

THE PROSPECTS FOR MEASUREMENT IN INFINITE DIMENSIONAL PSYCHOLOGICAL SPACES

Subtitle: Modern Notions for Geometric Person Measurements in Finite and
Infinite Dimensional Spaces

James T. Townsend, Devin Burns and Lei Pei, Indiana University

I: INTRODUCTION

This chapter is on certain spatial types of measurement of people, as carried out in the social and biological sciences. The global issue for us is perceptual classification and we specialize on spatial, for instance, geometric or topological, aspects. We will offer more detail below, but for now, we simply note that classification is the assignment of an object, typically presented to one or more of our sensory modalities, to one of some number (perhaps infinite!) of classes or designated sets, often names. Because in general terms, this is no more nor less than a function or mapping of a set of things (stimuli, memories, etc.) into another (the names or classifications), makes this concept extremely broad. Thus, classification covers many specific activities of psychological interest. Some important ones are: 1. A famous form of classification is that of "Categorization", where a bunch of objects is partitioned so that each object is separately assigned to a single, (typically) unique category. There are many kinds of categorization but one of special importance is usually given its own title: 2. "Identification", where each category has exactly one member (this may be rare in the real life). We can perform an experiment where each person has exactly one name and each name is given to exactly one person, but of course, in the real world, the name of "John" is

given to thousands of people. In most of the interesting psychological situations, the classification is made difficult by any number of factors. This could be because learning of the assignments is incomplete, such as in a recognition memory experiment. In this case the presented stimulus may be very easy to perceive, but the learned patterns may have become quite noisy, making the recognition task challenging. Other situations include some kind of added noise, randomness, or even very brief exposure periods 3. Yes-No Signal Detection is the specification by an observer of whether a signal has been presented or not. There are only two responses; "YES" vs. "NO" could be used either with faulty memory (the celebrated "Old" vs. "New" memory-recognition experiment) or with simply "hard-to-detect" physical stimuli. 4. Psychological Scaling: Typically here, the observer give ratings on one or more aspects of the stimuli. For instance, one might simply be asked to report a number that seems to represent the psychological magnitude of an aspect of a stimulus, such as how happy a particular face looks (see below for more on this). Or, as is often the case in multidimensional scaling (explained a bit later), the observer might be asked about the psychological similarity of two or more objects. An old anthology, but one still held in high esteem, with several chapters that are relevant to our present enterprise is Volume I of the Handbook of Mathematical Psychology (Luce et al., 1963). The philosophy of science of the present authors regarding measurement and modeling of people and an expansion of some of themes, can be located in Townsend & Thomas (1993).

All the chapters in this volume on "person measurement" will present and discuss valuable types of measurement of people, a number of them involving classi-

fication, usually pertaining to a finite number of aspects or dimensions. Section II below starts by discussing the foundational aspects of measurement on people and considers such finitistic situations. Section III concerns multidimensional scaling, and also focuses on a finite number of psychological dimensions. It is meant to not only set the stage for what is to come, but also provide an elementary tutelage on the subject and to guide the prospective user of these tools to some of the classical literature. That being said, it is our belief that almost none of the interesting aspects of perception refer at their most basic level to finite spaces. Rather, we feel that infinite dimensional spaces are required for a suitably theoretical milieu for human perception and beyond that, to cognition and action as well! For clarity we shall emphasize solely perception in this chapter. The topic of finite vs. infinite dimensional spaces will be taken up in Section IV. As consequence of this stance, an important topic is the mathematical issues that might arise when we inquire about the ability of people to extract finite dimensional information, even the value on a single dimension, from objects that by their very nature reside in infinite dimensional spaces. This question will be addressed in Section V.

A huge literature exists in mathematics (mostly for physics and engineering) on infinite dimensional spaces with particular relevance to functions in spaces that possess coordinates at right angles to one another. It is called "functional analysis". Such spaces typically are inherently flat, without curvature. A revolutionary movement took place in the late nineteenth and early twentieth centuries concerned with spaces which could be curved, even with spaces whose curvature could change from point to point. We refer to Riemannian manifolds. Section

VI introduces the concept of manifolds as gently as possible. This kind of space proved to be exactly the appropriate setting for Einstein's theory of general relativity. We go out on a limb and suggest that manifolds may be important for human perception and cognition and action as well. Section VII takes up the potential extension of probability theory and stochastic processes to infinite dimensional spaces: Everything we know about people, from the level of a single neuron (and lower) to the actions of the people of an entire nation are probabilistic. Hence, any theory or model whether finite or infinite must sooner or later broach this topic. There is a knowledge base for probabilities in infinite dimensional spaces that goes back to the early part of the twentieth century, with tremendous developments occurring mid-century. This knowledge applies immediately to the types of spaces found in functional analysis, mentioned just above. There is some work going on, in addition, developing laws of probability for finite dimensional manifolds and really cutting edge work in probability and stochastic processes on infinite dimensional manifolds. Section VIII treats the very special, and extraordinarily important case of models of classification based on functional analysis, dynamic systems theory, differential equations, and stochastic processes. We then delve more deeply into the mathematics and take a look at applying these techniques to the field of classification in section IX. The chapter concludes with section X, which is a brief review of where we've been and why, and an attempt to set these new methods in their proper context.

The early parts of this chapter leave out the mathematics. This is not only because of somewhat limited space but also because there are scores and sometimes

hundreds of articles and books on the topic. As we move along, we delve into material that is likely new to many investigators and students devoted to measurement on people. Although our greater use of mathematics as we progress may make the going a little tougher for some readers, it also affords a more rigorous and well-defined landscape for those willing to make the journey. We will accompany the verbal and quantitative developments with illustrations to aid in the readers' interpretations and intuitions. In this way, we hope that everyone will gain something from our presentation.

II: FOUNDATIONAL MEASUREMENT

A central characteristic of measurement in these regions, as opposed to vast regions of physics, is the absence of the opportunity to gather data based on strong measurement scales (e.g., Roberts et al., 1969). This single characteristic has been the subject of hundreds of papers and books, with roots in first, from the physical science point of view Campbell's famous book in which he develops a theory of fundamental measurement as a mapping between numerical systems and those things we wish to measure. In his theory, for something to be measurable, those things must have an ordering and the combination of those things must follow the same properties as the addition of numbers, which is often not satisfied by the things we wish to measure in psychology. A fundamental measurement system as defined by Campbell implies a ratio scale (Suppes & Zinnes, 1963). One of the earliest proposals that the scales typically employed by physics, that is "ratio scales" or "extensive measurement" might not be the only possibilities for rigorous application of numbers in the social or biological sciences was put forth by

Stevens (1951). Later, this idea was picked up by mathematical psychologists, mathematicians and logicians with the ensuing development of an axiomatic basis not only for extensive measurement but also the weaker (and in order of decreasing strength), absolute scale, ratio scale (extensive measurement), interval scale, ordinal scale and so-called nominal scale. The only other scale of note is the absolute scale, which is the strongest of all (e.g., for more on these topics, see the early work and citations in Suppes & Zinnes, 1963; Krantz et al., 1971; Pfanzagl, 1968). A seminal concept in the theory is that numbers considered as measurements should reflect the regularities of the natural phenomena that are being measured, while retaining properties of the number system (perhaps including arithmetic conditions and the like). Hence, for some purposes, an interval scale, like Celsius or Fahrenheit will suffice to study particular properties of a phenomenon. However, for work, say in thermodynamics, it may well be that the Kelvin scale, a form of , may be required. A fascinating consequent duality in the theory is that the stronger the scale, the fewer mathematical operations can be carried out on the observed measurements without distortion to the relationships between the numbers and the phenomena. An example of a permitted transformation on a ratio scale, like mass, is multiplication by a positive constant (e.g., changing pounds to kilograms). This leaves a ratio of any two measurements invariant—it is a "meaningful operation". However, calculating the ratio of yesterday's temperature with today's temperature, measured in Celsius, will not be the same as when the units are altered to Fahrenheit—meaningfulness is lost. In this interval scale we need to instead take a ratio of differences to retain meaningful results. To tersely complete

the picture, absolute scales allow no transformation whatsoever and they possess no units. Counting is perhaps the prototypical example of an absolute scale—one doesn't change a "3" when one counts 3 candy bars or anything else. Of course, a change of the name of "three" from one language to another is okay! The next strongest scale is afforded the name "ordinal scale". Ordinal measurement permits any monotonic (order-preserving) transformation to be made on the measurement numbers, but only order can be established: comparisons such as ratios or differences are meaningless. It has been debated whether or not the social or biological sciences may be forever confined to ordinal measurement. For now, we shall be mainly investigating aspects of measurement that do not touch on this axiomatic theory but also a few that do. Obviously, even a modest glimpse at the entire edifice (not to mention the debates) would take several hundred pages. One of the short-comings of the approach has been the absence of a generally accepted error theory, which would permit the erection of a statistical theory of inference within the approach. We should also mention that many statisticians, psychometricians and others have either argued against the idea that measurement scales are needed (e.g., Lord & Novick, 1968; Townsend & Ashby, 1984) or simply ignored it. Nonetheless, from the authors' point of view, it is important to keep these considerations in mind.¹

III: A BRIEF INTRODUCTION TO MULTIDIMENSIONAL SCALING

A popular method of inferring geometrical structure from data in psychology (and related disciplines) is called Multidimensional Scaling, or MDS. MDS is a set of mathematical techniques that can be used to reveal the underlying structure

in data. These methods provide a spatial configuration of the data points, which allows a researcher to interpret patterns and regularities relatively easily. There are numerous (but not quite infinite!) variations of these techniques, and here we will only touch on some of the most important, general, and famous methods.

