On the structural realist view what Newton really discovered are the relationships between phenomena expressed in the mathematical equations of his theory". Current philosophical debates between entity realists like Cartwright (1982) and structural realists like French and Ladyman (2010) exemplify Charkravartty's bottom-up versus top-down fundamentalist dichotomy.

I suggest we break out from the dichotomous choices on offer. I maintain that we need a non-fundamentalist, pragmatist option that jointly uses both strategies for warranting the real. This eschews a static representationalism of the real either as what is presumed to be "in the world" by experimental practice or as referents of theoretical models. What we are warranted in claiming about what is real are not just structures and not just entities, rather it involves the integration of both human interventions and conceptualizations. Neither of the two fundamental strategies alone will capture the judgments required for warranting claims of realism[63]. I suggest that real phenomena are those things in the world that are sufficiently stable to afford the coordination of the results of both causal detection *and* structural representation.

4: Affordances, an interactionist metaphysics

My view is inspired by the work of the 20th century ecological psychologist J. J. Gibson who coined the term "affordance" to explain human and animal behavior. (Gibson 1979). Gibson suggested we need a term to identify properties or entities that would convey the joint contribution of both the actor and the environment in which they act. Gibson recognizes the similarity of his proposal to the idea of an ecological niche[64] which identifies the features of the

---

[63] Chakravartty's conclusion is to resist the tightrope and embrace a plurality of consistent claims from top down and bottom up reasoning regarding what is real. Here I suggest that by identifying the locations of incompatibilities (in data interpretation, theoretical constraints, or philosophical commitments) we can better understand how compatibility is forged in scientific practice.

[64] How to define an ecological niche, and its significance to ecology is still much disputed. See Justus, J. (2019).

external world that are salient to particular species' dependence on their capacities to interact with it. Gibson coined the term "affordance" to accommodate that interactive relationship in the psychology of perception. For Gibson, affordances are invariant features of the external environment, that are perceived, classified, etc. dependent on an animal's capacities to interact with it.

> The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill. The verb to *afford* is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. (1979, 119)

Affordances for Gibson are real, objective properties of the environment-plus-organism that make specific behaviors possible. For example, Gibson (1979) indicated that "to be graspable, an object must have opposite surfaces separated by a distance less than the span of the hand." (133). The affordance, however, is not "out there" for the organism to engage, like a disposition or causal capacity. Rather it is constructed by the engagement with the organism.

> An important fact about the affordances of the environment is that they are in a sense objective, real, and physical, unlike values and meanings, which are often supposed to be subjective, phenomenal and mental. But, actually, an affordance is neither an objective property nor a subjective property; or it is both if you like. An affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. (Gibson 1979, 129)

I propose that we consider what is justifiably real in science to have this type of interaction relation. The joint contributions of causally grounded experimental data and theoretically structured representational models together specify what is real and what is not. If we substitute "scientist" for "animal" and "what is real" for "environment" we get a framework that better locates the metaphysical claims warranted by scientific practice. Borrowing some of Gibson's language:

The affordances of real phenomena are what they *offer* the scientist, what they *provide* or *furnish to experiment and representation*. I (Mitchell) mean by it something that refers to both the causal properties in nature and the representational framework of the scientist. It implies the complementarity of both in establishing what is real.

This interactionist approach to realism rejects both top-down and bottom-up fundamentalism. There is no fixed ontology of entities or fixed ontology of structures. Instead, what we project back as being constitutive of nature is contingent on a complementary relationship between what we can detect and how we represent it.[65] Nature is independent of us. The universe and our planet existed prior to humans and probably will exist post the human era. But what is viable as a metaphysics of what is **real** according to our science is what we, as limited beings, are justified in claiming about nature. All we can justifiably say yields a contingent, pragmatist ontology of stable, detectable and representable, entities and relations, i.e. affordances built from both experiment and theory.

Real affordances can be represented as entities with stable dispositional causal properties that we engage in experimental practices. Real affordances can also be represented as stable structures described in the relations constituting our best theories. Some form of stability is a minimal metaphysical presumption for doing science. In what follows I will investigate the kinds and character of stability required by experiment and model. Specifically, I will address a key question for a pragmatist form of realism: What are the roles of causes and concepts in constructing the real?

5: Reliability of data, theory and frameworks

I have suggested that affordance metaphysics exposes the interactive roles of theories, conceptual frameworks, and causal-experimental models in warranting metaphysical inferences. An assumption of any experimental detection procedure is that there is a phenomenon[66], independent of the detecting device, that is being causally engaged. Real phenomena afford

---

[65] See Chirimuuta (2017) for an analogous argument about color vision.

