If mathematical psychology did not exist we would need to invent it: A case study in cumulative theoretical development

Danielle J. Navarro[1]

[1] School of Psychology, University of New South Wales

Abstract

It is commonplace, when discussing the subject of psychological theory, to write articles from the assumption that psychology differs from physical sciences in that we have no psychological theories that would support cumulative, incremental science. In this brief paper I discuss one counterexample, namely Shepard's (1987) universal law of generalization and the various Bayesian extensions that it inspired over the last three decades. Though not disputing the claim that good theories in psychological science are rarer than one would like, I suggest that the subdiscipline of mathematical psychology is a good place to look to find them.

Keywords: psychological theory, inductive generalization, mathematical psychology, cognitive modelling
Word count: 4336 in main text, 576 in references

> *We generalize from one situation to another not because we cannot tell the difference between the two situations but because we judge that they are likely to belong to a set of situations having the same consequence. Generalization, which stems from uncertainty about the distribution of consequential stimuli in psuchological space, is thus to be distinguished from failure of discrimination, which stems from uncertainty about the relative locations of individual stimuli in that space* (Shepard 1987, p. 1322)

Introduction

In 1987 Roger Shepard published a short paper in *Science* with the ambitious title "Toward a universal law of generalization for psychological science" (Shepard, 1987). Drawing on the empirical literature on stimulus generalization in several domains and species, he asserted the claim that the form of any stimulus generalization function should be approximately exponential in form, when measured with respect to an appropriately formulated stimulus representation. His paper begins with the following remark (p. 1317):

> *The tercentenary of the publication, in 1687, of Newton's "Principia" prompts the question of whether psychological science has any hope of achieving a law that is comparable in generality (if not in predictive accuracy) to Newton's universal law of gravitation. Exploring the direction that currently seems most favorable for an affirmative answer, I outline empirical evidence an a theoretical rationale in support of a tentative candidate for a universal law of generalization*

Shepard's claim was remarkable in scope. He drew on data from multiple species (e.g., humans, pigeons, rats) and stimulus domains (e.g., visual, auditory) data that had, until that point, been assumed to be quite different to one another. To spot the invariance that holds across these data sets, Shepard used statistical insights from the similarity modelling literature. He noted that the apparent noninvariance of observed stimulus generalisation functions stemmed largely from the fact that response data had previously been analysed with respect to the physical dissimilarities of the stimulus. When the same responses were replotted as a function of distance in a psychological space contructed by multidimensional scaling, he found that the form of the stimulus generalisation was remarkably regular in shape, as shown on the left side of Figure 1.