To be able to represent psychological data in a geometrical form, we first need to establish a notion of "distance". There are many different ways of doing this, and careers have been spent debating which is the most appropriate. One of the most straightforward methods is to conduct a similarity experiment in which subjects rate how dissimilar different stimuli are from each other. The distance should then be a monotonic transformation of this dissimilarity, so that the more similar objects are the less distance there is between them. After establishing a matrix of all the pairwise distances between stimuli, computer programs can be used to infer where each stimulus is located in a space in relation to the others. Most modern statistical computing packages, such as SPSS, will contain a program (ALSCAL is a popular example) which will display these results for you.

But what results does it display exactly? This brings us to another topic of debate in MDS, that of dimension (The concept of dimension will be discussed in more detail below). When the computer infers the positions of the data points in space, we have to tell it which space to put them in. Euclidian space is the most commonly used and simplest, but there is no reason why the data should necessarily be Euclidean, and various spaces and metrics have been imposed for different domain applications. Sticking with Euclidean space for now, it must be decided by the investigator how many dimensions to use. When fitting the points,

MDS programs attempt to minimize the "stress" in the system, which roughly means how well the extrapolated geometric distances conform to the distances derived from the dissimilarity data. A high stress value means that the model is not fitting the data. Clearly, a model in a higher dimensional space will always have less stress than a more constrained model. We can easily see that lower dimensional models are included as subspaces of higher dimensional models, so this must always be true.

So how does one know how many dimensions to use? Theoretical expectations can be brought to bear here. If your stimuli vary along only, say, n dimensions, then an n -D model may be most appropriate (ignoring more complicated instances of interaction for now). An exception could occur if the brain is unable to separate the n dimensions. In that case, a subspace consisting of $k < n$ dimensions will suffice. In no case would we expect the estimated dimension to be $k > n$, although an inappropriate metric might make it appear to be so e.g., suppose the psychological space corresponding to two physical dimensions is a 2-D hemisphere embedded in 3-D coordinate system. Then, the inappropriate Euclidian metric could make it seem that three dimensions is better than two. If one has the correct metric in hand, the stress of the two dimensional model should be almost as small as that for three or greater dimensional models. This subjective comparison between the stresses of different models can be used as the deciding factor even in the lack of theoretical expectations. In this uninformed case, the experimenter should pick the model such that all subsequent models of greater dimensionality yield only a small reduction in stress relative to the previous reductions.

If all of psychological space were nicely Euclidean with separable dimensions, then all of this would be fine and we could construct high fidelity maps for every data set. As I'm sure you've guessed, this is sadly (but more interestingly!) not the case. What if psychological dissimilarity data are not directly associated with a metric? For instance, it could be that decisional biases intrude that violate properties associated with any metric (e.g., Townsend, 1971). In such a case, it is sometimes possible to extricate the true dissimilarity apart from the decisional influences. In any event, we need to know what the properties of a valid distance measure are. The three axioms that any metric must satisfy are:

Minimality: $d(a,b) \geq d(a,a) = 0$: No pair of objects are more similar to each other than any object is to itself, and the latter distance is 0.

Symmetry: $d(a,b) = d(b,a)$: Nancy's face is as similar to Joan's as Joan's face is to Nancy's.

The triangle inequality: $d(a,c) \leq d(a,b) + d(b,c)$: The political distance between Obama's world-view and George W. Bush's is less than or equal to the distance between Obama's and Franklin D. Roosevelt's plus that between Roosevelt's and Bush's.

It could be that one or more of the axioms is violated due to perturbation by a decisional bias, or it could happen in a more fundamental, systematic sense. The mathematical psychologist Amos Tversky pointed out that all three of these axioms can be routinely violated in basic psychological experiments (?). Minimality is violated whenever stimuli differ in their self-similarity. Symmetry is often violated in similarity data as well, especially if one stimuli is seen as more

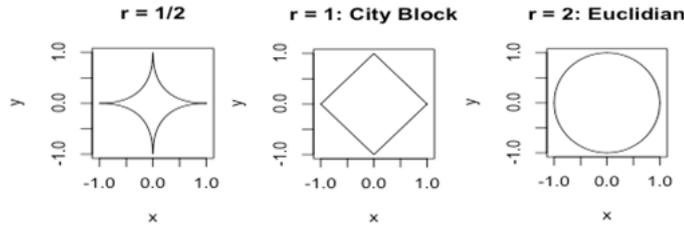


Figure 1: Each graph shows all points with a unit distance from the origin

broad or general than the other, perhaps including the latter as a subclass. For example, the word "Poodle" would be seen as more similar to the word "dog" than the reverse. The triangle inequality is harder to refute, because it is a quantitative statement, and similarity data is inherently ordinal. It has been shown that trivial manipulations of data can produce satisfaction of the first two assumptions, so the triangle equality often plays a major role in testing for the presence of a metric. Although there have been more rigorous expositions of the subject, Tversky gives a quick intuitive argument for why we shouldn't believe that the triangle equality will necessarily hold in all cases. Assume that the perceptual distance from Cuba to Russia should be small for political reasons (remember, 1977!), and the perceptual distance from Cuba to Jamaica should be small for geographic reasons. The triangle inequality would then force the distance from Jamaica to Russia to also be fairly small, which we would not expect.

One reaction to these arguments has been to impose a different metric on the space. Instead of using the standard Euclidian metric to compute distances, many other functions have been used. A common family of metrics takes the form:

$$d_{ij} = \sum_{m=1}^n |(x_{im} - x_{jm})|^r \Big|^{1/r}. \text{ These are commonly called power metrics. When } r = 2$$

we have the standard Euclidian metric, and $r=1$ is what is known as the city block metric. In this metric distances are computed as the sum of the projected distances in each orthogonal dimension (like in a city, when you can only travel in orthogonal dimensions, instead of "as the crow flies"). In this metric, the triangle inequality becomes an equality. If we consider values of $r < 1$, the triangle inequality is now reversed. This can be interpreted to saying that traveling along one dimension is "faster" than traveling along a diagonal path. These metrics can be better understood if we consider graphs of unit distance, shown in figure 1. In these graphs every point corresponds to an equal distance from the origin.

Tvesky offered up a different approach when he rigorously developed the "feature contrast model". In this model, the similarity between two objects is a function of the features that the objects have in common, minus the features that differ between them. The feature contrast model captures the intuitive idea that identical shared features decrease psychological dissimilarity between two objects, a property unattainable with metric-based differences. This relatively simple model is sufficient to account for each of the previously mentioned violations of the metric axioms.

One of the primary difficulties involved in modeling psychological data is that the perceptual phenomena that we seek to describe are confounded by the decisional processes inherent in any experimental situation. The classic theory designed to tease these aspects apart is known as Signal Detection Theory (SDT). This theory was established in 1966 by John Swets and David Green, building on earlier work done by radar researchers. In this methodology, participants are

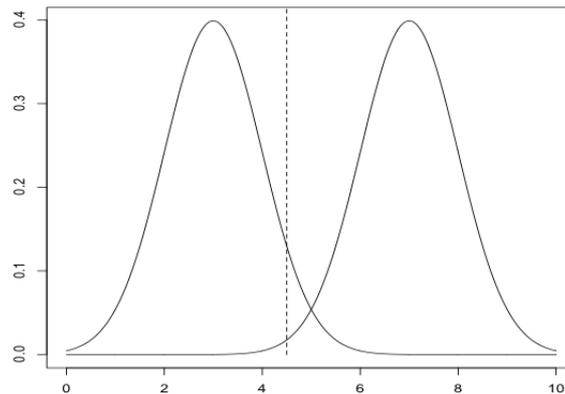


Figure 2: A typical SDT graph of signal and noise densities. The dotted line is the decision criterion

asked to discern "signal" trials from "noise" trials. The experimenter can then find values for the participant's perceptual discriminability and their decision criterion. These variables can be independently manipulated to study various decisional or perceptual qualities. In SDT, stimuli are considered to be perceived probabilistically. Instead of each stimulus corresponding to a single, fixed point in some space, as in MDS models, stimuli have corresponding probability density functions.

A typical example is shown in figure 2. The left curve corresponds to a noise trial, and the right to a signal trial. The vertical dotted line represents a decision criterion, where a participant will change from calling everything on the left "noise" to calling everything on the right a "signal". We can see that the participant will be incorrect for the stimuli belonging to the noise distribution that fall on the right of the criterion (called false alarms) and also for the stimuli from the

signal distribution that fall to the left of the criterion (called misses). Shifting the criterion left or right (which can be achieved by altering the experimental instructions) will result in a trade off between these two kinds of wrong answers. The only way to increase the total number of correct answers is to increase the distance between the means of the two distributions, which is referred to as d' . This theory offers a natural explanation for the confusions between stimuli while also elucidating the differences between perceptual and decisional effects, which are measured with d' and the criterion, respectively.