[66] By phenomena I follow Bogen and Woodward (1988) to mean the stable relations or entities that are not directly observable candidates for what is real.

stable, repeatable causal interactions. They might be detected by means of a variety of different procedures, each producing idiosyncratic data as emphasized in Bogen and Woodward's distinction. The agreement between a theoretical model-based prediction from phenomena to data and the experimental inference from data to phenomena are the joint source of justification for the existence of the phenomenon and relations specified in the model. If alternative models predict divergent data, and if the experiments are sufficiently precise to distinguish between the competing models, then the data provides the kind of contrastive confirmation that justifies accepting one model over another as a more accurate representation of what is real in nature. If the data of multiple, diverse experiments all agree with the model-based prediction, then that is taken as the strongest evidence that the multiply detected phenomenon is real (Salmon 1985). In what follows, I will dig deeper into the structure of assumptions that ground judgments of realism from the convergence of experimental results.[67]

When phenomenal stability is conjoined with reliable detection, then convergent data are taken to support claims of realism. Causal reliability is a judgment about the causal process generating data in an experiment which contributes to the inference from data to phenomenon for each type of experiment.[68] When multiple different types of experiment generate data from which "the same" phenomenon is inferred, then realism is the conclusion of a no-miracles argument. When there are substantially different, if not strictly independent, ways of causally detecting the source of the signals in the experiments, and they all yield the same measurement, then, as the no miracles argument goes, it must be because the source is a real phenomenon independent of each individual experiment with its own idiosyncratic assumptions or models of the experiment. Experimental convergence is taken as the strongest "bottom-up" evidence for realism.

What is required for this argument to support a claim of realism? First, the causal processes (phenomenon to data) realized in the experiments should, themselves, be reliable. Second the

---

[67] In Mitchell (2020a) and (2020b), I have presented arguments for how divergent data from experiments can also be a route to increased representational accuracy.

[68] See also Woodward (2010) for a description of reliability in inferences from statistical evidence.

multiple experiments should be independent of each other. I will discuss first the issue of independence, then turn to reliability.

5.1 Independence of experiments

There are three types of independence that are invoked in arguments for convergence or robustness of experimental results to ground a claim of realism. First, the predictions from the theory should be independent of the actual results obtained by running the experiment. Second, the experimental design and inference to the measured data in a particular experiment should be independent of the theory/model under test. Third, the multiple experiments whose results converge, should be independent of each other.

Clearly there are theories involved in designing and performing an experiment. The models of the experiments presuppose theories in order to set up the conditions in which the phenomenon would be detectable by the instruments, as Duhem emphasized. A central concern has been which conceptual and theoretical assumptions are required for the data from a single or multiple experiments to be a test of a prediction about the phenomenon. Suspicion is raised about whether the data from an experiment can serve as a test of a theory if, for example, the predictions are built after the fact from the observed data. Overfitting parameters or tweaking a model after the experiment has been run can be epistemically perilous (see Epstein and Forber, 2012). But it is also the case that data about how the experimental apparatus operates, which may be gleaned from results of running experiments, is required in order to generate a prediction from the theory in the first place. Which initial conditions need to be in place, which causes can be blocked from confounding the results? For an experiment to be "about" the phenomena described in a theory and the prediction to be "about" the experimental results, some coordination is required. To produce epistemically relevant information, the Duhemian semantic underdetermination relationship between theoretical predicted facts and practical experimental facts requires a connection between them be made. To do so undermines a view of the strict independence of data from theory for it to serve as evidence for that theory.

Second, instruments selectively interact with some features of the target phenomena, not all features, and in this sense, are causally biased (see Giere 2006). Judgments of convergence and its associated warrant for realism require a prior decision that the multiple experiments are of the same phenomenon. Measurements or some other representation of the causal effects of the instrument/phenomenon interaction encode conceptual and theoretical assumptions about the salience of the feature targeted in an experiment for the theoretical fact about the phenomenon. For example, experimental determination of protein structure is predominantly achieved by X-ray crystallography or nuclear magnetic resonance spectroscopy. X-ray experiments target the electron clouds of crystalline atoms in a protein which diffract a beam of incident X-rays. The angles and intensities of the diffracted beans are measured and a three-dimensional density map of the electrons produced. From this, the relative positions of the atoms in the crystallized protein can be determined. In contrast, NMR protein experiments target quantum mechanical properties of the nuclei of the atoms (nuclear spin polarization) in the protein. A protein is placed inside a large magnet, radio frequency signals are sent through the sample, and the absorption of those signals is measured which allows the relative distances between atoms in the protein sample to be inferred. When the two different experimental protocols agree in their inferred experimental model of the protein structure, this is taken to be evidence that the real structure has been detected.[69] The behavior of electrons and the behavior of nuclei are related by theory, which permits us to say that the two experimental methods are "about" the same target phenomenon, namely the atomic structure of the same protein, even though the specific causal interactions in the experiment involve very different affordances – the interaction of atomic *electrons* with monochromatic X-rays and the interaction of atomic *nuclei* with certain kind of variation in a strong magnetic field. The affordances are constructed jointly from the theoretically "same" external target (atoms in a protein) *and* the different experimental environments and detecting devices. The two different affordances produce two different, theoretically linked, and reliable views of realities of protein structure.

