Journal of Mathematical Psychology 56 (2012) 281–285

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Notes and comment

Bayesian generalization with circular consequential regions

Thomas L. Griffiths *, Joseph L. Austerweil

Department of Psychology, University of California, Berkeley, United States

ARTICLE INFO

Article history:
Received 19 May 2012
Received in revised form
6 July 2012
Available online 26 July 2012

Keywords:
Generalization
Bayesian inference
Rational analysis

ABSTRACT

Generalization – deciding whether to extend a property from one stimulus to another stimulus – is a fundamental problem faced by cognitive agents in many different settings. Shepard (1987) provided a mathematical analysis of generalization in terms of Bayesian inference over the regions of psychological space that might correspond to a given property. He proved that in the unidimensional case, where regions are intervals of the real line, generalization will be a negatively accelerated function of the distance between stimuli, such as an exponential function. These results have been extended to rectangular consequential regions in multiple dimensions, but not for circular consequential regions, which play an important role in explaining generalization for stimuli that are not represented in terms of separable dimensions. We analyze Bayesian generalization with circular consequential regions, providing bounds on the generalization function and proving that this function is negatively accelerated.

Generalizing a property from one stimulus to another is a fundamental problem in cognitive science. The problem arises in many forms across many different domains, from higher-level cognition (e.g., concept learning, Tenenbaum (2000)) to linguistics (e.g., word learning, Xu and Tenenbaum (2007)) to perception (e.g., color categorization, Kay and McDaniel (1978)). The ability to generalize effectively is a hallmark of cognitive agents and seems to take a consistent form across domains and across species (Shepard, 1987). This consistency led Shepard (1987) to propose a "universal law" of generalization, arguing that the probability of generalizing a property decays exponentially as a function of the distance between two stimuli in psychological space. This argument was based on a mathematical analysis of generalization as Bayesian inference.

Shepard's (1987) analysis asserted that properties pick out regions in psychological space ("consequential regions"). Upon observing that a stimulus possesses a property, an agent makes an inference as to which consequential regions could correspond to that property. This is done by applying Bayes' rule, yielding a posterior distribution over regions. The probability of generalizing to a new stimulus is computed by summing over all consequential regions that contain both the old and the new stimulus, weighted by their posterior probability. Shepard gave analytical results for generalization along a single dimension, where consequential regions correspond to intervals of the real line, proving that generalization should be a negatively accelerated function of distance, such as an exponential. He also simulated results for generalization in two dimensions, examining how the pattern of generalization related to the choice of consequential regions. The resulting model explains generalization behavior as optimal statistical inference according to a probabilistic model – a rational analysis of generalization (Anderson, 1990; Chater & Oaksford, 1999) – and is one of the most important precursors of the recent surge of interest in Bayesian models of cognition, which include extensions of the Bayesian generalization framework beyond spatial representations (Navarro, Dry, & Lee, 2012; Tenenbaum & Griffiths, 2001).

One of the valuable insights yielded by Shepard's (1987) analysis was that different patterns of generalization could be captured by making different assumptions about consequential regions. People use two different kinds of metrics when forming generalizations about multi-dimensional stimuli: *separable* dimensions are associated with exponential decay in "city-block" distance or the L_1 metric, while *integral* dimensions are associated with exponential decay in Euclidean distance or the L_2 metric (Garner, 1974). These different metrics also have consequences beyond generalization behavior, influencing how people categorize objects varying along different dimensions (Handel & Imai, 1972) and whether people can selectively attend to each dimension (Garner & Felfoldy, 1970). Additionally, there is evidence that people can learn which metric they should use for generalization based on concept learning (Austerweil & Griffiths, 2010).

In the Bayesian generalization model, the difference between separable and integral dimensions emerges as the result of probabilistic inference with different kinds of consequential regions (Davidenko & Tenenbaum, 2001; Shepard, 1987, 1991).

* Correspondence to: University of California, Berkeley, Department of Psychology, 3210 Tolman Hall # 1650, Berkeley, CA 94720-1650, United States.
E-mail addresses: tom_griffiths@berkeley.edu (T.L. Griffiths), joseph.austerweil@gmail.com (J.L. Austerweil).