A limitation of the SDT method as typically employed is that stimuli are only allowed to vary along a single dimension, making the methodology applicable to a paucity of psychological experiments. In 1986, Gregory Ashby and James Townsend developed a multidimensional extension of SDT that they call General Recognition Theory (GRT). In GRT, the probability density functions associated with stimuli are multidimensional, so stimuli can vary along as many dimensions as desired. Because of the difficulty involved in making four or more dimensional graphs, let us consider the case where stimuli vary along just two dimensions. Figure 3a shows the probability densities of two stimuli that vary along dimensions x and y (Ashby & Townsend, 1986). The plane shown passing through both functions describes the equal probability contours of the two distributions. Because the shape of the intersection of this plane with a given density (a circle in this case) does not change depending on the height of the plane (it will only expand or contract), it is useful to graph just these intersections. This is shown for our example case in figure 3b.

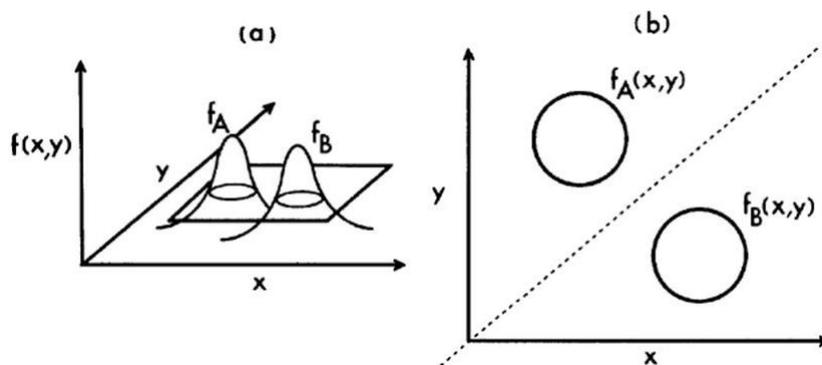


Figure 3: (a) shows the three dimensional densities of two distributions, and (b) is an equal likelihood plot of the same two densities

We can see that this latter graph can easily be mapped onto our earlier understanding of signal detection theory. The dotted line once again represents the decision criterion, the only difference is that it now depends on both dimensions (the decision on the x dimension depends on the level of the y dimension). Even though the decision for each dimension is dependent upon the other, in this case both stimuli are perceptually independent. What we mean by this is that the perceptual effects of one dimension do not influence those of the other dimension. If this property were violated, we would see that in the graph of equal probability contours. Instead of seeing circles, which portend perceptual independence, we would see skewed ellipses which would point up-right for positive correlation and down-right for negative. Having positively dependent variables means that the greater the perceptual effects are on one dimension, the greater they will be on the other.

It is important to note that perceptual independence is logically distinct from

having a decision criterion on one dimension that is not influenced by the other dimension. This latter property is referred to in GRT as decisional separability. A third, also logically distinct formalization of the idea of independence is called perceptual separability. This property means that the perceptual effects of one dimension are not dependent on the other. This definition sounds quite similar to that of perceptual independence, but the difference is that independence is a within stimulus condition, while perceptual separability is between stimuli.

Although GRT approaches the modeling of psychological phenomena in a fundamentally different way from MDS, because of its generality and versatility it has been shown that standard Euclidian MDS is actually a special case contained within the GRT framework (Ashby, 1988). When constrained in this manner, GRT will necessarily be required to assume the metric axioms that we examined earlier. However, in the fully generalized GRT model there is no need to make these assumptions. In GRT, the overlapping regions of multiple stimulus densities correspond to where they are confused with each other, rather than relying on a distance metric. Because these confusions are a function of both the means and standard deviations of the densities, the probability of a correct recognition is not necessarily monotonic with the distance between the perceptual means. Other metric violations can also be accounted for to yield an accurate description of the data in a wide variety of circumstances (e.g., Ashby, 1992; Kadlec & Townsend, 1992; Thomas, 1999, 2003).

IV: DIMENSIONS: FINITE AND INFINITE

The concept of dimension really only began to assume a rigorous treatment in

the late 19th century. There are now several mathematical approaches to dimensionality. First, let us just define "space" as some set of points, where a point is a "primitive", that is, an undefined entity. The "point" may in fact, be given in more immediately comprehensible terms, but need not—often it can gain its meaningfulness through a list of axioms about what structure it exists within (e.g., on the space itself) or operations that can be done on it, and so on. The easiest approach to understand is probably the one that defines dimension as "the minimum number of real numbers that can be employed to define a point in the space".⁴ Then an infinite dimensional space is one which requires a denumerable (i.e., can be put in one-to-one correspondence with an infinite set of integers) or non-denumerable set of numbers (e.g., the irrational numbers, products of such sets and so on), to indicate a specific point in the original space. Of course, this definition requires some sort of function that relates the points in the space to numbers, and that is far from always the most natural tack to take with some spaces. Nonetheless, it will be the most straightforward definition for our purposes. Our definition of a physical or psychological dimension is that it be representable by a possibly bounded interval on the real line.

Infinite dimensional spaces, while frequently obeying many precepts found in finite spaces, sometimes demand especial care, tactics and occasionally simply act in seemingly bizarre ways relative to finite spaces. A natural question from readers is likely to be why we need infinite dimensional concepts and especially in spatial terms. Thus, there appear to be a finite, if unbelievably huge, number of fundamental particles in the universe (However particles are defined by mod-

ern physics, and in spite of the particle-wave duality of quantum theory. And the definition of "fundamental" has altered over the past century with, string theory notwithstanding, no end in clear sight.). However, infinite dimensional spaces are a necessity for theory in modern science, including physics. One of the arguments for infinite dimensional models is that mathematical descriptions when the "points" number in the millions or billions, are simpler or more elegant, depending on the uses of the model. The same goes for spaces of very high dimension. Thus, Newtonian mechanics enjoys the artifice of continuous trajectories (where a trajectory contains a non-denumerable number of points) of objects in, say, 3-D space, though a modern quantum description might look quite different.

Most functions that even high school students meet, are defined on an infinite space, that of the real numbers (the latter being non-denumerable as we saw above, for it is made up of a denumerable set, the integers plus the rational numbers and the irrational numbers). Furthermore, the most useful of functions, such as the set of all continuous functions on the real line, are themselves infinite, with the latter set having the same dimensionality as the real line itself! In fact, the ubiquitous appearance of continuous (and often, smooth, i.e., differentiable to a high or infinite degree) functions in science by itself forms a powerful argument for the inclusion of infinite dimensional spaces in psychology. Whether or not the theorist looks to functions, and we welcome it (e.g., see Townsend et al., 2001; Townsend & Spencer-Smith, 2004; Townsend et al., 2005), even a kind of common sense consideration of such objects as the set of all faces seems to call for infinite dimensional spaces (see especially the first of the list immediately

above). Even though the number of faces on earth is finite, it is clearly possible in principle to create an infinite number of faces. Interestingly, even the modern approaches to approximation and numerical analysis, employ as a fundamental underpinning, the structures of infinite dimensional function spaces. And, psychologists and statisticians are beginning to work out statistical procedures and theories that are appropriate for interesting sets of functions (see, e.g., Ramsay & Silverman, 2006).

V: INSERTING OR EXTRACTING FINITE PSYCHOLOGICAL DIMENSIONS INTO (OUT OF) AN INFINITE DIMENSIONAL SPACE

Now, there are many directions we could take from here. An important one is: Is it possible to think of finite dimensional spaces as sub-parts of infinite dimensional spaces? This would be a boon for psychophysical scaling since it means that it would not be unnatural to think and work with finite psychological spaces, even if the "real" description demands infinitude. In fact, given that almost all real stimuli appear to be objects from infinite dimensional spaces, it is clear that our 150 years or so of finite dimensional psychophysics would lie on a very shaky foundation indeed if, say, a dimension like visual area or hue were not, first of all, mathematically separable in some sense from the more complex (infinite dimensional) signal of which it is a part, and secondly, the biological entity (e.g., the reader) were not able to somehow pluck this information from the signal. Thus, visual area of the surface of a dinosaur is not only computable from the complicated space of all dinosaurs but the visual system of a human can approximate that size, thus extricating the size dimension from this 'point' of that space. In

the case of hue, the actual signal may be a continuum of wave lengths from the light spectrum yet, the visual system computes a composite hue obeying the laws of color perception. Due to space concerns here (not infinite dimensional!), we shall only concern ourselves with the situation where a psychological dimension corresponds to a physical dimension. Of course, a physical dimension itself may or may not be very useful in science. For instance, products of powers of measurements, with the power being a rational number of basic physical dimensions (usually, mass, length, and time) constitute new dimensions but only a relatively small number of these are useful in physical laws (see, e.g., Chapter 10 in Krantz et al., 1971).

Consider a space X which is infinite dimensional according to the above discussion. Then to elicit a physical dimension that might (or might not) make up a psychological dimension, we simply need to map an interval (a, b) into X as in $f(x)$ where x is contained in (a, b) and $f(x)$ is contained in X . We write this more compactly as $f : (a, b) \rightarrow X$. While perfectly logical, this definition of a dimension as contained in X is not very intuitive. To unpack this situation a bit more, consider a depiction of infinite dimensional spaces based on a generalization of a so-called Cartesian product of dimensions. Thus, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, the combination of 2 selections from the set of real numbers. We can think of these combinations as a 2-place vector. This concept can be extended first to arbitrary finite combinations of \mathbb{R} , where we use the 'pi' symbol to indicate "Cartesian product", that is, all conceivable combinations as in $\prod_1^n \mathbb{R}^i$ and in this particular case, $\mathbb{R}^i = \mathbb{R}$ for $i = 1, 2$. In general then, consider an index set that, instead of $i = 1, 2, \dots, n$,

n is either countably (i.e., equal to the number of integers) or uncountably (equal to the number of all real numbers) infinite. Instead of using products of \mathbb{R} , we can use any kind of space; these spaces do not even have to be the same as one another. We can even take a combination of some finite and some infinite spaces. Then we have our space $X = \prod_{\alpha} X_{\alpha}$, which means that we take all the possible combinations of the spaces X_{α} , one value for each possible value of α . In addition to the apparent description of a generalized "point" in this space as an infinite dimensional vector (with position in the vector indicated by the value of α), we can also think of it as a function $f : \{\alpha\} \rightarrow X_{\alpha}$. For instance, when each X_{α} is \mathbb{R} and $\{\alpha\}$ is also in \mathbb{R} , and f is continuous, then each "vector" is a member of the set of all continuous functions defined on the interval (a, b) .