---

[69] The story is much more complicated than this, of course. Each experimental protocol generates an ensemble of conformational structures that are consistent with their data, not a single structure. So agreement has to do with the overlap of the ensembles.

Tal argues that: “To attain the status of an outcome, a set of values must be abstracted away from its concrete method of production and pertain to some quantity objectively, namely be attributable to the measured object rather than the idiosyncrasies of the measuring instrument, environment, and human operators.” (2017, 35). But what is it for the quantity to be objective? On my view, it is for there to be *integrated affordances*. I have argued elsewhere (Mitchell 2020a, 2020b) that multisensory perception provides an informative analogy. When more than one sense modality (vision and hearing) uptakes information from the environment, a signature consequence of their neuronal integration is the superadditivity of their effects. That is, the number of neurons firing in multisensory integration is more than the sum of the firing rate of the two individual modalities. Superadditivity also is seen in multi-modal integration in time-to-task completion studies (e.g. grasping an object). Like Gibsonian affordances, multisensory integration processes are context and modality dependent. Vision is more reliable for spatial representation while audition is more reliable for temporal representation. In specific contexts, visual input will dominate over auditory input, in other contexts, the reverse ordering will occur. Yet, integrating the two provides superadditive accuracy of representation (Van Atteveldt, et. al. 2014). Just as brains integrate input from the modalities by which they acquire information, selectively targeting different types of signals in ways that permit more accurate representations and more effective action, so too, I argue does the integration of the multiple causal, theoretical and conceptual “modalities” of scientific practice.

Theories are essential for both performing an experiment, in the form of a causal theory of the experiment, and for identifying when two convergent (or divergent) experiments are about the same phenomenon. Strict independence of theory and experiment is not possible. Does that mean we have to accept an all embracing Quinean holism where every scientific claim is connected to every other scientific claim? In his seminal work of the history of the thermometer, Hasok Chang (2004) argues that scientists can avoid the most extreme forms of holism by adopting a “principle of minimalist overdetermination.” Chang means by overdetermination the type of convergence discussed above, namely, agreement by different methods on the measurement value ascribed to some phenomenon, e.g. a temperature determined by both theoretical inference calculation and measurement by a mercury thermometer or measured by two different types of thermometer. He agrees that invoking auxiliary hypotheses is necessary in order to make predictions, build apparatus and interpret the results of an experiment, but by minimizing those assumptions, Chang argues, the damage can be contained. “The heart of minimalism is the removal of all possible extraneous (or auxiliary) non-observational hypotheses.” (Chang 2004, 94). His approach suggests we should prefer the most direct, or least theory-mediated inference, between the causal experimental interaction and the representational measurements.

One might interpret “minimalist” to require that the theories of the experiment should rely on *as few assumptions* about the function that associates the target feature with the measurement as possible. By minimizing the number of auxiliary assumptions Chang’s thesis would suggest we can provide the strongest grounds for taking experimental results to be confirmations or refutations of the hypothesis being tested. The theory of the causal interaction between phenomena and instrument should, if this interpretation of Chang is correct, rely on no more assumptions implicated by the theory under test than is necessary. This approach has the virtue of recognizing the ineliminabilty of influence of theoretical assumptions in acquiring measurements in an experimental set up while attempting to avoid *ad hoc*, unreliable confirmation. However, this interpretation implies that using the fewest number of theoretical assumptions in producing data [minimization] is what matters. Given Chang’s extensive appeal to scientific practice and his recent pragmatist account of realism (2022), it is clear that this interpretation is not right. I suggest, and I suspect Chang would agree, that what matters most is *what* is assumed in an experimental set up, not just *how many* assumptions are used. By specifying what parts of the theory under test are assumed, the decision that convergence of multiple experiments has been achieved and realism can be justified can be made openly and reflectively.

5.2 Experimental reliability in data-to-phenomena inference.

I have suggested that affordances are what science can justify as being *real*. Neither causally detectable phenomena alone nor theoretical structures alone, are sufficient to characterize the causal interactions with multiple detection devices that produce data which support claims of realism. Such claims always involve a translation from Duhemian practical facts, or minimal descriptions, of the upshot of the causal process enacted in an experiment into Duhemian theoretical facts which support or fail to support theories describing the relations that hold among real phenomena. There are assumptions and decisions throughout this chain of inferences that shape the claims of what is real and depend on both our goals and our epistemic and technical limitations. I now will turn to the assumptions that enter into this inferential process through judgments made about the reliability of data.