0022-2496/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2012.07.002

When consequential regions are aligned with the axes of the space, such as rectangles or ellipses that have their major axes parallel to the dimensions in which stimuli are expressed, a pattern of generalization similar to that seen for separable dimensions emerges. When consequential regions are indifferent to the axes of the space, such as circles or randomly-oriented rectangles or ellipses, a pattern of generalization similar to that seen with integral dimensions appears. Shepard (1987) noted: "For stimuli, like colors, that differ along dimensions that do not correspond to uniquely defined independent variables in the world, moreover, psychological space should have no preferred axes. The consequential region is then most reasonably assumed to be circular or, whatever other shapes may be assumed, to have all possible orientations in the space with equal probability" (p. 1322).

Despite the importance of considering different kinds of consequential regions in multidimensional spaces to Shepard's (1987) theory, the result that the generalization function should be negatively accelerated was only proved in the unidimensional case. Subsequent analyses have shown that negatively accelerated functions can be obtained with rectangular consequential regions (Myung & Shepard, 1996; Tenenbaum, 1999b,a) and generalized the argument to discrete representations (Austerweil & Griffiths, 2010; Chater & Vitanyi, 2003; Russell, 1986; Tenenbaum & Griffiths, 2001). However, the case of circular consequential regions – which are particularly important for representing integral dimensions, as noted above – has not been investigated in detail. In this article, we derive bounds and prove that the function produced by Bayesian generalization with multidimensional circular consequential regions is negatively accelerated, extending Shepard's original result to this multidimensional case.

The strategy behind our analysis is as follows. We begin by formulating the problem of generalization as Bayesian inference for an unknown consequential region. Next, we reparameterize the problem to allow us to simplify the probability of generalizing to a new stimulus to the integral of a simple function. Unfortunately the integral has no known closed form solution, leading us to attack it in two ways. First, we derive bounds on the integral that approximate the true solution. Second, we prove through analysis of the derivatives of the integral that the solution to the integral is convex and must be monotonically decaying in the Euclidean distance between the two stimuli.

1. Problem formulation

Assume that an observation $\mathbf{x}$ is drawn from a circular consequential region in $\mathbb{R}^2$. Then we have

$$p(\mathbf{x}|\mathbf{c}, s) = \begin{cases} \frac{1}{\pi s} & \|\mathbf{x}-\mathbf{c}\|^2 \leq s \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\mathbf{c}$ is the center of the consequential region, with s the square of its radius. We can then consider the set of all possible consequential regions from which the observation might have been drawn, which is here the set of all possible circles, and use Bayes' rule to calculate the probability of that consequential region given the observation of $\mathbf{x}$. Specifically, we have

$$p(h|\mathbf{x}) = \frac{p(\mathbf{x}|h)p(h)}{p(\mathbf{x})} \tag{2}$$

where h is some hypothetical consequential region, here consisting of a pair $\mathbf{c}, s$. To evaluate the denominator, we simply compute $\int_{h\in\mathcal{H}} p(\mathbf{x}|h)p(h)dh$, where $\mathcal{H}$ is the set of all hypotheses under consideration, here being all pairs $\mathbf{c}, s$. From this we can obtain the probability that some other point $\mathbf{y}$ is in the true consequential region from which $\mathbf{x}$ was drawn

$$\begin{aligned} p(\mathbf{y}\in C|\mathbf{x}) &= \int_{h\ni y, h\in\mathcal{H}} p(h|\mathbf{x})dh \\ &= \frac{\int_{h\ni y, h\in\mathcal{H}} p(\mathbf{x}|h)p(h)dh}{\int_{h\in\mathcal{H}} p(\mathbf{x}|h)p(h)dh} \end{aligned} \tag{3}$$

where C is the true consequential region. We focus on the numerator for now (the denominator will follow as a special case).