Now that we have a fairly intuitive idea of this highly useful type of space, we can simply form a new function which assigns a value in X for each member of the interval (a, b) . In a special case, it might be that, say X_{237} contains a 1-1 image of (a, b) , $f : (a, b) \rightarrow X_{237}$, and this either is a useful dimension for psychologists (e.g., loudness) or physicists, or both. So far, even with the foregoing explanation, the situation may seem abstract and of little use. But consider the more interesting example of the space of all faces and call it F . We shall postulate that F is contained in the space of all 2-D surfaces, considered as continuous functions of a bounded subset of some cross-product of intervals. It won't hurt to simply pose our mapping as a function from some domain D realized as the pair of intervals $(0, 1) \times (0, 1) = D$ into this face space, F . We shall focus, for simplicity, on the front half of the face. Basically, we can do the same kind of thing on the rear of

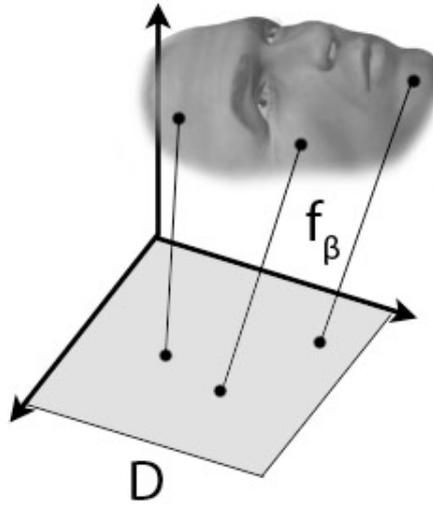


Figure 4: f_β is the mapping from the domain D onto a specific face in the face space

the faces.

Let the face function be $f_\beta : D \mapsto F$, over a non-denumerable set $\{\beta\}$, and we can without harm think of a particular face as being associated with a particular map, say f_β . Each face will be a different f_β mapping the domain into the face space such that every point in the domain corresponds to a point on the face. Figure 4 shows a face constructed in this fashion. Now, we mustn't lose track of our first goal, which was indicating how psychological dimensions can be exhibited in an infinite dimensional space, like our current face space F . Suppose we wish to experiment with the psychological dimension of mouth size (ms) and the set of all mouth sizes: $MS = \{ms\}$, contained naturally in \mathbb{R} . Of course, there is a true physical dimension of surface area of the mouth in this case, but that

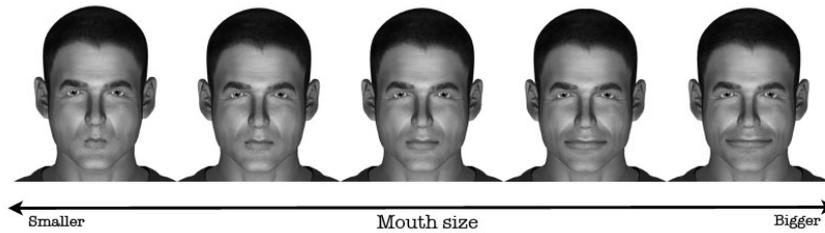


Figure 5: Here is a series of faces that vary across one isolated dimension, mouth size, but there are infinitely many other dimensions in the space of all faces

may well not be linearly related to the psychological dimension. We can include $ms = 0$, even though we may never actually observe that in reality. Anyhow, we embed MS in F by, for instance, taking all members of F but without a mouth and then forming the total $F = F_{-mouth} \times MS$. This operation constructs all possible faces with all possible mouth sizes. In actuality, some faces may not accommodate some, perhaps particularly large mouth sizes. This type of constraint does no real harm to our deliberations here, and in fact, a more sophisticated theory based on the idea of manifolds can readily encompass this type of constraint (see, e.g., Townsend et al., 2001). In any event, any 1-D curve in F can, for every face-point traversed, be indexed by a value of ms . In fact, any such trajectory through F can be modified so as to 'measure' mouth size at each of its points (i.e., at each face). Of course, a subset of trajectories will possess a constant ms while other facets of the faces vary, and so on. From the reverse perspective, we can think about how as attention is directed toward a mouth, perceptual operators are filtering the psychological mouth size from each face to which it is attending.

The present theory may seem intriguing (or too complex to exert any effort

for), but is there any hope for disentangling psychological dimensions from an infinite dimensional space? The answer is apparently "yes" in any cases we have seen. The Shepard-Kruskal approach and its many relatives might seem to apply most immediately to stimuli that are themselves exhibiting and varying the finite dimensions under study. A real life often employed case is a set of rectangles with varying length and width (Schonemann et al., 1985). Although the Shepard-Kruskal approach has never been given an axiomatic underpinning, nonetheless it seems clear that humans and probably many in the animal world can attend to a small number of dimensions at a time from stimuli that are inherently from an infinite dimensional space, especially when a finite set of 1-D dimensions are varying across the stimuli. In fact, even the set of rectangles can be interpreted as a very special subset of the set of all 2-D surfaces. The dimensions of psychological length and width would then be extracted by the Shepard-Kruskal procedures. It should be noted, though, that some individuals abstract the dimensions of size (length \times width) and shape (length \div width). In addition, the fact that infinite dimensional spaces are not incompatible with 1-D psychological dimensions helps weld together the various psychological tasks and processes, from unidimensional psychophysics (e.g., Fechner et al., 1966; Falmagne, 2002; Baird, 1997; Link, 1992) to higher order mental functions such as symbol and language identification (e.g., Townsend & Landon, 1982, 1983; Pelli et al., 2003; Rumelhart & McClelland, 1987), to categorization (e.g., Nosofsky, 1986; Ashby, 1992) and beyond.

Perhaps even more interestingly, by bringing to bear "principal component analysis" (closely related to "singular value decomposition"), it is feasible in prin-

ciple to dissect a set of perceived patterns, even faces, into a set of objects from the same dimensional space, that can (at least so the wish goes) be interpreted as a set of faces that serve as a foundation from which all the faces in the stimulus set, and hopefully many more, can be built up. In the example case of our face space, these foundational "basis" faces are called "eigen faces". More rigorously, this approach originally stems from the theory of vector spaces. Many of the present readers may have taken a basic linear algebra course, where they learn about sets of independent basis vectors (possibly, but not necessarily at right angles to one another), weighted sums of which can produce any vector in the space (of course, we have to omit a lot of detail here!). It turns out that, say, the space of continuous, differentiable functions, which can be multiplied by numbers and then added, subtracted, etc., form a valid, but infinite dimensional vector space. Nonetheless, the property of infinitude does not rule out the possibility of finding a set of eigen vectors (in this case, actually eigen functions!), infinite in number, which can, with the proper numbered weights, exactly reproduce any of the original stimulus functions. In fact, in continued analogy to the common vector spaces to which the readers are undoubtedly accustomed, there are many ways of selecting a usually infinite, but discrete, set of basis vectors which perform the foregoing service. Furthermore, it also happens that these functions (or in the present case eigen faces) can be ordered in their importance for producing the human behavior, for instance, similarity judgements, confusion probabilities and so forth. Hence, the investigator will take a finite set of these eigen faces as being the most important, down to a certain but arbitrary level of precision, and

use them as hypothetical representations of the basic faces of which "all" others can be reconstructed. Naturally, since we end with a finite number, these will only approximate the original faces to some degree. Anyhow, this seemingly kind of exotic and unpromising approach has actually been carried out with great success by a number of laboratories. An auspicious example is found in the research of Alice O'Toole and her colleagues (see, e.g., O'Toole et al., 2001; Deffenbacher et al., 1998), where eigen faces have been employed to discover many intriguing aspects of face perception. Our only word of caution here is that often the eigen faces may not resemble real faces to a high degree. This facet is important in our overall discussion, since even though the space of continuous functions is a legitimate vector space (e.g., a weighted sum of two continuous functions is again a continuous function), the space of all faces is not. This is because not all faces would necessarily be representable as some sort of "combination" of basis faces. This lack of nice vector space qualities, leads us to consider a more general framework, in fact, one where locally, that is, within a small region, one has vector space characteristics (in fact, Euclidean properties), but where globally, the space will not be either Euclidean or a vector space. A natural setting to consider is that of a manifold, which is also the framework within which Einstein's theory of general relativity came to be expressed. In order to help the reader develop some intuition for these rich spaces, we actually need to take a step back to rather primitive, but extremely useful concepts, such as a "topology".