For experiments to provide grounds for positing something about unobservable phenomena, we need to assume that the data is reliably reflecting the causal input from the target phenomena. There are important issues about how we reason about eliminating random and systematic influences in data production to filter out the non-target contributions to the initial data. One role for multiple experiments is to expose the systematic, non-target contributions of the individual experimental protocols and operations (Kuorikoski and Marchionni 2016, Mitchell 2020b). Random and systematic sources of non-target signal require different treatment in reaching the decision that the "cleaned" data provides information about the target phenomenon. (Suppes 1962) identified descriptive features of the data that could provide clues about the target and non-target sources. Random noise is identified by characteristic instability in the data over multiple experimental runs (Steinle 2002), and its effects can often be averaged out. Experimentalists compare iterations of experiments to distinguish random and systematic sources of data (e.g. Morris 1992). Systemic, non-target effects are not simply averaged, but may be managed in a variety of ways including redesigning experimental protocols to block non-target causes, to re-interpreting measurements to adjust for known biases in the instruments.[70] Experimental repetition and comparison can expose features of data that can be associated with random and systemic, non-target, contributions to the measured signals. But to serve this

---

[70] How wonderful that some of the programs for cleaning data in structural biology Xray crystallography experiments are named POINTLESS and AIMLESS. See Evans and Murshudov 2013 for a serious discussion of the criteria for deciding the `resolution' of a measured data set.

function, both theoretical assumptions about what *kind* of phenomenon is being tested and conceptual assumptions about causality itself are involved.[71]

Here I will examine the roles played by the theoretical characterization of the phenomena and the philosophical theory of causation in judgments about the reliability of experimentally produced data. Both are essential to the justification of claims of realism. If there is a real phenomenon, there will be some form of stability across experimental detection. And if the detection process and measured data are causally and inferentially reliable, then convergent data can be taken as justification of realism. But, I will argue, the judgment of the reliability of detection depends upon both the kind of phenomena investigated and the type of causation invoked.

Woodward's (2005) influential account of causation characterizes causal stability as invariance under intervention. On his view, invariance need not be exceptionless (see also Mitchell 1997, 2000). To count as causal, "A generalization can be invariant within a certain domain even though it has exceptions outside that domain. Moreover, unlike strict lawfulness, invariance comes in gradations or degrees." Woodward 2000, p 199). Woodward's interventionist account of causation is not intended to be merely methodological, but is a philosophical, or conceptual account of causation: "… for Y to change under an appropriate intervention on X just is what it is for X to cause Y." (2000, fn1 p 204-5).[72] One alternative account of causation that Woodward rejects, is a regularity account that requires strict laws. On the strict law view, for X to cause Y is for the generalization describing the causal relationship to be universal, exceptionless, true and naturally necessary. There are other philosophical theories of causation including probabilistic, primitivist, eliminativist and process accounts (Gallow 2022, Frisch 2022, Hitchcock 2021). For the purposes of this paper, I will consider only interventionism and strict lawful regularity accounts.

---

[71] The theoretical differences between Quetelet and Galton, for example, treated variation in the values of the measured data of a population in completely opposite ways. Quetelet, by appealing to the mean-squared law of astronomers, assumed there was a "real" singular height of Scottish soldiers and the measured variation could be averaged away to find it. Galton, in contrast, assumed the variation was the real phenomenon, and the average was a statistical description of the data – not the world. See Stigler (1986).

[72] Woodward's most recent work has explicated the psychological foundations for his counterfactual account of what humans mean by causation (Woodward 2021).

Causal reliability in experimentation is a judgment about how well the causal signal from the phenomenon is reflected in the data of the detecting device, and hence how well it supports the reliability of the inferences from data to phenomena and subsequently in support of theoretical models. Sometimes reliability is parsed in terms of validity. Campbell (1957) introduced the distinction between internal and external validity and further elaborations have been developed, but basically experimental validity refers to the accuracy of an experimental measurement. Validity has been identified with replicability of results and applicability to targeted real-world situations. For example, Guala characterizes internal and external validity as follows "the result of an experiment E is internally valid if the experimenter attributes the production of an effect B to a factor (or set of factors) A, and A really is the (or a) cause of B in E. Furthermore, it is externally valid if A causes B not only in E, but also in a set of other circumstances of interest, F, G, H, etc. (Guala 2003, p. 1198). But the "really is" language begs the question I want to explore, namely, how can we justify claims of what things "really are" on the basis of the reliability of experiments? We incur a problematic circularity if reliability requires we know in advance that reality for which the experimental practice is designed to provide evidence.