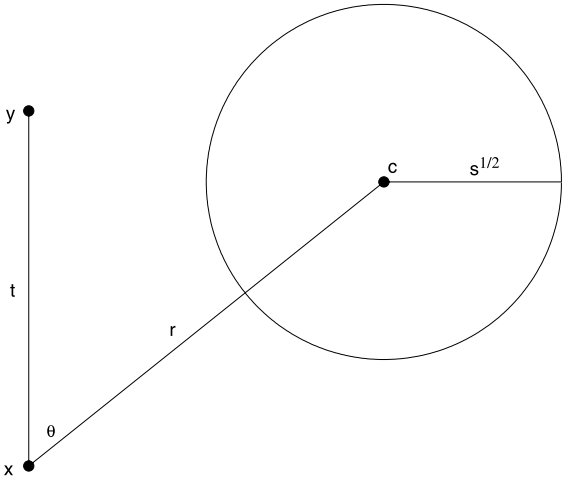

Fig. 1. Parameterization used to compute $P(\mathbf{y}\in C|\mathbf{x})$.

We can think about this problem in terms of the graphical representation shown in Fig. 1. Taking $\mathbf{x}$ as the origin, we can express the location of $\mathbf{c}$ in polar coordinates (r, θ), where θ is such that $\mathbf{y}$ is located at $(t, 0)$. Let r be the distance between $\mathbf{x}$ and $\mathbf{c}$, $\|\mathbf{x}-\mathbf{c}\|$ in Eq. (1), and t be the distance between $\mathbf{x}$ and $\mathbf{y}$. This is a nice parameterization for the problem, because it allows us to integrate over all circles containing both $\mathbf{x}$ and $\mathbf{y}$ (beginning with the smallest circle containing both of them). We can divide the plane into four quadrants, with one axis passing through $\mathbf{x}$ and $\mathbf{y}$, and a perpendicular axis that crosses halfway between $\mathbf{x}$ and $\mathbf{y}$. Due to the resulting symmetries between the circles containing both $\mathbf{x}$ and $\mathbf{y}$ in these four quadrants, we need only consider one of the quadrants. In Fig. 1, $\mathbf{c}$ is located above the midpoint of $t/2$, $\mathbf{y}$ will always be in h if $\mathbf{x}$ is in h. Thus we need only consider those circles for which $s \geq r^2$, where s is the variable of integration and represents the area of the circular consequential region. The resulting generalization gradients will be only a function of t, the distance between $\mathbf{x}$ and $\mathbf{y}$, and the denominator of Eq. (3) follows from the case where $t = 0$, as with generalization in one dimension.

For reasons that will become clear in a moment, we use $u = r^2$ instead of r directly. This choice of parameterization allows us to write

$$\begin{aligned} &\int_{h\ni y, h\in\mathcal{H}} p(\mathbf{x}|h)p(h)dh \\ &\quad\propto \int_0^{\pi/2}\int_{u_0}^{\infty}\int_u^{\infty} p(\mathbf{x}|\theta, u, s)p(\theta, u, s)\, ds\, du\, d\theta \end{aligned} \tag{4}$$

where u_0 is the minimum value of u to place $\mathbf{c}$ in the desired quadrant, which will be a function of θ. The first two integrals are over the possible centers of circles (that place $\mathbf{c}$ in the desired quadrant) and the third integral is over the possible circle sizes (ranging from the smallest circle including both $\mathbf{x}$ and $\mathbf{y}$). This is equivalent to integrating over the entire domain because $p(\mathbf{x}|\theta, u, s) = 0$ for the circles that do not contain $\mathbf{x}$ and the integral is constrained to include $\mathbf{y}$.

The expression in the above equation requires us to specify a likelihood, $p(\mathbf{x}|h) = p(\mathbf{x}|\theta, u, s)$ and a prior distribution, $p(h) = p(\theta, r, s)$. The likelihood is uniform over all points in the circle specified by h, which is defined by Eq. (1). By taking s to be the squared radius, we have

$$p(\mathbf{x}|h) = p(\mathbf{x}|\theta, u, s) = \begin{cases} \frac{1}{\pi s} & \mathbf{x} \in h \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

This implements the "size principle" that plays an important role in Bayesian generalization (Tenenbaum, 1999a; Tenenbaum & Griffiths, 2001).