Finally for this topic, we observe that the question of independence of psychological dimensions has long been of interest in perceptual science. For instance,

the dimensions of loudness and intensity (commonly, and somewhat oddly called "volume" in English) are mutually dependent, although their physical sources are not. Within a more complex setting, social psychologists have found that perceived intelligence and attractiveness are correlated although, of course, they are not in reality. Ashby & Townsend (1986) proposed a general theory of perceptual dependence among psychological dimensions embedded in a multi-dimensional pattern recognition setting. Many supplements to the theory and associated methodology have been made and applications in various areas of perception and action (e.g. Kadlec & Townsend, 1992; Thomas, 2003; Maddox, 2001; Wenger & Ingvalson, 2003; Amazeen, 1999).

VI: MANIFOLDS FOR PSYCHOLOGICAL SPACES

We put aside for now, the property of having an infinite number of dimensions to pursue a different course. Of the virtually infinite number of possible generalizations of Euclidean geometry, there is another one that stands out which has been hardly explored at all in psychological domains. We are referring to "manifold theory". The concept of a space plus a topology is so important that we have to deal with it, at least informally, before beginning in earnest. First, a "space" is any set (collection) of entities we call points. These can, as in the kinds of spaces with which the reader is likely most familiar, be approximated by dots. In more complicated situations, however, such as the space of all continuous functions, a "point" would be an entire continuous function. The space is then the collection of all of these "points". Anyhow, a mathematician usually quickly imposes more structure on the space. This catapults us to the idea of topological spaces.

A topological space X is a set of points with a special collection of subsets of points which are called "basis sets". One then takes all possible finite numbers of intersections of these sets along with all possible finite or infinite unions of these sets to produce the so-called "set of open sets" (we also must include X , the set of all points that make up the space, plus the empty set, \emptyset). That is, every set in this possibly very big set of sets is by definition "open". The so-called "closed" sets can then be elicited by taking the complement $X - O_\alpha = C_\alpha$ where O_α is open and C_α is then closed. With this structure and little else, one can immediately define continuous functions from one topological space X to another Y , along with many other valuable concepts. If X and Y are topologically equivalent, one can find a bicontinuous function that carries every point in X to one in Y and vice versa, and one already has the justification for the oft repeated topology joke that "A topologist is a person who can't tell the difference between a tea cup and a doughnut", since, of course, the stretching and shrinking allowed in such a function preserves the topological properties (e.g., how many holes a space possesses) of either space. Topology is extremely powerful for the relatively small set of assumptions on which it rests. For even more interesting properties, we now proceed to topological manifolds.

The essence of a topological manifold M^n rests on three vital properties: 1. M is Hausdorff. This means that for any two points, two sets can be formed containing those points which have no intersection (overlap). 2. M has a countable basis. 3. M is locally Euclidean. That is, any point in M is contained in a small open set O_M (called a "neighborhood") which can be mapped bicontinuously (both the

original function and its inverse are continuous) and in a 1-1 (each point in O_M maps to exactly point in an open set O_E of E^n) and onto (all points in O have an inverse point in O_M) fashion. This means that a small region of our manifold can be treated approximately (and totally in the limit as the original set gets smaller and smaller) like a Euclidean space. We can now use the Euclidean metric locally, $d(x_1, x_2) = \sqrt{(\sum_1^n (x_{1,i} - x_{2,i})^2)}$ where $x_1 = (x_{1,1}, x_{1,2}, \dots, x_{1,n})$, an n-place vector in Euclidean n-D space, and similarly for x_2 . It can be shown that a topological manifold can always be granted a global metric. However, without more assumptions, that metric may not be the one we wish—a Riemannian metric, named after the great mathematician Bernard Riemann. Riemann furthered the drive toward non-Euclidean geometry, and worked out a type of metric which generalized the one invented by another mathematical genius, Karl Friedrich Gauss for special cases, and which included as special cases, those posed by Lobachevsky and Bolyai (negative curvature), but long before the invention of the curvature concept. Riemann's manifolds and his measure capture both positive as well as negative curvature.

Some bright mathematicians met this need and bestowed the drive for a generalized metric with great force, by inventing a kind of differentiation that works on manifolds instead of just in Euclidean space. This topic is much more recondite than what we can detail here but basically we can view the operations in a simple case to get a feeling for what is going on. Basically (and very roughly), we can think of bundles of vectors that sit on a manifold and starting at any single point, direct us to the next point. With some care, we can provide differentiation

operations on these vectors such that the 'output' of the differentiation operator remains in the original manifold (e.g., in a manifold that appears as a surface in \mathbb{R}^3 , the derivative vectors might stick out away from the surface we are studying). Without care taken to 'confine' them to the manifold at hand, we can't represent speed and acceleration to provide for a Newton-like mechanics in our novel space. Closely associated with our setup is a matrix (or quadratic operator) which maps velocity vectors into a speed number. This result can be integrated over a path to produce a path length. Even better, we can put conditions on the paths such that we can evoke the closest possible analogue to straight lines in Euclidean spaces (formally called "geodesics"). We define the "Riemannian metric" as an $n \times n$ matrix (where n is the dimension of the manifold), which is usually written $g_{ij}(x)$, with i, j running from 1 to n and x being a point on the manifold. It tells how fast distance is accumulating at different places in the space (signified by x) and how the different dimensions, i and j , affect that accumulation at each point. For instance, suppose that our manifold is shaped like a mountain and that we wish our metric to take into account not only distance in our usual sense but also perhaps the effort (e.g., power) that is expended in moving around on this manifold. Hence, when the grade is fairly flat, we can make $g_{ij}(x)$ small but when the going gets steep we can make it much larger. Suppose too that our journey starts at a point where, when we move in the direction of both increasing dimension i as well as dimension j we are heading up the steepest part of the slope. If we momentarily hold one dimension constant, the effort is considerably less (due to not actually climbing upward). This situation is picture with a large $g_{ij}(x)$ close to the starting

point and $g_{ii}(x)$ and $g_{jj}(x)$ being smaller (in Euclidean space, these are the only entries in the matrix that are appropriate, and all are equal to 1 for all points x). Note the strong interaction between dimensions here, which is absent in Euclidean space. Interestingly, this (relatively) simple function is all that is required to investigate the curvature in our space, a fact that was prefigured by Karl Friedrich Gauss and made general and rigorous by Bernard Riemann (Gauss was a senior professor and sat on Riemann's Habilitation exam [kind of like an advanced doctoral dissertation in Germany, but being dissolved as we speak]). Even with this brief (and quite non-rigorous) tutorial, we can perhaps see that manifolds increase the generality and scope of the kinds of spaces we could consider for psychology magnificently.

As a simple, finite dimensional example, consider the traditional shape used in color theory to capture some important aspects of color: A. Hue. B. Brightness. C. Saturation (or its opposite, Complexity). When we don't worry too much about the precise shape, the surface looks like two ice cream cones with their tops (the biggest part of the cone) stuck together. This is a manifold which can, of course, be embedded in Euclidean 3-D space. The middle, widest portion, is used to depict the optimal brightness level, where the full range from gray (in the very center of the circle) to the brightest hues is possible. As we descend toward the bottom apex, all is dark, and up at the other apex the stimulus light is so bright that no hue can be discerned. This little space is very well behaved (except at the apexes, where strange geometric things happen, so we put those aside for now). Using Riemann's techniques, we quickly learn that although 'in the large'

the space is not globally like a Euclidean space (e.g., it is bounded rather than unbounded), locally around a given point, its curvature is 0 just like in Euclidean space. Intuitively, this is because our double cone, with a cut from bottom to top, and one through the center, unrolls to make a portion of a Euclidean plane.

VII: PROBABILITY THEORY IN INFINITE DIMENSIONAL SPACES

So, now we are in possession of elementary topology, "manifolds", metrics, differentiability (what mathematicians like to call "smoothness"), and finite and infinite dimensionality. Next we need a way to induce "randomness", that is "probability", on our finite or infinite dimensional manifolds. Fortunately, there is a well-trodden pathway that allows us to retain our desirable generality. This comes by way of utilizing the resident topology to our own ends. First, we observe that probability is a form of the rigorous concept of "measure". A "measure" is constituted by a function that maps sets (often called "classes") of sets into the real numbers in a regular fashion. The conditions to satisfy the tenets of a measure, and again we cannot reach detail here, are things like: 1. Additivity of the measure of non-overlapping sets in the topology. 2. Finiteness of the measure, i.e., the measure is always bounded by some fixed real number. There are several others that we won't talk about here, but with the added stipulation that if we have a probability measure, the measure on the entire set of points (i.e., the points in the topological space) is equal to 1 (i.e., something has to happen). Hence, we simply form an appropriate topology, then take its sets and assign a probability measure to them. A class of sets (along with certain operations) meeting these and other technical tenets is called a "sigma field". Basically, one can assess the

probability of an event by computing the measure associated with appropriate sets in the sigma field. This general argument applies to finite or infinite dimensional spaces. As might be expected, peculiarities can arise in the latter case, but usually they don't unduly perturb the pathways to the theorist's goals. These deliberations should convince us that infinite dimensional spaces can possess probability distributions, particularly when a metric is present. In many cases, some of which are standard in such fields as electrical engineering (although the rigorous underpinnings might be found only at the graduate school level), one may employ tools from other fields, such as functional analysis, or stochastic processes, to avoid the explicit production of an appropriate sigma field. More on this topic later.