I argue that the affordances constructed from philosophical accounts of causation, theories of the phenomenon, and judgments of experimental reliability support a non-foundationalism in claims about realism, hence making room for the interactionist affordance alternative I propose.[73]

On Woodward's account of causation, causal relations and the generalizations that describe them can differ in degrees of invariance and still count as causes. Whether data counts as reliable

---

[73] Reiss 2019 argues against both the evidential reasoning built on the distinction between internal and external validity and methodological foundationalism. He claims the latter assumes some forms of experimentation are intrinsically reliable, as seems to be the case in medicine's overwhelming endorsement of the "gold standard" of randomized controlled studies as exclusively justificatory. Reiss suggests this form of methodological foundationalism can be seen as an attempt to realize the ideal interventionist logic of Woodward's invariance account of causation. Reiss suggests that "Reasoning concerning target systems should begin with a hypothesis about that system and ask what types of evidence we need to establish that hypothesis." Reiss, J. (2019). I share Reiss's conclusion, but arrive at it by a different route and deploy it for different purposes.

evidence of a real phenomenon or not depends on the kind and degree of stability or invariance that is attributed to the causal structure by a theory about the phenomenon. Contrary to the claim that the theory of the phenomenon is or should be independent of the experimental data, for data to be deemed reliable, that theory has to be invoked.

The roles of the theory of the phenomena and causal theory in determining reliability is somewhat invisible in the easy cases of the most stable phenomena, like the melting point of lead, discussed in Bogen and Woodward (1988). In what follows I will illustrate the roles in a case of emergent biological phenomena.

What is the relationship between the reliability of an experiment and the degree of invariance of the causal relation? The melting point of lead has the properties that fit the Bogen-Woodward conception of phenomena, i.e. stability over a wide range of experimentally enacted causal conditions. On Woodward's interventionist account (2005) of what causality is, that means the functional relation among the variables explicitly represented in the theoretical model predicting the melting point is *invariant under intervention*. For Woodward's account, the functional relation is also independent of other causal factors operating, i.e. it satisfies the condition of *modularity* or independent disruptability. And the functional relation is *insensitive* to a host of conditions of varying, non-represented or exogenous background conditions. The melting of lead is predicted to occur when the forces associated with the thermal motion of the free electrons exceed the electromagnetic forces holding the electrons and nuclei in a solid lattice structure or, as Bogen and Woodward put it, "the melting of lead occurs whenever samples of lead are present at the appropriate temperature and pressure, and results from a characteristic change in the crystalline structure of this metal." (1988, 319). There are *ab initio* theoretical predictions for melting points of metals[74] and there are experimental measuring procedures. While experimental measures deliver a range of data points, their average is of 327.5 degrees Celsius, though no single measurement may precisely matching that value. When does the relatively close

---

[74] For an example of an ab initio prediction of the melting temperature of a metal see: "Thermodynamic calculation of the melting point of aluminum" (Belan-Gaiko, et al 1980).

agreement between the theoretically predicted value and data obtained from multiple different measuring techniques warrant belief in the reality of the phenomenon?

For the inference to realism from convergence to be justified, the range of variation of the measured values within and between experiments has to be sufficiently narrow for the practical facts of the experiment to be interpreted as satisfying the theoretical fact of the prediction. But what if a new type of measuring technique is developed that in multiple replications indicates a wildly different value? Based on the past stability of measurements we might infer that the new protocol is not valid. Or, if there is theoretical justification for the new technique, it might explain why the value is far from the older techniques by identifying some systematic bias shared by all the older techniques. We might accept the new results and revise the account of the phenomenon based on it.

On the strict lawful regularity account of causation, science cannot tolerate inconsistent results. This account of causality requires a strict reading of Bogen and Woodward's 1988 claim the phenomenon will occur "*whenever* samples of lead are present at the appropriate temperature and pressure." Of course one could question whether the "appropriate" conditions are met. But presuming a shared judgment of what is appropriate, the strict laws view of causation would demand there be no exceptions. However, Woodward's interventionist account permits causation with degrees of invariance under intervention, rather than all or nothing strictness. If the conditions of the new experiment push the phenomenon outside the range of its causal invariance, both the new and old measurements can be deemed accurate, and hence both experimental processes would be causally reliable. As I will argue below, to decide if an experimental result inconsistent with the theoretical prediction is reliable (and hence a refutation for the strict law view of causality) or reliable but merely reflecting the boundaries of the range of invariance (not necessarily a refutation of on the invariance view of causality) or not reliable at all, requires a theory about the type or features of the phenomenon being investigated.