Taken by itself Shepard's reanalysis would have been impressive. However, Shepard went on to provide a theoretical explanation for *why* we should expect to find this invariance. The theory was surprisingly simple: the learner presumes there exists some unknown *consequential* region of the stimulus space across which roughly the same properties hold (e.g., things that look like apples will probably taste the same as one another). Encountering a single stimulus that entails a particular consequence, the learner's task is to infer the location, shape and size of the consequential region itself. Naturally this is an underconstrained problem, as there are an infinite number of possible regions that might correspond to the true consequential region. Nevertheless, Shepard showed that under a quite range of assumptions that the learner might make about the nature of consequential regions, the shape of the *generalization* function across the stimulus space ends up approximately
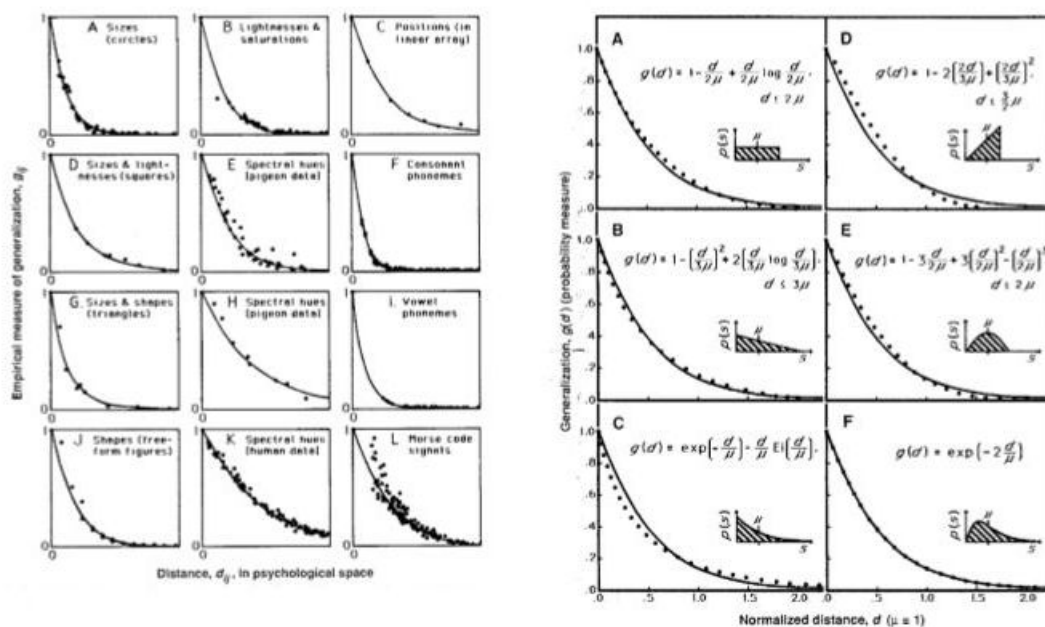
Figure 1. Figures from Shepard's (1987) paper on stimulus generalisation. On the left (Figure 1 in the original) each of the 12 panels displays a generalization gradient, plotting the distance d_{ij} between the observed stimulus and the test stimulus in psychological space on the x-axis, and the empirical measure of generalization g_{ij} on the y-axis. The panels represent data from a variety of different species and stimulus domains. On the right (Figure 3 in the original), six theoretical generalization functions derived under different assumptions about $p(s)$ (dotted lines), and the corresponding exponential decay. For additional detail see the original paper. *[Note: if accepted, higher resolution versions will be provided in the final version and permission will be obtained to reproduce the originals]*

exponential, shown on the right side of Figure 1.

Although brief, Shepard's paper has been highly influential in the cognitive science literature. It presented no new empirical data, and in substance it is mostly devoted to the derivation of a formal relation between one unobservable quantity (psychological distance) and another (generalisability). The universal law featured prominently in a special issue of *Brain and Behavior Sciences* in 2001 and a first person retrospective (Shepard, 2004). In this special issue devoted to the prospects for theoretical development in psychology, I can think of no better case study to illustrate points I would like to make about the nature of theoretical work in psychology. My goal in this paper is to discuss the role that theory can play in the development of psychological science, using Shepard's law to illustrate.

How should we think about psychological theories?

I tentatively suggest that because these regularites reflect universal principles of natural kinds and of probabilistic geometry, natural selection may favor their increasingly close approximtion in sentient organisms wherever they evolve. (Shepard, 1987, p. 1323)

Reading Shepard's original 1987 paper and the 2004 retrospective, some surprising characteristics of his theoretical work stand out. First, the theoretical development was largely post hoc. The paper does not collect new data, and indeed the main empirical results reported in the paper (left hand side of Figure 1) were based on a reanalysis of existing data. Second, the paper reports no hypothesis tests or indeed any statistical inference of any kind. There are no p-values, no Bayes factors, nor any confidence intervals or their Bayesian equivalents. Third, the paper does not outline any specific novel prediction about the result of future experiments. Though the paper makes a strong claim that the exponential law should hold broadly it does not prescribe how tests of this prediction should be constructed.