Any topological manifold can do this, but we are especially interested in (infinitely differentiable) Riemann manifolds, which are all metrizable, and in fact, we wish to only work with those that are complete metric spaces. A nice property of metrics is that any metric generates a topology but not all topologies admit a metric. The finite cases are dealt with elsewhere (e.g., Townsend et al., 2001), so we want to see what happens with infinite dimensional manifolds or other infinite dimensional spaces. Perhaps the most straightforward infinite dimensional space is the space of all continuous functions on an interval I on the real line, which may itself be infinite (i.e. $[0, 1]$), say $f : I \rightarrow \mathbb{R}$. Of course, the graph of each of these functions yields the usual picture of a typical function as taught to us in elementary math. Interestingly, it has been shown that this class of objects (i.e., the set of functions) is of the same dimensionality (the jargon term is "cardinality") as the points on the real line itself! There are lots of metrics we could use but an

extremely useful type is the so-called \mathcal{L}^2 metric $d(f, g) = \int_0^1 (f(x) - g(x))^2 dx]^{1/2}$, where f and g are two such functions. It should be obvious that this is the analogue of the Euclidean metric in finite spaces. Even more intriguing is the fact that it is the only power metric in function space that satisfies the conditions required to be a Riemannian metric. Thus, the infinite dimensional 'twin' of the city block metric (see above), $d(f, g) = \int_0^1 |(f(x) - g(x))| dx$ is not a type of Riemannian metric on any manifold!

We back up for a moment in order to examine the finite n -dimensional metric and a path length in n -dimensional space. The general infinitesimal displacement in n -dimensional Riemannian manifold is $ds = \left\{ \sum_{i,j=1}^n [g_{ij}(x_1, x_2, \dots, x_n) dx_i dx_j] \right\}^{1/2}$. Observe that g_{ij} is simply an $n \times n$ matrix and that it can depend, in general, on the specific point in space, (x_1, x_2, \dots, x_n) , where we are currently. However, in the special case where $g_{ij} = I$, the identity matrix at all points in the space, this expression reduces to the Euclidean metric. Likewise, the path length of a path through Euclidean n -space is $\int \dots \int \left\{ \sum_{i,j=1}^n [g_{ij}(x_1, x_2, \dots, x_n) dx_i dx_j] \right\}^{1/2}$, where n integrals are taken tracking the appropriate path. Perhaps it is even more intuitive when the displacements dx_i are converted to velocities, dx_i/dt , for then we can simply consider the path in terms of the velocities of each coordinate. Here we evaluate, say dx_i , as being a tiny motion in the x_i direction, for $i = 1, 2, \dots, n-1, n$. This obviously pushes us to a new point $(x'_1, x'_2, \dots, x'_n)$ where g can take on a new value, and so on. All this will come in handy in the infinite dimensional case just following.

Moving back to functions in one variable, since it is more intuitive to confine ourselves to this "space of differentiable functions in one dimension", let us indicate the general Riemannian metric on this elementary manifold:

$$ds = \int_0^t \left(\int_0^1 \int_0^1 g_{\alpha\beta}(h) (\partial f(\alpha, t) / \partial t) (\partial f(\beta, t) / \partial t) dx_\alpha dx_\beta \right)^{1/2} dt.$$

Now, we have to adjust our thinking a bit. Instead of considering dx_α as a miniscule motion in the x_α direction, we can think of it in the following way. Take a look at the two facial profiles in Figure 5, drawn as a function on x from 0 to 1. Now, for every α , x_α is a number between 0 and 1. However, we now need to think of dx_α as a motion from, say face f_1 to face f_2 , that is moving up or down vertically between the two designated faces. Because of the possibility of noise, we can only assume that each function along the path, that is, h is "face-like", not that it is necessarily a true face. Indeed, in a pattern identification situation one or both of f_1 and f_2 could be noisy renditions of faces. The extra complexity here must be explained a bit more. We are representing a function as a gigantic (and dense!) vector, $\gamma = f(x), x \in \mathbb{R}$, and since f is our current 'point', we have to let the metric g depend on it, which is why it appears in the argument of g . In addition, instead of dx_i for a finite $i = 1, 2, \dots, n$, we must extend this notation to dx_α where α runs over the interval $[0, 1]$, and the same is true for β . So, the overall idea is that we are depicting the movement of f by way of how it changes with each 'coordinate' of f . And, since we are using a general quadratic metric of the coordinates, whether finite or infinite, we have to take the two-way products of

all possible coordinate changes, that is dx_α and dx_β . Finally, then the path length through this function space, explicitly using our velocity representation, is just

$$d^R(f, g) = \int_0^{t^*} \int_0^1 \int_0^1 g_{\alpha\beta}(h) (dx_\alpha/dt) (dx_\beta/dt)]^{1/2} dt.$$

Of course, when we seek the shortest path length between two faces, we can define that as the distance and the ensuing path as the geodesic between them. Again, the reader may wish to confirm that when α, β are both identical to 1 and don't depend on "f", we get back to the analogue to the Euclidean metric, the \mathcal{L}^2 metric (and see Townsend et al., 2001, for more on this and other issues pertaining to face-geodesics). In any event, all this can be (with some tedium) expanded to functions in any finite dimensional space. It is important to observe before we go further that just because we use the term "path" here, as is common, we do not imply a temporal factor. All the above, for instance computing the distance between two faces, might take place simultaneously along the path, although some paths might nonetheless take longer to compute than others. A more abstract and general notation is possible (e.g., Boothby, 1975; Townsend et al., 2001). Such notations are very useful because they capture the principal ideas in a markedly clear fashion without sometimes mind-numbing profusions of indices. On the other hand, when the scientist wishes to actually compute something, the evil indices must be present and accounted for.

VIII: A SPECIAL CASE OF GREAT IMPORTANCE

Of course, infinite dimensional manifolds have not yet seen much application in most of science, especially the life sciences. Yet, it seems worthwhile to mention a special case of our quantitative apparatus that has been of enormous

value in the basic and applied sciences. The theory goes by many names, but one term is the "mathematical theory of communication" (which we will simplify to "MTC"), used for example by Norbert Wiener. Like virtually everything, including mathematics, there are roots in the distant past for this theory, but a real explosion of new results from mathematicians, physicists, and engineers, occurred in the mid-twentieth century. To a major extent, this work accompanied the amazing scientific and technological effort expended in and around the time of World War II. We recall the names of formidable mathematicians such as Norbert Wiener, John von Neumann, and electrical engineer Claude Shannon. The sphere of mathematical communications theory intersects huge regions of mathematics and physics, including functional analysis, stochastic processes and probability, wave-form analysis (e.g., Fourier and Laplace transforms), electrical engineering, differential equations and deterministic and stochastic dynamic systems, and so on. In the present writers' opinion, the whole of this work is as important to modern technology and science (if not usually so lethal in application) as the splitting of the atom.

In most applications of MTC, we employ various types of function spaces. These may be deterministic or probabilistic, but usually do not require manifold theory per se. When needed, differentiation can be extended to so-called Frechet' or Gâteaux derivatives, and usually ordinary Lebesgue or Riemann integration suffices for integration. For many purposes, the function space is itself a vector space, sometimes with a norm (as in Banach spaces—the norm of a function is the analogue of the magnitude of a vector in a finite dimensional vector space) or an

inner product (the analogue of the dot product in a Euclidean vector space). The inner product, the analogue of the dot (or ineptly called by non-mathematicians the "cross product") of two functions of a real variable is just $\int_{-\infty}^{\infty} f(x)g(x)dx$. The accompanying metric, which by now will probably not shock the reader is simply $d(f, g) = [\int_{-\infty}^{\infty} (f(x) - g(x))^2 dx]^{1/2}$, which naturally looks very much like our good friend, the Euclidean metric, but stretched out along the real-line continuum. We can then think of $d^R(f, g) = \int_0^{t^*} [\int_0^1 \int_0^1 g_{\alpha\beta}(f) (dx_{\alpha}/dt) (dx_{\beta}/dt)]^{1/2} dt$, which we formulated above, as weighting distinct parts of figures such as faces differently, depending on their importance. For instance, it is well known that the eyes play a critical role in face perception and hence we would expect $g_{\alpha\beta}$ to be large when transversing that part of the face.

IX: APPLYING THESE IDEAS TO CLASSIFICATION

Now, suppose that a person is confronted with a pattern from an infinite dimensional space. As pointed out above, almost everything seen or heard is from such a space, not a finite dimensional one. In applying the above concepts to this situation, we have to make sure that our space is 'big enough' to contain not only patterns arising in a noiseless, perfect perception environment, but also the patterns that are perturbed by some type of noise, or that are simply randomized in some way. Thus, the theoretical system we propose is an analogue to the fact that probabilistic identification in a noisy environment (e.g., Ashby & Townsend, 1986), demands a very similar type of theoretical structure as does categorization of a set of patterns which are associated with a probability distribution (e.g.,

Ashby, 1992). First, let us agree that some kind of representation is constructed of each pattern to be classified, whether it is categorized by a set of exemplars (see Nosofsky, 1988) or a template. In identification, it might be a memory of each object to be uniquely identified. Let us concentrate on the case of identification since one can readily generalize that situation to categorization or reduce it to signal detection. The reader may refer to detailed developments and discussion of this material for the finite vector case in Townsend & Landon (1983).