The significance of theory in determining reliability of experiments and inference to the reality of the phenomenon becomes clearer when we consider experimentation on dynamically complex phenomena. Consider cases where a system-level property is caused by behaviors of the components of that system. Lead melts at a temperature (its melting point) when the thermal motions of the free electrons in the atoms composing the lead breaks the electromagnetic forces holding the electrons and nuclei in a solid lattice. There may be exceptions, i.e. lead fails to melt at its "melting point", even when the functional relationship between the variables for thermal motion and electromagnetic forces satisfy the theory, but this case is pretty close to a strict generalization.[75]

Complex systems phenomena, like the genotype-phenotype relationship where a gene is taken to be the cause of the trait (e.g. "gene for" language in the gene for Huntingdon's chorea, or the gene for melanism in peppered moths, etc.), are less stably realized, more fraught by exceptions, than melting points of metals. Depending on the account of causation, and the theory of the phenomenon, the reliability of experimental detection of the phenomenon will be judged differently. This is vividly displayed in the history of gene knock-out experiments (see Mitchell 2009) where intervening to remove (or silence) particular genes in an organism is used to identify their phenotypic function or effect.[76] Comparing the phenotypes of organisms with, and without, a particular gene follows the logic of a controlled experiment. Intervene on only one gene, leaving the rest of the genome intact, and any changes to the phenotype should expose the causal effect of that gene. In roughly 30% of experiments with viable knockout mutants, there is there is little or no phenotypic difference. For some, including Mario Capechi who invented the knock-out technique, the inference from these "failed" experiments was that the experiment was not done correctly and, therefore, the results are not reliable. For others, the best inference was that the target gene having little phenotypic effect in a relatively high percentage of cases, should not be identified as the cause of the phenotype. A third alternative was proposed (Greenspan 2001) that is based on the theory that there is a robust dynamics of the network of genes that

---

[75] Of course, there is the standard move in the strict law camp to introduce ceteris paribus conditions to preserve universal exceptionless. The interventionist might appeal to something similar in determining what is the "appropriate" conditions, but if interventionism is to be an alternative that permits "degrees of invariance" in contrast to strictness it must permit some attribution of causation in cases where the strict view would not. See also Woodward (2001) and Mitchell (2002).

[76] Scientific techniques have advanced in precision since knock-out experiments with the development of CRSPR but the issues of the warrant for reliability and realism occur with all experiments.

interact in the causal production of a phenotype. On Greenspan's account, when one gene is silenced, there is a reorganization of the network of genes such that, in many of knockout mutants, the normal phenotype is still produced, but by a different causal pathway. This is not a case of redundancy, where there are back up copies of genes that step in when one token is removed, but a case of "degeneracy," (Edelman and Gally 2001) where new causal networks of interaction arise in the absence of the knocked-out node. "The relationships that have been described as pathways are no doubt real, but they need not be invariant. Their relationships are embedded in broader and more plastic networks that can be reconfigured depending on the immediate circumstances." (Greenspan 2001 p 386).

If data from different experiments converge, then the inference is that the experiments are reliable and the phenomenon is real. But if data from different experiments diverge, then, as in the case of knockout experiments, what is inferred exposes how judgments about reliability and reality are shaped by theories of the phenomena and philosophical accounts of causation.

Woodward (2009) further elaborated his analysis of causation broadening the scope of judgments that can be made about causation beyond a dichotomous framework in which something either is a cause or is not a cause. He suggests that instead there are important distinctions among causes with respect to degrees of stability, level of causal description and causal specificity drawing on examples of genetic causation. I have also argued that a more nuanced account of causal dependence (Mitchell 1997, 2000, 2009) better captures scientific practice and permits us to see the similarities and differences among causal claims made in different sciences. Here, I want to emphasize how a more nuanced view exposes the interactive roles of multiple assumptions in making causal judgments, and thus in shaping and justifying claims of realism.

The reliability of knockout experiments is indicated by but not constituted by their replicability. What counts as a successful or failed replication depends on both a theory of the type of phenomenon being tested and the causal features and pathways that investigators associate with experimental interventions. Are traits caused by individual genes, or feedback sensitive reorganizing networks of genes? Are causes maximally stable, expressing universal laws or are causal relationships stable only in some domains of invariance, but unstable outside of those domains? If a knockout experiment to determine the genetic cause of a phenotypic trait does not show a difference between organisms with or without that gene, or if multiple knockout experiments diverge, then inferences to what is real will track the answers to these questions. On a strict regularity laws view of causation, any failure of stability, no effect, or diverging effects entail that either one or more experiments is unreliable, or there is no real causal phenomenon. These were both responses to the surprising result that 30% of knockout experiments fail to display a stable gene-phenotype relationship. On an invariance under intervention view of causation, in the same situation, one could infer that one or more experiments was unreliable, if the expectation of the theory of the phenomenon was for invariance under the conditions of the experiment. But one could legitimately infer that either the conditions of the experiment were outside the domain of invariance (the experiment was reliable) or the phenomenon was complex, robust and hence its stability would not be captured by the gene knockout experiments (the experiment was reliable but not for that type of phenomenon).