Viewed through the lens of the methodological reform culture (documented by Flis, 2019) all three of these properties might seem strange, and might even amount to a form of "questionable research practice". To illustrate, consider the question of strong inference. In the current zeitgeist, strong claims cannot be justified without preregistered confirmatory tests (Wagenmakers, Wetzels, Borsboom, Maas, & Kievit, 2012) but Shepard presents no tests at all while making sweeping claims. His theorising is entirely post hoc, and would at first blush appear to be a form of "HARKing" (hypothesizing after results are known; Kerr, 1998). Judged by the standards currently proposed for the assessment of psychological research, Shepard's paper might be expected to fare quite poorly. One might even be tempted to ask how Shepard managed to get the paper published at all.

Something seems awry in this description. Few researchers in the mathematical psychology community would consider Shepard (1987) to be "bad science" yet even a cursory inspection suggests it fails to meet some of the basic quality control criteria proposed as a remedy for the discipline's ills. The resolution to the discrepancy, I suggest, is that methodological criteria in psychology are typically developed to guide statistical inferences about empirical data, and – as I have argued before (Navarro, 2019) – are not well suited to the evaluation of scientific theories. Or to put it another way, one might argue that the success of Shepard's (1987) paper tells us a great deal about what a theory is not. Novel empirical data, no matter how fascinating or carefully collected, are not a theoretical contribution, and Shepard's theoretical work in his 1987 paper did not rely on a single new data set. Equally, a statistical test does not encapsulate what a theory does, and accordingly Shepard's theoretical contribution was neither reliant on nor presented as a statistical claim of any kind. Perhaps surprisingly, discovery need not lie at the core of a theoretical claim either: the theoretical component to Shepard's paper was not the discovery of an exponential law, but rather the explanation proposed to account for it.

Empirical data, statistical inference and discovery are all vital to the scientific process – of that there is little doubt – but none of them are substitutes for or synonymous with theoretical development. If so, what might a theoretical explanation be? On my reading, what Shepard's paper did was *systematise* an existing body of empirical findings. It separated those aspects to the data that are invariant across studies from those that are not. To do so it provided a precisely stated mathematical formalism of those invariants, and thus provides a conceptual framework that licence radical inferences beyond the scope of the original data. That is, while Shepard's theory asserts that the form of a generalisation

curve should be exponential, this exponential law is not the theoretical primitive. Rather, the theoretical primitives are the assumptions Shepard made about consequential regions, psychological spaces, and the relationship these things have to stimulus generalisation.

It is precisely because his theory is built from such abstract primitives that Shepard justifies the generality of his claim. In short, the exponential law is an *entailment* of the theory, not its substance. If the exponential law were merely observed in a few terrestrial species there would be little reason to suggest that the law should be expected to hold beyond our own planet. Such inference would be statistically unjustifiable, even as a "tentative suggestion". Shepard makes this claim, not by relying on any statistical tool, but by noting that the exponential law emerges *as* an entailment of more primitive rules that could be reasonably expected to hold in vastly different environments. It is this kind of "radical generalisation" that I argue is the key to good theorising.

A good theory can escape its measurement context

In the remainder of this paper I want to use Shepard's law to motivate a discussion of what a good theory can achieve. Some desiderata for scientific theories seem easy to list. A scientific theory should be independent of its creator, for instance. It is difficult to make much use of a theory otherwise. In practice this typically means a theory is mathematical or computational in nature. Similarly, psychological theories should of course make some connection with empirical data, giving some account of the generative mechanism that gave rise to those data. Theories should be usable, in the sense of providing other scientists guidance for future research. Other criteria could also be named, including falsifiability, simplicity, compatibility with existing literature, generalizability, predictive ability and so on. However, while it is easy to list desiderata and even easier to argue over which elements to such lists are the most important, such "discussions in the abstract" rarely provide much guidance to the would-be theoretician. From the perspective of the working scientist, it is perhaps more useful to give concrete examples, and to that end I return to an examination of Shepard's (1987) paper.