For the prior, we assume a uniform distribution over the location of the center of the circles and an Erlang distribution (with parameters $k = 2$ and $\lambda = \pi$) over their area.[1] This is similar to the maximum entropy "expected-size" prior that captured human judgments well for multidimensional axis aligned concepts (Tenenbaum, 1999b) and takes the same form as the prior that yielded an exponential generalization function in Shepard (1987). Hence we have

$$p(\theta) = \frac{1}{2\pi} \tag{6}$$

$$p(u) \propto 1 \tag{7}$$

$$p(s|k = 2, \lambda = \pi) = \pi^2 s e^{-\pi s} \tag{8}$$

$$p(h) = p(\theta, u, s) \propto \pi s e^{-\pi s} \tag{9}$$

which is an improper prior—the integral over all $\mathcal{H}$ diverges when $\mathcal{H}$ includes all circles in the plane. The use of this improper prior motivates the choice of $u = r^2$ rather than r in the above parameterization.

Another justification for the prior is provided by thinking in terms of a generative process that creates circles by generating a circle location and size independently. To do this, we want a uniform distribution over the locations of the circles. In polar coordinates, this means that we are going to be doing the equivalent of choosing points from a very large circle. For a circle of radius R, a point in that circle is chosen with probability $p(\mathbf{s}) = \frac{d\mathbf{s}}{\pi R^2}$. Transforming to polar coordinates, $p(\mathbf{s}) = p(\theta, r) r\, d\theta\, dr = p(\theta) p(r) r\, d\theta\, dr = \frac{1}{2\pi} \frac{2r}{R^2}\, d\theta\, dr$. If we now transform to coordinates (θ, u), where $u = r^2$, we have $du = 2r\, dr$, so $dr = \frac{du}{2r}$. This means that $p(\theta) = \frac{1}{2\pi} \Rightarrow \theta \sim U(0, 2\pi)$, and $p(u) = \frac{2r}{R^2} \frac{du}{2r} = \frac{du}{R^2} \Rightarrow u \sim U(0, R^2)$. We are thus choosing points from a simple uniform distribution for both parameters, which allows us to define the improper prior given above and use it in exactly the same way as the unidimensional proof by Shepard (1987).

Our desired integral now becomes

$$\begin{aligned} \int_{h \ni y, h \in \mathcal{H}} p(\mathbf{x}|h) p(h) dh &\propto \int_0^{\pi/2} \int_{u_0}^{\infty} \int_u^{\infty} \frac{1}{\pi s} \pi s e^{-\pi s}\, ds\, du\, d\theta \\ &\propto \int_0^{\pi/2} e^{-\pi u_0} d\theta. \end{aligned} \tag{10}$$

We can then solve for u_0, the minimum squared distance from $\mathbf{x}$ required to place $\mathbf{c}$ in the quadrant where it is guaranteed to be closer to $\mathbf{y}$ than to $\mathbf{x}$. This means that for a given value of θ, we have to find the length of the squared hypotenuse u in a right triangle with $t/2$ as the side adjacent to θ. By the definition of the cosine of an angle as the ratio of the adjacent side to the hypotenuse, this is just $\frac{t^2}{4 \cos^2 \theta}$. Making the substitution $v = \frac{1}{\cos^2 \theta}$, we obtain

$$d\theta = \frac{1}{2} \frac{\cos \theta}{\sin \theta} \cos^2 \theta\, dv \tag{11}$$

$$\begin{aligned} \int_{h \ni y, h \in \mathcal{H}} p(\mathbf{x}|h) p(h) dh &\propto \int_0^{\pi/2} e^{-u_0} d\theta \\ &\propto \int_1^{\infty} \frac{\cos \theta}{\sin \theta} \cos^2 \theta \exp\left\{-\frac{\pi}{4} t^2 v\right\} dv \\ &= \int_1^{\infty} \frac{\cos \theta}{\sqrt{1 - \cos^2 \theta}} \frac{1}{v} \exp\left\{-\frac{\pi}{4} t^2 v\right\} dv \\ &\propto \int_1^{\infty} \frac{\exp\left\{-\frac{\pi}{4} t^2 v\right\}}{v\sqrt{v - 1}} dv \end{aligned} \tag{12}$$

which evaluates to π for $t = 0$ (as $\int_1^{\infty} \frac{dv}{v\sqrt{v-1}} = \pi$), giving us the denominator of Eq. (3). This gives us the final expression

$$p(\mathbf{y} \in C|\mathbf{x}) = \frac{1}{\pi} \int_1^{\infty} \frac{e^{-v\pi t^2/4}}{v\sqrt{v - 1}} dv, \tag{13}$$

which unfortunately does not have an elementary solution.