Let us conceive of the input as a random function in some n -dimensional space (e.g., a perturbed 2-dimensional surface plus noise placed in a 3-dimensional Euclidean space with an \mathcal{L}^2 metric). It is standard to use an \mathcal{L}^2 metricized function space for physical signals so we'll continue that here (but cf. Cornish & Frankel, 1997). Let U_i be the random function describing the probabilistic input for stimulus S_i and $U = \{U_i\}_1^N$ the set of N randomized inputs. We'll take X as the perceptual image space. X is considered as an infinite dimensional function space, but may be embeddable in some finite dimensional space. So, the perceptual map is $U_i \rightarrow X$, and each realization f of the set F , where $F \subseteq X$ is assumed to be continuous. The memory pattern set Y , with which the input is matched, will be taken as a deterministic set of functions. That is, $g \in Y$ is well memorized with no noise attached. In certain cases, not having both be random can substantially simplify matters like calculations.

Next, we need an evidence space Z to represent the evidence for each resulting comparison of an input with the N memory items from Y . The simplest, yet very natural space to employ here is just \mathbb{R}^+ , that is, the positive real numbers. Hence,

we must have a map $e : (F, G) \rightarrow \mathbb{R}_0^+$, that is, $e(f, g) = r \in \mathbb{R}_0^+$, with the "0" subscript indicating that 0 is included (naturally, this set can easily be generalized to include the negative real numbers if required). The evidence function could be calculating distance, Bayesian likelihood, or some other measure of acceptability of a memory alternative. In any event, $z \in Z = \mathbb{R}_0^+$ will be a real-valued random variable, and the image of a particular (f, g) .

Now, within X every point function, except for ties, will be in favor of one response alternative over the others, say the "jth". Then, for a perceptual signal f in such a set, $e(f, g_j) = \text{MAX}_1^n[e(f, g_k)]$, again except for ties. This means that outside of ties, X will be partitioned into sets of points (i.e., functions) that favor some one of the alternatives. Then the probability of responding R_j when the stimulus was S_i requires determining the probability assessed across the various percepts, f , that can appear when S_i is presented. This probability will be $P(R_j|S_i) = P[e(f, g_j) = \text{MAX}_1^n[e(f, g_k)]|S_i]$.

Interesting facts about the evidence functions:

1. As noted, under very weak conditions (e.g., there exist points in X where $e(f, g_k)$ favors R_k for all k ; and so on) "e" will partition the space into mutually exclusive regions that favor each of the alternatives plus points where ties may occur.

2. Consider pairs of, say, g_i and g_j , distinct memory patterns and the set of $f \in F$ such that $h(f, g_i, g_j) = e(f, g_i) - e(f, g_j) = 0$. Then, in most circumstances, the border separating whether R_i wins vs. where R_j wins, will be a closed submanifold in X . We can call the set $\{f\}$ such that $h(f, g_i, g_j) = e(f, g_i) - e(f, g_j) = 0$, the

"kernel of h , and again they are "equidistant" from faces g_i and g_j . On either 'side' of this boundary one or the other 'wins'. Now, the same things happens when any pair is considered and we can also look for ties among three, four or more of the set of N faces. Under fairly weak conditions, the distinguished set $\{f\}$ (e.g., designated by the 'tied' distance) will even be a nice sub-manifold of the original face space, inheriting its topology from that of the 'parent' space. Sometimes the region of face space where, say, g_i "wins" over its competitors, will be connected. However, devotee's of signal detection in a single dimension will recognize that even there, the region of points where "YES" dominates the "NO" decision will not be connected: When the distributions of signal + noise vs. noise-alone are normal with unequal variances, this "disconnection" always occurs under a "maximum likelihood" decision rule. Yet another nice property ensues if the equidistant boundaries all have probability = 0 of occurring, for then we don't have to worry about jumping outside of our evidence space to adjudicate ties. This is quite a natural occurrence for finite dimensional spaces.

What happens when the situation is stochastic? Suppose the human observer or signal processor, let's call her Sheila, is at least to a first approximation a linear filter, and that she is attempting to recognize one of N patterns. Suppose that the observer is deterministic (that is, her filtering mechanisms act the same way each time they are called into play) and that a specific signal "i" ($i = 1, 2, \dots, N$) is itself a continuous function ($s_i(t)$) with Gaussian noise ($\tilde{N}(t)$) added in. Then, the signal pattern can be expressed as $U(t) = s_i(t) + \tilde{N}(t)$. The observer's filter for each signal possibility (e.g., the "jth") can be written also as a function of t , $h_j(t)$. It

turns out (see, e.g., Luenberger, 1979; Padulo & Arbib, 1974) that her output on her j^{th} perceptual 'template' is $x_{ij}(t) = \int_0^t h_j(t-t')[s_i(t') + \tilde{N}(t')]dt'$, under some reasonable conditions. That is, Sheila is using a template across time represented by h to filter or compare with the input. In fact, when the noise has certain properties, $h_j(t) = s_j(t)$, which means the filter-template is a replica of the j^{th} pattern itself and the filtering action is basically a correlation of the input with each one of these N stored replicas. This action produces $x_{ij}(t) = \int_0^t s_i(t')[s_i(t') + \tilde{N}(t')]dt'$. There are several decision structures that could be imposed, but given our space considerations, we take the most straightforward: we suppose that Sheila samples information for a fixed interval, say, $[0, t^*]$ and then selects the maximum 'correlation' provided by $x_{ij}(t^*)$, as calculated across "j" for a given present signal "i". Under some restrictions, this strategy is optimal and is called a "matched filter". Indeed, if $\tilde{N}(t')$ is stationary Gaussian white noise with variance $\sigma_N^2 = 1$ and mean $\mu_N = 0$, $x_{ij}(t)$ will be itself normal with mean $\mu_x = \int_0^t s_i^2(t')dt'$ and variance $\sigma_x^2 = \int_0^t s_i^2(t')dt'$, the very same thing!¹¹

What happened to our metric? Well, with the same sampling rule but now using the simplest Riemannian metric, the L^2 metric, we would compute $d(s_i(t'), x_{ij}(t^*)) = [\int_0^{t^*} \{s_i(t') - [s_j(t') + \tilde{N}(t')]\}^2 dt']^{1/2}$. Now this quantity is monotonic with its square which is easier to deal with so we examine $d^2(s_i(t'), x_{ij}(t^*))$ instead. We find that

its mean of expectation (signified by the operator "E") is

$$\begin{aligned}\mu_d &= E \left[\int_0^{t^*} \{s_i^2(t') - 2s_i(t')s_j(t') + s_j^2(t') - 2(s_i(t') - s_j(t'))\tilde{N}(t') + \tilde{N}^2(t')\} dt' \right] \\ &= E \left[\int_0^{t^*} \{s_i^2(t') - 2s_i(t')s_j(t') + s_j^2(t') + \tilde{N}^2(t')\} dt' \right]\end{aligned}$$

due to the fact that the mean of $\tilde{N}(t') = 0$. Next, $\int_0^{t^*} s_i^2(t') dt'$ is a constant and the same for all comparisons (recall that $s_i(t')$ is the presented signal) and $\int_0^{t^*} \tilde{N}^2(t') dt'$ is a random variable but is also the same on any one trial for all the comparisons and so neither of those quantities can discriminate the various decision/response alternatives. Hence, the only operative quantity is $\int_0^{t^*} \{-2s_i(t')s_j(t') + s_j^2(t')\} dt'$ with mean $\int_0^{t^*} \{-2E[s_i(t')s_j(t')]\} dt' + \int_0^{t^*} E[s_j^2(t')]\} dt'$. Notice, in particular that the term $\int_0^{t^*} \{-2s_i(t')s_j(t')\} dt'$ is directly proportional to the critical term from our matched filter approach above, namely involving the integral of $s_i(t')s_j(t')$. Therefore, minimizing the distance is tantamount to maximizing the correlation between the input and the decision alternatives. The only difference is that our distance approach includes a biasing term, $\int_0^{t^*} s_j^2(t') dt'$, which provides a relative bias for alternative "j" according to its energy or magnitude. We won't detail the derivation of the variance this time but it is also closely related to that of the matched filter expres-

sions. Interestingly, both these rules are also equivalent under certain constraints to a rule based on maximizing the likelihood that the presented signal pattern was "j", given the observation.

X: CONCLUSION

Our itinerary on the geometric aspects of psychological classification has taken us from foundational, axiomatic measurement theory through (finite) multidimensional scaling, to concepts of dimensionality, including infinite dimensional spaces. From there, we reassured ourselves that finite dimensional subspaces of infinite dimensional spaces are legitimate mathematical concepts, meaning that models which perform dimensional reduction can readily be applied as hypotheses about the human abstraction of psychological dimensions. This is something that is taken for granted, but its truth is not at all obvious, although we certainly find some way to filter interesting dimensions. Next, we moved on to ideas that are still relatively new in mathematics (i.e., only around 150 years old!), that of non-Euclidean and Riemannian geometries, and even more ambitious, infinite dimensional Riemannian spaces. In the social and biological sciences, probability and stochastics are a necessity, and we briefly surveyed the prospects for placing probability distributions on finite and infinite dimensional manifolds. It can be said that mathematicians are still in hot pursuit of the best ways of carrying out this program. Nonetheless, the next two sections exhibit important special cases and applications to classification illustrating theory and methodologies that have been around since at least the 1940's and 50's in engineering and computer science, and have been expanded and deepened in the meantime. We take it as a

plausible working hypothesis that the gargantuan corpus of questions in social, biological, and even such areas as cognitive science, including machine intelligence and human-machine interactions, ultimately cannot all rest comfortably or rigorously within the simpler types of geometric spaces which have dominated those areas (with the exception of the special cases treated in the last two sections, which have seen extensive implementation in engineering and computer science). That is, in the final analysis, an empirical question, but we believe our present and future researchers should accouter themselves with the powerful tools that can aid in answering fundamental questions such as these. Over the centuries, the symbiosis between physics and mathematics has tremendously enriched both fields. Up to now, the social and biological sciences have largely been on the 'borrowing' rather than 'loaning' side of the interactions, but there are strong signs that this is changing, for the better!