Consider first the problem of measurement. If one is to build a strong theory, it must be tethered in some way to observational or experimental data, and to do so one must have an appropriate measurement tool. For example, one of the key insights in Shepard's (1987) paper is the recognition that stimulus generalization functions that appear extremely irregular at first glance are in fact highly regular when stimuli are represented in an appropriate fashion, as depicted in Figure 2. Generalization to tones is irregular when tones are described in terms of their ordering around the chroma circle, but very smooth when described around the circle of fifths. Similarly, stimulus generalization to colours is irregular with respect the physical wavelenghts of light but is smooth with respect to perceptual colours space (e.g., Ekman, 1954). In retrospect this seems obvious, but at the time Shepard developed the theory, he was faced with a substantive problem of how to extract the *appropriate* stimulus representation to apply a theory to. Nonmetric multidimensional scaling (MDS; Kruskal, 1964) as a measurement instrument tool served this purpose for Shepard in 1987, and the smooth empirical curves on the left of Figure 1 all use MDS-estimated psychological spaces to supply the relevant measure of distance on the x-axis.
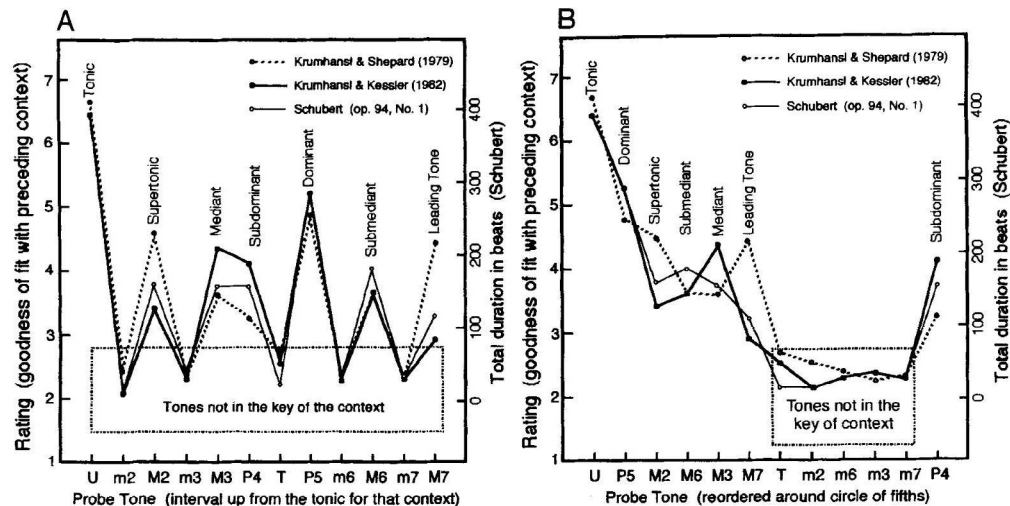
Figure 2. Illustrations taken from Shepard's (2004) retrospective account of the universal law, plotting stimulus generalization data from three previous experiments in two different formats. On the left, data are shown in a raw format, and on the right data are recast in terms of constructs relevant to the theory. See original for more detail. *[Note: a higher resolution version will be provided in the final version and permission will be obtained to reproduce the originals]*

As this discussion illustrates, without MDS as a measurement tool Shepard would have found it almost impossible to formulate the empirical regularity with any confidence. However, it is equally clear that MDS is merely a tool used to help define the phenomenon to be explained. It does not define the theory itself, and in the relevant literature it quickly became apparent that Shepard's law applies even in situations where MDS does not: shortly after the publication of Shepard's original paper, Russell (1988) demonstrated that the same law holds for stimuli defined in terms of discrete features as well as to the continuous spaces for which Shepard's work was defined, a connection that was later extended by Tenenbaum and Griffiths (2001). While the theoretical framework could not have come into existence without the scaffolding provided by the MDS measurement model, it quickly outgrew any need for this support. Illustrating this, the generalization gradient shown in Figure 3 is neither exponential in superficial form, nor is the underlying stimulus representation a metric space extracted by MDS, yet as Tenenbaum and Griffiths (2001) note, it too can be seen as an entailment of Shepard's theoretical framework.