2. Approximations to the generalization function

Since the generalization function is not analytically tractable, we attempt to get a clearer picture of its properties by obtaining bounds on the function. We do this in two ways—defining simple fixed bounds, and deriving parameterized variational bounds.

2.1. Simple bounds

We can obtain simple upper and lower bounds by bounding the integrand in Eq. (13). As an upper bound, we can bound the integral by noting that the domain of integration restricts $v \geq 1$ (and thus, removing v from the exponent can only increase the result of the integral) in the following manner:

$$\begin{aligned} \frac{1}{\pi} \int_1^{\infty} \frac{e^{-v\pi t^2/4}}{v\sqrt{v - 1}} dv &\leq \frac{1}{\pi} \int_1^{\infty} \frac{e^{-\pi t^2/4}}{v\sqrt{v - 1}} dv \\ &= \frac{1}{\pi} e^{-\pi t^2/4} \int_1^{\infty} \frac{dv}{v\sqrt{v - 1}} \\ &= e^{-\pi t^2/4}. \end{aligned} \tag{14}$$

This gives an upper bound on the generalization function $p(\mathbf{y} \in C|\mathbf{x}) \leq \frac{1}{\pi} e^{-\pi t^2/4}$. As a lower bound, we can integrate $\frac{e^{-v\pi t^2/4}}{v^{3/2}}$, which gives a lower bound on the generalization function $p(\mathbf{y} \in C|\mathbf{x}) \geq \frac{2}{\pi} e^{-\pi t^2/4} - t(1 - \text{erf}(t\sqrt{\pi}/2))$, where erf is the error function. These bounds are plotted with dotted lines in Fig. 2 and are similar in their tightness to those found by Tenenbaum (1999a) for Bayesian generalization with axis-aligned rectangular consequential regions.

2.2. Variational bounds

The simple bounds are unfortunately very poor, and give little idea of the shape of the generalization function. We can obtain better results with variational upper and lower bounds, based upon a decomposition of the integral. For a lower bound, we can introduce a variable v_0, and use the fact that $\frac{e^{-v_0\pi t^2/4}}{v\sqrt{v-1}} < \frac{e^{-v\pi t^2/4}}{v\sqrt{v-1}}$ for $v < v_0$. This means that the expression

$$\frac{1}{\pi} \left[\int_1^{v_0} \frac{e^{-v_0\pi t^2/4}}{v\sqrt{v - 1}} dv + \int_{v_0}^{\infty} \frac{e^{-v\pi t^2/4}}{v^{3/2}} dv \right] \tag{15}$$

[1] The Erlang distribution is a special case of the Gamma distribution, where the scale parameter is constrained to be integer valued.

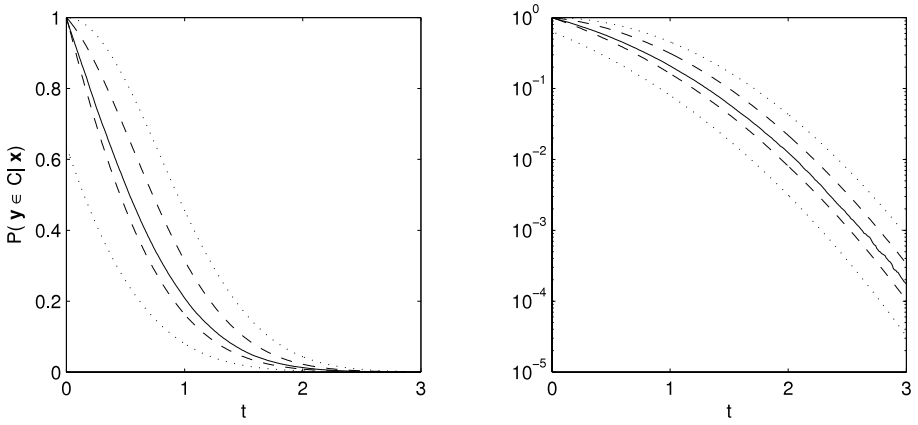

Fig. 2. The generalization function, shown on a linear scale on the left and a logarithmic scale on the right. Dotted lines show simple bounds, dashed lines show variational bounds, and the solid line is an approximate generalization function computed via 10 million Monte Carlo samples from the prior.