FOOTNOTES

1. The primary disciplines of the authors are psychology, and cognitive science. We shall use "psychology" as a proxy for any of the social or biological fields that cannot be readily specified in terms of physics (an example of a candidate for one that can be so described might be quantum properties of a neurotransmitter in re-uptake dynamics).

2. Modern psychology now includes a tremendous effort in neuro-sciences, especially neuro-imaging and of course, also affords a rich domain of mathematical research possibilities—thus, for the latter, see the websites for the US Society for Mathematical Psychology and the European Mathematical Psychology Group.

3. There has been much work mostly theoretical in the interim. Narens (1985) studies the critical notion of 'meaningfulness' in foundational measurement. Luce & Weber (1986) provide a in-depth account of axiomatic decision making from an axiomatic measurement theory viewpoint.

4. Mathematical psychology is a subfield of psychology where, in place of verbal theorizing, mathematical theories or as they are often called, "models", are utilized to make one's assumptions rigorous, and to make strong predictions that are testable by observational data. Townsend & Kadlec (1990) offer a brief overview of the major branches of mathematical psychology and Townsend (2008) discusses challenges and prospects for mathematical psychology in the twenty-first century.

5. ? was not the first to think of identical features as affecting the psychological similarity of two objects. He traces his theory to seminal ideas of Restle (1961) who proposed set theoretic means of assessing similarity which could include the former. Other uses of this concept in models of pattern recognition can be found in work by Townsend & Ashby (1982). Nonetheless, Tversky's theory was by far the most quantitatively thorough and was developed in the context of the previously discussed foundational measurement theory.

6. The theory (or rather theories) of dimensionality has grown over the years. Relatively deep earlier treatments along some avenues can be found in Hurewics & Wallman (1941) and Nagata (1965).

7. A limitation of principal components analysis is that it assumes that the basis vectors be orthogonal to one another. It might well be that the basis vec-

tors for a psychological vector space are linearly independent but not orthogonal. A methodology which assumes independence but not orthogonality which has gained much attention lately, especially in the analysis of fMRI signals (literally "functional magnetic resonance imaging, a neuroimaging method based on measurement of blood oxygen levels and their changes during psychological tasks) is called "independent components analysis" (ICA). However, this approach too, has its limitations.

8. It bears mentioning that relativity theory, even the special theory, requires adding time as a negative number in figuring the distance in the space (e.g., Minkowski, 1908).

9. The Ashby group has developed a general theory of categorization based on general recognition theory (e.g., see Chapters in Ashby, 1992).

10. A volume on applications of tensor analysis (a typically applied branch of differential geometry) by Schouten (1951) is a landmark in the applications of tensor theory to physics. However, it stands as one of the more challenging-to-read treatises in science, due to the maze of indices.

11. This nice event comes about due to stationarity, the properties of white noise, and the values of the noise mean and variance.

References

Amazeen, E. L. (1999). Perceptual independence of size and weight by dynamic touch. *Journal of Experimental Psychology: Human Perception & Psy-*

- chophysics*, 25(1), 102-119.
- Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics*, 44, 195-204.
- Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154-179.
- Baird, J. C. (1997). *Sensation and judgment: Complementarity theory of psychophysics*. Lawrence Erlbaum Associates.
- Boothby, W. M. (1975). *An introduction to differentiable manifolds and riemannia*. Academic Press, New York.
- Cornish, N. J., & Frankel, N. E. (1997, Aug). The black hole and the pea. *Phys. Rev. D*, 56(4), 1903–1907.
- Deffenbacher, K. A., Vetter, T., Johanson, J., & O'Toole, A. J. (1998). Facial aging, attractiveness, and distinctiveness. *Perception*, 27(10), 1233-1243.
- Falmagne, J. (2002). *Elements of psychophysical theory* (Vol. 01). Oxford University Press, Incorporated.
- Fechner, G., Adler, H., & Boring, E. (1966). *Elements of psychophysics*. Holt, Rinehart and Winston.

- Hurewics, W., & Wallman, H. (1941). *Dimension theory*. Princeton University Press.
- Kadlec, H., & Townsend, J. T. (1992). Signal detection analysis of multidimensional interactions. In F. G. Ashby (Ed.), *Probabilistic multidimensional models of perception and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement volume i: Additive and polynomial representations*. Academic Press, New York and London.
- Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Luce, R., Bush, R., & Galanter, E. (Eds.). (1963). *Handbook of mathematical psychology* (Vol. I-III). Wiley.
- Luce, R. D., & Weber, E. U. (1986). An axiomatic theory of conjoint, expected risk. *Journal of mathematical psychology*, 30(2), 188-205.
- Luenberger, D. G. (1979). *Introduction to dynamic systems : Theory, models, and applications*. Wiley.
- Maddox, W. T. (2001). Separating perceptual processes from decisional processes

- in identification and categorization. *Perception & Psychophysics*, 63(7), 1183-1200.
- Minkowski, H. (1908). Space and time.
- Nagata, J. (1965). *Modern dimension theory*. Amsterdam: North-Holland Publishing Co.
- Narens, L. (1985). *Abstract measurement theory*. Cambridge, Massachusetts-London: The MIT Press.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1), 39-57.
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14(4), 700-708.
- O'Toole, A. J., Wenger, M. J., & Townsend, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedent and possibilities. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (p. 50). Erlbaum Press.
- Padulo, L., & Arbib, M. (1974). *System theory*. Saunders Company.
- Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word. *Nature*, 423(6941), 752-756.
- Pfanzagl, J. (1968). *Theory of measurement*. Wiley.

- Ramsay, J., & Silverman, B. (2006). *Functional data analysis*. Springer-Verlag, New York.
- Restle, F. (1961). *Psychology of judgement and choice*. New York: Wiley.
- Roberts, L. D., Patterson, D., Thomson, J. O., & Levey, R. P. (1969, Mar). Solid-state and nuclear results from a measurement of the pressure dependence of the energy of the resonance gamma ray of *au197*. *Phys. Rev.*, *179*(3), 656–662.
- Rumelhart, D. E., & McClelland, J. L. (1987). *Parallel distributed processing*. The MIT Press.
- Schonemann, P. H., Dorcey, T., & Kienapple, K. (1985). Subadditive concatenation in dissimilarity judgements. *Perception & Psychophysics*, *38*, 1-17.
- Schouten, J. (1951). *Tensor analysis for physicists*. Oxford: Clarendon Press.
- Stevens, S. (1951). *Mathematics, Measurement and Psychophysics* (S. Stevens, Ed.). New York: J. Wiley.
- Suppes, P., & Zinnes, J. (1963). Basic measurement theory. In R. D. RD Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Stanford, California,.
- Thomas, R. D. (1999). Assessing sensitivity in a multidimensional space: Some problems and a definition of a general d' . *Psychonomic Bulletin & Review*, *6*, 224-238.

- Thomas, R. D. (2003). Further consideration of a general d' in multidimensional space. *Journal of Mathematical Psychology*, 9, 40-50.
- Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 9, 40-50.
- Townsend, J. T. (2008). Mathematical psychology: Prospects for the 21st century: A guest editorial. *Journal of mathematical psychology*, 52, 269-280.
- Townsend, J. T., Aisbett, J., & Busemeyer, J. B. (2005). General recognition theory and methodology for dimensional independence on simple cognitive manifold. In H. Colonius & E. N. Dzhafarov (Eds.), *Measurement and representation of sensations: Recent progress in psychophysical theory*. Washington, D.C.: American Psychological Association.
- Townsend, J. T., & Ashby, F. G. (1982). An experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 834-864.
- Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: the misconception misconceived. *Psychological Bulletin*, 96(2), 394-401.
- Townsend, J. T., & Kadlec, H. (1990). Mathematics and psychology. In R. Mickens (Ed.), *Mathematics and science*. Singapore: World Scientific Publishing Co.
- Townsend, J. T., & Landon, D. E. (1982). An experimental and theoretical inves-

- tigation of the constant ratio rule and other models of visual letter recognition. *Journal of Mathematical Psychology*, 25, 119-163.
- Townsend, J. T., & Landon, D. E. (1983). Mathematical models of recognition and confusion in psychology. *International Journal of Mathematical Social Sciences*, 4, 25-71.
- Townsend, J. T., Solomon, B., Wenger, M. J., & Spencer-Smith, J. (2001). The perfect gestalt: Infinite dimensional riemannian face spaces and other aspects of face cognition. In J. T. Townsend & M. J. Wenger (Eds.), *Computational, geometric and process issues in facial cognition: Progress and challenges*. Erlbaum Press.
- Townsend, J. T., & Spencer-Smith, J. (2004). Two kinds of global perceptual separability and curvature. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition*. Mahwah, NJ: Erlbaum.
- Townsend, J. T., & Thomas, R. D. (1993). Foundations of perceptual theory. In S. C. Masin (Ed.), (chap. On the need for a general quantitative theory of pattern similarity). Elsevier Science Publishers.
- Wenger, M. J., & Ingvalson, E. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 1106-1118.