A good theory can escape its formal description

The second point to consider is the role of formalism. The mathematical form that Shepard used to implement his ideas is not unique, and the theory can be rewritten in different notation. Previously, Cooper and Guest (2014) have argued that theoretical work need not be constrained to a particular "implementation" (or formalism), but is better captured by a more abstract notion of a "specification". As a concrete example, it is worth considering
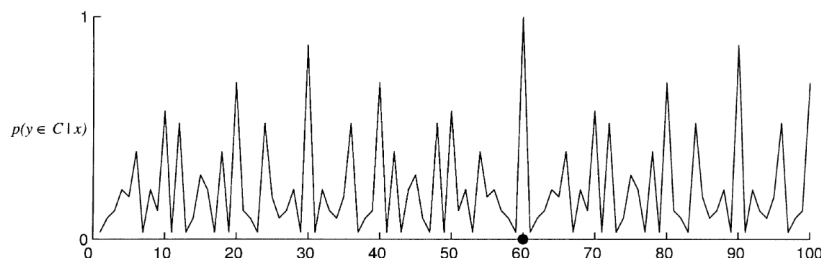
Figure 3. A Bayesian generalization gradient for stimuli (numbers) that are more naturally represented in terms of discrete consequential sets, reproduced from Tenenbaum and Griffiths (2001). In this example, the learner knows that a property holds for the number "60", and must judge the probability that this novel property holds for other numbers. The hypothesis space $\mathcal{H}$ consists of sets of consecutive numbers (to capture the notion that people are sensitive to magnitude), but also includes a variety of mathematical properties (even numbers, odd numbers, multiples of three, etc). See original for more detail. *[Note: a higher resolution version will be provided in the final version and permission will be obtained to reproduce the originals]*

the manner in which Shepard's law was later reformulated as a Bayesian model,[1] and the effect this rewriting had on how the model could be applied.

In his original notation Shepard proposed that the generalization strength for a stimulus located at coordinate $\mathbf{x} = (x_1, \ldots, x_K)$ in a K-dimensional space,[2] when the learner has previously encountered a consequential stimulus located at the origin be expressed as follows:

$$g(\mathbf{x}) = \int_0^\infty p(s) \frac{m(s, \mathbf{x})}{m(s)} ds \qquad (1)$$

where the ratio $m(s, \mathbf{x})/m(s)$ is a measure describing the probability that a consequential region of size s contains $\mathbf{x}$ given that it also contains the consequential stimulus at the origin, and $p(s)$ is a probability measure that assigns value to regions of size s. Much of Shepard's paper is devoted to solving this integral for various choices of $p(s)$ and showing that the resulting generalization function $g(\mathbf{x})$ is approximately exponential in form, and is precisely exponential when $p(s)$ is the Erlang distribution.

In and of itself Equation 1 does not describe a Bayesian model. The connection to Bayesian inference was made explicit by Tenenbaum and Griffiths (2001) who recast Shepard's formalism using different notation and expressing the underlying ideas rather

---

[1]Because Bayesian inference is often considered to be a normative standard for learning, it should be noted that some care is required in assigning normative status to Bayesian models of cognition; see Tauber, Navarro, Perfors, and Steyvers (2017).

[2]The space is usually endowed with one of the Minkowski distance metrics, most typically the Euclidean distance measure or the city block distance. One of the major contributions Shepard made was to provide a theoretical explanation for why the distance metric varies across stimulus domains, but is a little beyond the scope of the current paper.