The claim that a phenomenon is real acquires empirical warrant from the reliable replication of a single type of experiment and even more so from reliable reproduction of results from diverse types of experiments. Underlying this warrant are judgments about the reliability of the experiments to causally engage certain features of the phenomenon which it targets. Instrumentally "observed" causal processes are perspectival, detecting only those invariant affordances which make a difference to the actions of detection devices. To go beyond the Duhemian practical facts that reliable causal interactions produce, in order to attribute reality to more than local, idiosyncratic results requires invoking abstract, theoretical representations and abstract structures, i.e. theories of the phenomenon (simple, complex, robust, etc.) Only then can multiple experiments of the "same" phenomenon accrue the added-value support of convergence inferences for realism. Clearly, what is permitted, condoned or forbidden by a concept of and criteria for causality itself will also guide judgments of reliability and hence of realism. To understand why there is a scientific disagreement about whether or not a purported phenomenon is real or not, we need to investigate all three contributing factors in establishing the real affordances of nature – experimental data, theoretical representations and the causal framework. Real phenomena, external to an observer afford detection by means of their causal interactions as characterized by theories of the phenomena – biased or perspectival though they be, and afford representation by functional or structural theories of causal relations, partial and conceptually bound though they be. As experiences, experimental techniques, conceptual re-orientations and theoretical innovations change, our warrant for claiming which phenomena are real will track those changes.

6: Conclusion

What justifies a claim of realism in science is a function of the coordination of experimental data (causal reliability), theoretical expectation (types of phenomena with differing degrees or types of stability) and a theory of causation (strict laws, invariance, etc.). What is real is an *affordance whose identification is built from the coordination of reliable interaction with what is external to us and the theoretical and philosophical frameworks constraining and structuring what we can represent*. Ian Hacking (1992) argued for anti-realism by appeal to a similar coordination claiming that "Stable laboratory science arises when theories and laboratory equipment evolve in such a way that they match each other and are mutually self-vindicating (1992, 56) and but that when that happens "we have not read the truth of the world." (58). Rather,

> There usually were not some preexisting phenomena that experiment reported. It made them. There was not some previously organized correspondence between theory and reality that was confirmed…The process of modifying the workings of instruments – both materially…and intellectually…furnishes the glue that keeps our intellectual and material world together. It is what stabilizes science. (Hacking 1992, 58)

I suggest that the bearable thinness of being permits a pragmatist realism that reflects the role of human judgments about reliability, types of phenomena and causation in identifying the real affordances that nature permits. This non-fundamentalist, pragmatist approach to realism embraces the contingency of judgments about what is real that depend on what makes a difference both for causal detectability (bottom-up) and structural representability (top-down). This is the kind of realism our scientific practices can bear.

Acknowledgments: I would like to thank Holly Andersen, Joel Smith and Dana Matthiessen for extremely helpful feedback on this chapter, and all those who participated in the PragMaps workshops for stimulating discussions.

References:

Acton, T. B., Xiao, R., Anderson, S., Aramini, J., Buchwald, W. A., Ciccosanti, C., ... & Montelione, G. T. (2011). Preparation of protein samples for NMR structure, function, and small-molecule screening studies. In *Methods in enzymology* (Vol. 493, pp. 21-60). Academic Press.

Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. *Science*, *294*(5540), 93-96.

Belan-Gaiko, L. V., Bogdanov, V. I., & Fuks, D. L. (1980). Thermodynamic calculation of the melting point of aluminum. *Soviet Physics Journal*, *23*(12), 1009-1011.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The philosophical review*, *97*(3), 303-352.

Boyd, Nora Mills and James Bogen, "Theory and Observation in Science", *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2021/entries/science-theory-observation/>.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological bulletin*, *54*(4), 297.

Cartwright, N. (1982). When explanation leads to inference. *Philosophical Topics*, *13*(1), 111-121.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press.

Chakravartty, A. (2021) Realist Representations of Particles: The Standard Model, Top-Down and Bottom-Up. in T. D. Lyons & P. Vickers (eds.), Contemporary Scientific Realism: The Challenge from the History of Science, pp. 350-373 (2021), Oxford University Press

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford University Press.

Chang, H. (2022). Realism for Realistic People: A New Pragmatist Philosophy of Science.

Chirimuuta, M. (2017). *Outside color: Perceptual science and the puzzle of color in philosophy*. MIT press.

Darling, K. M. (2002). The complete Duhemian underdetermination argument: scientific language and practice. *Studies In History and Philosophy of Science Part A*, *33*(3), 511-533.

Darling, K. M. (2003). Motivational realism: The natural classification for Pierre Duhem. Philosophy of science, 70(5), 1125-1136

Deng, H., Jia, Y., & Zhang, Y. (2018). Protein structure prediction. *International Journal of Modern Physics B*, *32*(18), 1840009.

Duhem, P. M. M. (1991). *The aim and structure of physical theory* (Vol. 13). Princeton University Press.

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, *98*(24), 13763-13768.