will be a lower bound on Eq. (13) for $v_0 \geq 1$. This gives

$$p(\mathbf{y} \in C|\mathbf{x}) \geq \frac{2}{\pi} e^{-v_0 \pi t^2/4} \left(\operatorname{atan}\left(\sqrt{v_0 - 1}\right) + \frac{1}{\sqrt{v_0}} \right) - t(1 - \operatorname{erf}(t\sqrt{\pi v_0}/2)). \tag{16}$$

Taking the maximum at each point of the functions resulting from varying the settings of the variational parameter v_0 gives the lower bound shown with a dashed line in Fig. 2. Applying a similar procedure to the upper bound an initial portion of the function and the tail separately, we can take

$$\frac{1}{\pi}\left[\int_1^{v_0} \frac{e^{-\pi t^2/4}}{v\sqrt{v-1}}\, dv + \int_{v_0}^{\infty} \frac{e^{-v_0 \pi t^2/4}}{v\sqrt{v-1}}\, dv\right] \tag{17}$$

to obtain the variational upper bound

$$p(\mathbf{y} \in C|\mathbf{x}) \leq \frac{2}{\pi} e^{-\pi t^2/4 \operatorname{atan}(\sqrt{v_0 - 1}+1)} + e^{-v_0 \pi t^2/4} \left(1 - \frac{2}{\pi} \operatorname{atan}\left(\sqrt{v_0 - 1}\right)\right) \tag{18}$$

with $v_0 \geq 1$. Taking the lowest value of this function across a range of settings of v_0 gives the upper bound shown with a dashed line in Fig. 2.

3. Monotonicity and concavity through derivatives

Finally, we can get an idea of the form of the function $g(t) = P(\mathbf{y} \in C|\mathbf{x})$ by computing its derivatives. We can exchange the integral and derivative to give $\frac{d}{dt}\int_1^{\infty} f(v,t)dv = \int_1^{\infty} \frac{\partial}{\partial t} f(v,t)dv$ if $\frac{\partial}{\partial t} f(v,t)$ is continuous on $[1,\infty)$ and $\int_1^{\infty} f(v,t)dv$ has the property of dominated convergence: there is some function $\phi(v)$ such that for all $v \geq 1$, $|f(v,t)| \leq \phi(v)$ and $\int_1^{\infty} \phi(v)dv$ converges. In the case of $g(t) = \frac{1}{\pi}\int_1^{\infty} \frac{e^{-v\pi t^2/4}}{v\sqrt{v-1}}\, dv$, we have

$$\frac{\partial}{\partial t} \frac{1}{\pi} \frac{e^{-v\pi t^2/4}}{v\sqrt{v-1}} = -\frac{t}{2} \frac{e^{-v\pi t^2/4}}{\sqrt{v-1}} \tag{19}$$

which is continuous on $[1,\infty)$ (over t, which is required to interchange derivatives and integrals),[2] and the simple upper bound derived above uses a function $\phi(v)$ that can be used to establish dominated convergence.

Having shown that we can differentiate under the integral, we now obtain

$$g'(t) = -\frac{t}{2}\int_1^{\infty} \frac{e^{-v\pi t^2/4}}{\sqrt{v-1}}\, dv. \tag{20}$$

We can evaluate this integral by substituting $w = v - 1$,

$$g'(t) = -\frac{t}{2}\int_0^{\infty} dw \frac{e^{-\frac{(w+1)}{4}\pi t^2}}{\sqrt{w}} = -\frac{t}{2} e^{-\pi t^2/4} \int_0^{\infty} dw \frac{e^{-\frac{w}{4}\pi t^2}}{\sqrt{w}}. \tag{21}$$