differently. Where Shepard referred to the notion of a "consequential region" located within a psychological space with all the geometric connotations that entails, Tenenbaum and Griffiths took a more general view and framed their analysis in terms of "consequential sets". Any specific candidate for the true consequential set was labelled a "hypothesis" h, presumed to belong to a broader "hypothesis space" $\mathcal{H}$. In their notation, the measure on regions sizes $p(s)$ was replaced by a Bayesian prior $p(h)$ defined over the hypothesis space, and Shepard's ratio $m(s, \mathbf{x})/m(s)$ recast in probabilistic terms also. In their notation, the probability of generalising to a novel stimulus y given that a consequential stimulus x has already been observed is formulated as the following Bayesian marginal probability:

$$p(y \in C|x) = \sum_{h:y\in C} p(h|x) \tag{2}$$

where $p(h|x)$ is the posterior probability a Bayesian learner assigns to hypothesis h in light of observing the consequential stimulus x,

$$p(h|x) = \frac{p(x|h)p(h)}{\sum_{h'\in\mathcal{H}} p(x|h')p(h')} \tag{3}$$

where $p(x|h)$ is the likelihood that the learner would encounter the data x if hypothesis h were true, and the summation in the denominator is taken over all hypotheses in the learner's hypothesis space.

The Bayesian reformulation of Shepard's theory presented by Tenenbaum and Griffiths was more than a superficial change: it allowed them to show how Shepard's theory could itself be generalized in three distinct ways. First, as mentioned earlier, they showed (much like Russell 1988) that Shepard's theory could encompass stimuli that were not representable as points in a metric space: in their notation, this is as accomplished by substituting a new hypothesis space $\mathcal{H}$. Second, this formulation allowed the theory to naturally accommodate inductive generalization problems in which the learner has encountered more than one consequential stimulus. If a set of n stimuli $\mathbf{x} = (x_1, \ldots, x_n)$ are generated in a (conditionally) independent fashion, the likelihood $p(\mathbf{x}|h)$ factorises and thus

$$p(\mathbf{x}|h) = \prod_{i=1}^{n} p(x_i|h). \tag{4}$$

Finally, this formalism called attention to a potentially limiting assumption in Shepard's original paper. Shepard (1986, p. 1321) argued that "in the absence of any information to the contrary, an individual might best assume that nature selects the consequential region and the first stimulus independently". According to this account, which Tenenbaum and Griffiths termed *weak sampling*, this independence gives a uniform likelihood function:

$$p(x|h) \propto 1 \tag{5}$$

They contrasted this with a *strong sampling* account in which stimuli are presumed to be sampled uniformly from the consequential set. If $|h|$ denotes the size of the consequential

set[3] then a strong sampling process produces the likelihood

$$p(x|h) = 1/|h| \qquad (6)$$

if x is an element of the consequential set, and 0 otherwise.

While a complete description of the relationship between Shepard's formalism and Tenenbaum and Griffiths' is beyond the scope of this paper, some things seem clear enough. In one sense the two formalisms capture the "same" underlying theoretical idea, but that does not make them equivalent. As I outlined above, the Bayesian reformulation encourged researchers to apply the model in different ways. By "escaping" the original notation Shepard used to describe it, the theory could be extended to quite different problems, as I describe in the next section.

A good theory can escape its empirical domain

The third point I wish to consider is the way in which theories can extend their domain of applicability. Looking at their paper in retrospect, one of the most important contributions made by the Bayesian formulation adopted by Tenenbaum and Griffiths (2001) is that it allowed the underlying theory to be applied in a much broader range of scenarios. Shepard's (1987) original construction, though purportedly to be a very general law itself, was formulated with respect to a narrow class of psychological problems: inductive generalization from a single observation. Moreover, because the origins of his work lay in the study of human perception and the animal learning literature, it was not immediately clear – at least it was not clear to me – how the theory should be extended to higher order cognition. By framing the problem in a more general way, Tenenbaum and Griffiths' construction enabled a large proportion of my later research. Examples of research questions that I would have been unable to properly formulate without this theory:

- What kind of individual differences exist in the sampling assumptions people adopt in simple generalization problems? (Navarro, Dry, & Lee, 2012)
- Do categorization problems and inductive generalization problems lead people to adopt different sampling assumptions? (Hendrickson, Perfors, Navarro, & Ransom, 2019)
- Do people's beliefs about the origins of data and shape how we interpret the evidentiary value of the quantity (Ransom, Perfors, & Navarro, 2016) and diversity of evidence (Hayes et al., 2019b)?
- Can we use the Bayesian theory to unify and extend previously puzzling results (Lawson & Kalish, 2009) on how sampling affects reasoning? (Hayes et al., 2019a)
- Can we build a Bayesian model of how people "take a hint" from ostensibly uninformative data? (Voorspoels, Navarro, Perfors, Ransom, & Storms, 2015)
- Completing a "virtuous cycle" of sorts, recent work has taken our results from the reasoning context (Voorspoels et al., 2015) and applied it to make novel predictions in the associative learning (Lee, Lovibond, Hayes, & Navarro, 2019)

---

[3]For finite sets this $|h|$ is simply the number of entities in the set; for infinite sets, a more careful approach to measuring the size of sets is required

This is of course a highly selected portion of the relevant literature, focusing as it does on work in which I was personally involved. The purpose of choosing these examples is that – as the person developing the theoretical extensions in each case – I am in a position to comment on exactly *how* the theoretical formalism from Tenenbaum and Griffiths (2001) paper enabled this research when the original formalism from Shepard (1987) would have made it difficult. Because Shepard's formalism adhered closely to the specific problem he sought to explain (generalization from a single observed stimulus represented in a metric space) it was not clear to me prior to 2001 that the learner is necessarily reliant on an assumption (explicit or implicit) about how the stimulus was called to the learner's attention. Under the Bayesian reformulation of the theory, this assumption was rendered visible via the likelihood $P(x|h)$. When written in this form, it was painfully clear to me that the inductive inferences any learner makes about a novel item y is necessarily dependent on more than the observed data x – it also depends on the assumptions embodied in the likelihood $P(\cdot|h)$ as to how those data were collected. This link, present but obscured in Shepard's (1987) paper, was the central motivation behind this entire body of work and would not have been possible without the more abstract formalism introduced by Tenenbaum and Griffiths (2001).

Conclusion

Those outside the field, on hearing the term mathematical psychology, often react with the raised eyebrows reserved for oxymora (Shepard, 2004, p. 1)

Mathematical psychology is something of an oddity in the discipline. It does not eschew empirical research, but neither does it view the goal of science to be the accrual of empirical effects. Quite unlike most areas of psychology with which I am familiar, mathematical psychologists place a high value on theoretical development, particularly when such theories can be stated in a formal manner. My goal in this paper was to highlight the manner in which cumulative theoretical work has developed in this discipline, using Shepard's law as an example. From its origins in associative learning and stimulus generalization, to its reformulation as a Bayesian model and its extension to a variety of novel contexts, a single theoretical claim can be shown to connect to a variety of empirical findings in superficially distinct domains. The breadth of situations to which the law can be unambiguously applied is greatly to be desired in a theoretical explanation.

Although I have focused on Shepard's law and its extensions in this paper, the underlying pattern is – I would suggest – quite general. I could have chosen the Rescorla-Wagner model of associative learning as the basis for this discussion (Rescorla & Wagner, 1972), or the Generalized Context Model of human categorization (Nosofsky, 1986). I could have chosen to focus on models such as ALCOVE that sought to unify associative learning and categorization (Kruschke, 1992), or models such as the hierarchical Dirichlet process that seek to unify a wide variety of category learning models within a common theoretical language (Griffiths, Canini, Sanborn, & Navarro, 2007). I could have revisited Ebbinghaus' 1885 work on formal theory in experimental psychology (reprinted as Ebbinghaus, 2013). I could have examined sequential sampling models of choice reaction time (Luce, 1986) and the rich theoretical tradition that mathematical psychologists have developed in that domain also.