Epstein, B., & Forber, P. (2013). The perils of tweaking: how to use macrodata to set parameters in complex simulation models. *Synthese*, *190*(2), 203-218.

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution?. *Acta Crystallographica Section D: Biological Crystallography*, *69*(7), 1204-1214.

Franklin, A. and Perovic, S. (Summer 2021 Edition). Experiment in Physics. *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/physics-experiment/>.

French, S., & Ladyman, J. (2010). In defence of ontic structural realism. In *Scientific structuralism* (pp. 25-42). Springer, Dordrecht.

Frisch, Mathias, "Causation in Physics", *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2022/entries/causation-physics/>.

Galison, P. (1988). History, philosophy, and the central metaphor. *Science in Context*, *2*(1), 197-212.

Gallow, J. Dmitri, "The Metaphysics of Causation", *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/sum2022/entries/causation-metaphysics/>.

Gibson, J. J. (2014). *The ecological approach to visual perception: classic edition*. Psychology Press.

Giere, R. N. (1999). *Science without laws*. University of Chicago press.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, *172*(2), 269-281.

Greenspan, R. J. (2001). The flexible genome. *Nature Reviews Genetics*, *2*(5), 383-387.

Guala, F. (2003). Experimental localism and external validity. *Philosophy of science*, *70*(5), 1195-1205.

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

Hacking, I. (1992). The self-vindication of the laboratory sciences. *Science as practice and culture*, *30*.

Hitchcock, Christopher, "Probabilistic Causation", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/causation-probabilistic/>.

Humphreys, P. (1991), "Computer Simulations", in A. Fine, M. Forbes, and L. Wessels (eds.), PSA 1990, vol. 2. East Lansing, MI: Philosophy of Science Association, pp. 497-506., J

Justus, J. (2019). Ecological Theory and the Superfluous Niche. *philosophical topics*, *47*(1), 105-124.

Kuhn, T. S. (1976). Mathematical vs. experimental traditions in the development of physical science. *The Journal of Interdisciplinary History*, *7*(1), 1-31.

Kuorikoski, J., & Marchionni, C. (2016). Evidential diversity and the triangulation of phenomena. *Philosophy of Science*, *83*(2), 227-247.

Mitchell, S. D. (1997). Pragmatic laws. *Philosophy of science*, *64*, S468-S479.

Mitchell, S. D. (2000). Dimensions of scientific law. *Philosophy of Science*, *67*(2), 242-265.

Mitchell, S. D. (2002). Ceteris paribus—an inadequate representation for biological contingency. *Erkenntnis*, *57*(3), 329-350.

Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge University Press.

Mitchell, S. D. (2009). Unsimple truths. In *Unsimple Truths*. University of Chicago Press.

Mitchell, S. D., & Gronenborn, A. M. (2017). after fifty years, why are protein X-ray crystallographers still in business?. *The British Journal for the Philosophy of Science*.

Mitchell, S. D. (2020a). Through the Fractured Looking Glass. *Philosophy of Science*, *87*(5), 771-792.

Mitchell, S. D. (2020b). Perspectives, representation, and integration. *Understanding Perspectivism: Scientific challenges and methodological prospects*, 178-193.

Morris, G. A. (1992). Systematic sources of signal irreproducibility and t1 noise in high-field NMR spectrometers. *Journal of Magnetic Resonance (1969)*, *100*(2), 316-328.

Price, H. (2007). Quining naturalism. *The Journal of Philosophy*, *104*(8), 375-402.

Ramsey J. L. (1997) Between the Fundamental and the Phenomenological: The Challenge of 'Semi-Empirical' Methods, *Philosophy of Science*, 64 pp. 627-653.

Reiss, J. (2019). Against external validity. *Synthese*, *196*(8), 3103-3121.

Rydenfelt, H. (2021). Realism without representationalism. *Synthese*, *198*(4), 2901-2918.

Salmon, W. C. (1985). Scientific explanation and the causal structure of the world. In *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Schindler, S. (2011). Bogen and Woodward's data-phenomena distinction, forms of theory-ladenness, and the reliability of data. *Synthese*, *182*(1), 39-55.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.

Suppes, Patrick (1962) "Models of Data", in E. Nagel, P. Suppes, and A. Tarski (eds.), Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress. Stanford University Press, 252-261.

Steinle, F. (2002). Experiments in history and philosophy of science. *Perspectives on science*, *10*(4), 408-432.

Tal, E. (2017). Calibration: Modelling the measurement process. *Studies in History and Philosophy of Science Part A*, *65*, 33-45.

Teller, P. (2010). "Saving the phenomena" today. *Philosophy of Science*, *77*(5), 815-826.

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron*, *81*(6), 1240-1253.

Vetter, B. (2020). Perceiving potentiality: A metaphysics for affordances. *Topoi*, *39*(5), 1177-1191.