This integral is the normalizing constant for a Gamma integral with parameters $1/2$ and $\pi t^2/2$,

$$g'(t) = \frac{-\frac{t}{2}\Gamma(1/2) e^{-t^2\pi/4}}{(\pi t^2/2)^{1/2}} = -\frac{\sqrt{2}}{2} e^{-\pi t^2/4} \tag{22}$$

where we have used the identity that $\Gamma(1/2) = \sqrt{2}$. Thus, the derivative of $g(t)$ is always nonpositive. It follows directly that

$$g''(t) = \frac{\pi t\sqrt{2}}{4} e^{-\pi t^2/4} \tag{23}$$

so $g(t)$ has inflection points at $t = 0$ and $t = \infty$ but not at intermediate values, as shown by Shepard (1987) for the generalization function in one dimension. Additionally, $g'(t) < 0, \forall t \in (0,\infty)$ and $g''(t) > 0, \forall t \in (0,\infty)$ implies that $g(t)$ is strictly convex on $(0,\infty)$ (Berkovitz, 2002).

4. Conclusions

In this article, we analyzed the nature of the generalization function of the Bayesian generalization model with a hypothesis space of circular consequential regions. Though the generalization function did not have an elementary solution, we reached a form of the generalization function involving a single integral that allowed us to bound the generalization function using simple and variational upper and lower bounds. Finally, using derivatives we found that the generalization function decays monotonically with increasing distance between the stimuli and that it is convex.

[2] On the other hand, Eq. (19) is continuous in v on $(1,\infty)$, but not $[1,\infty)$. Regardless, it satisfies the necessary properties for interchanging derivatives and integrals.

Consequently, generalization will be a negatively accelerated function of distance in psychological space, extending Shepard's (1987) result for unidimensional consequential regions. Taken as a whole, the series of results bolsters our understanding of a fundamental problem in cognition, yields analytic approximations, and provides a formal basis for cognitive models involving generalization using integral stimuli.

Acknowledgments

We thank Michael Lee, Dan Navarro, Josh Tenenbaum, Ewart Thomas, and an anonymous reviewer for feedback on a previous draft of this manuscript. This work was supported by grant number FA-9550-10-1-0232 from the Air Force Office of Scientific Research.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
Austerweil, J. L., & Griffiths, T. L. (2010). Learning hypothesis spaces and dimensions through concept learning. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 73–78). Austin, TX: Cognitive Science Society.
Berkovitz, L. D. (2002). *Convexity and optimization in* $\mathbb{R}^n$. New York, NY: John Wiley & Sons.
Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Science*, *3*, 57–65.
Chater, N., & Vitanyi, P. (2003). The generalized universal law of generalization. *Journal of Mathematical Psychology*, *47*, 346–369.
Davidenko, N., & Tenenbaum, J.B. (2001). Concept generalization in separable and integral stimulus spaces. In *Proceedings of the 23rd annual conference of the cognitive science society*. Mahwah, NJ.
Garner, W. R. (1974). *The processing of information and structure*. Maryland: Erlbaum.
Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225–241.
Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception & Psychophysics*, *12*, 108–116.
Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, *54*, 610–646.
Myung, I. J., & Shepard, R. N. (1996). Maximum entropy inference and stimulus generalization. *Journal of Mathematical Psychology*, *40*, 342–347.
Navarro, D. J., Dry, M. K., & Lee, M. D. (2012). Sampling assumptions in inductive generalization. *Cognitive Science*, *36*, 187–223.
Russell, S. J. (1986). A quantitative analysis of analogy by similarity. In *Proceedings of the national conference on artificial intelligence* (pp. 284–288). Philadelphia, PA: AAAI.
Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.
Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: from an early convergence of evidence to a proposed theoretical basis. In *The perception of structure: essays in honor of Wendell R. Garner* (pp. 53–71). Washington, DC: American Psychological Association.
Tenenbaum, J. B. (2000). Rules and similarity in concept learning. In S. A. Solla, T. K. Leen, & K.-R. Muller (Eds.), *Advances in neural information processing systems 12* (pp. 59–65). Cambridge, MA: MIT Press.
Tenenbaum, J. B. (1999b). Bayesian modeling of human concept learning. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems 11* (pp. 59–65). Cambridge, MA: MIT Press.
Tenenbaum, J.B. (1999a). *A Bayesian framework for concept learning*. Ph.D. Thesis Massachussetts Institute of Technology. Cambridge, MA.
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–641.
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272.