These theoretical advances have something in common. In each of these areas psychological researchers have built up a considerable body of theoretical knowledge that is instantiated in formal models of psychological processes. In every case the underlying theoretical models are more than mere summaries of empirical results, and more substantive than a mere statistical model. In all cases the formalism can be used to generate novel predictions in experimental paradigms that differ markedly from the experimental contexts used to develop the model (and, remarkably, some of those predictions have even turned out to be correct). By a judicious combination of abstraction and formalism, mathematical psychologists have been able to develop a toolkit that allows anyone to derive theoretical predictions in completely novel paradigms. If it is indeed the case that psychology suffers from a kind of "theoretical amnesia" (Boorsbaum, 2013), perhaps the machinery of mathematical psychology can aid its memory. Perhaps fittingly, the words of Shepard (1987, p. 1323) seem an appropriate way to conclude

> *Undoubtably, psychological science has lagged by behind physical science by at least 300 years. Undoubtedly, too, prediction of behavior can never attain the precision for animate that it has for celestial bodies. Yet, psychology may not be inherently limited merely to the descriptive characterization of the behaviors of particular terrestrial species. Possibly, behind the diverse behaviors of humans and animals, as behind the various motions of planets and stars, we may discern the operation of universal laws*

References

Boorsbaum, D. (2013). Theoretical amnesia. *Center for Open Science.* Retrieved from http://osc.centerforopenscience.org/2013/11/20/theoretical-amnesia/

Cooper, R. P., & Guest, O. (2014). Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research, 27*, 42–49.

Ebbinghaus, H. (2013). Memory: A contribution to experimental psychology. *Annals of Neurosciences, 20*(4), 155.

Ekman, G. (1954). Dimensions of color vision. *The Journal of Psychology, 38*(2), 467–474.

Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology, 29*(2), 158–181.

Griffiths, T., Canini, K., Sanborn, A., & Navarro, D. (2007). Unifying rational models of categorization via the hierarchical dirichlet process.

Hayes, B. K., Banner, S., Forrester, S., & Navarro, D. J. (2019a). Selective sampling and inductive inference: Drawing inferences based on observed and missing evidence. *Cognitive Psychology, 113*, 101221.

Hayes, B. K., Navarro, D. J., Stephens, R. G., Ransom, K., & Dilevski, N. (2019b). The diversity effect in inductive reasoning depends on sampling assumptions. *Psychonomic Bulletin & Review, 26*(3), 1043–1050.

Hendrickson, A. T., Perfors, A., Navarro, D. J., & Ransom, K. (2019). Sample size, number of categories and sampling assumptions: Exploring some differences between categorization and generalization. *Cognitive Psychology*, *111*, 80–102.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*(2), 115–129.

Lawson, C. A., & Kalish, C. W. (2009). Sample selection and inductive generalization. *Memory & Cognition*, *37*(5), 596–607.

Lee, J. C., Lovibond, P. F., Hayes, B. K., & Navarro, D. J. (2019). Negative evidence and inductive reasoning in generalization of associative learning. *Journal of Experimental Psychology: General*, *148*(2), 289.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press on Demand.

Navarro, D. J. (2019). Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. *Computational Brain & Behavior*, *2*(1), 28–34.

Navarro, D. J., Dry, M. J., & Lee, M. D. (2012). Sampling assumptions in inductive generalization. *Cognitive Science*, *36*(2), 187–223.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39.

Ransom, K. J., Perfors, A., & Navarro, D. J. (2016). Leaping to conclusions: Why premise relevance affects argument strength. *Cognitive Science*, *40*(7), 1775–1796.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, *2*, 64–99.

Russell, S. (1988). Analogy by similarity. In *Analogical reasoning* (pp. 251–269). Springer.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shepard, R. N. (2004). How a cognitive psychologist came to seek universal laws. *Psychonomic Bulletin & Review*, *11*(1), 1–23.

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review*, *124*(4), 410.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640.

Voorspoels, W., Navarro, D. J., Perfors, A., Ransom, K., & Storms, G. (2015). How do people learn from negative evidence? Non-monotonic generalizations and sampling assumptions in inductive reasoning. *Cognitive Psychology*, *81*, 1–25.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Maas, H. L. van der, & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*, 632–638.