CHAPTER 6. Representing Concepts by Fuzzy Sets

6.1 Introduction

The concept of a fuzzy set, first proposed by Zadeh (1965), enjoyed an initial burst of enthusiasm in psychology. This enthusiasm waned after receiving criticism from Osherson and Smith (1981). Fuzzy set theory does not provide a descriptive model for how humans reason about concepts, which is unsurprising given that it is a mathematical theory, not a substantive one. These objections have been examined elsewhere in this volume (Chapters 1 and 5). We will leave these issues aside and will instead consider only whether fuzzy set theory is useful as a mathematical tool. That is, we primarily focus on preconditions for using fuzzy set theory as such. For a discussion of psychological models built on fuzzy set theory, we refer the reader to Smithson and Oden (1999).

The problem fuzzy sets are designed to handle involves the issue of *degree-vagueness* (Ackerman 1994). By this we mean a concept lacking a firm boundary between objects with or without the property. The problem is an old one, with the most notable example being the well-known Sorites Paradox. There are many variants, but the essential component is that membership of a given set does not have a sharp demarcation. Often this is due to perceptual limitation, when the units making up the whole are very small parts. For instance, a heap of sand is made up of many grains. Removal of one grain does not change its “heap-ness”, but if one continues this process, there will be a point at which point any observer would agree that a heap is no longer present. In other cases, degree-vagueness occurs not because of our inability to discriminate the state of the world, but because we cannot discriminate at the level of consequences. For instance, the addition of an extra student to a class is certainly discriminable to the instructor but hardly changes the membership of the class in terms of its perceived size. However, the addition of multiple students to the class definitely changes its perceived size. To speak of an application of fuzzy set theory, both categorical endpoints and gradation between these endpoints are essential. It is important that both properties be present. Many constructs of interest to behavioral scientists can be thought of as being subject to degree-vagueness between categorical endpoints. Here are some examples:

- Brehm and Gates (1993), in a study of principal-agent relations in a bureaucracy, considered the degree to which an individual agent (in their case a police officer) was working or shirking, i.e., engaged in behavior consistent with the principal-agent relationship or defecting from it. The data used in this study were part of a larger study by Reiss (1971). This proportion, as a value between 0 and 1, was judged by an observer assigned to the officer. It makes little sense to consider this proportion as a probability in the conventional sense. This proportion is better interpreted as a meaningful position between ‘fully shirking’ and ‘fully working’.

Brehm and Gates model this proportion by using regressors that predict compliance or defection, allowing them to test different game theoretic accounts of principal- agent behavior.

- The United Nations Development Programme's Human Development Index (HDI) scores nations on the general concept of development, as defined by combining three key variables: Wealth, health and literacy (UNDP 2009). The HDI was not constructed specifically as a fuzzy set but can be thought of in those terms (Verkuilen 2005; Smithson and Verkuilen 2006).
- Participants in cognitive psychology experiments frequently make choices between two objects in a two alternative forced choice paradigm. Often they are asked in addition to rate their confidence in their decision, usually on a scale that has well-defined endpoints ranging from "not at all confident" to "totally confident." Confidence ratings are often used to provide additional information about discriminable stimuli and could be the object of study themselves, as in Verkuilen and Smithson (in press). Other so-called graphical ratings are used in similar contexts, e.g., preference studies. One successful model for these data is the fuzzy logical model of perception (FLMP), which has been tested in great detail by Massaro and colleagues (e.g., Friedman and Massaro 1998).
- The concept of physical maturity summarizes the ensemble of specific

measurements that could be made about a child, regarding bone development, secondary sex characteristics, dental development, etc., all of which start in an immature state and eventually develop to a mature state, although at different paces. However, this concept ranges from “not mature” to “mature” in degrees. Healy and Goldstein (1976) showed how to generate a scoring system that maps measurements to a maturity scalein an optimal way. While they did not directly refer to a fuzzy set interpretation, interpreting their optimal maturity scale in such a light is quite natural

- Conventional psychiatric diagnosis schemes (e.g. the DSM-IV) can be seen as crisp set membership assignment functions that assign a 1 if a person fits a certain criterion or 0 if they do not. However, this does not sit well with the nuances of clinical practice, and ignores an intuition shared by experts and lay- men alike that it makes sense to entertain the possibility that a person is “more depressed” than some other person. A better perspective might be to consider psychiatric disorders as fuzzy, that is, to devise membership functions that as- sign values to individuals to represent a degree of depression, with meaningful endpoints (“Fully depressed” to “Not depressed”). This example will be considered in more detail later.

All of these examples illustrate continuous variation between qualitative poles of full membership or non-membership. In each case a single number represents the state of a subject at a given point in time. For instance, a child scored according to physical maturity will start at 0 (not mature) and end up at 1 (fully mature). Determining this value depends on combining the values of the various physical measurements under question by means of some procedure. Other times, subjective judgments of the subjects or of external judges are involved, as is the case for the confidence rating and police examples given above, respectively.

As Bilgiç and Türkşen (1999) note, getting things to work in practice requires facing up to two related but distinct issues. The first is that of *membership assignment*. Is it possible to provide meaningful values of membership for objects in a fuzzy set? This question is of course dependent on the purpose to which which the membership values are meant to be put. Obtaining meaningful, or at least useful, membership functions is essentially a unidimensional scaling problem. In this case, the disciplines of psychometrics and mathematical psychology offer potential solutions to the problem of assigning membership, and also provide insight into limitations on what can be expected. Presuming it is possible to assign meaningful fuzzy set values can be accomplished, the second problem relates to comparisons *between* sets, which is known as *property ranking*. Claims such as "object x is more in fuzzy set A than it is in fuzzy set B" require that we can make meaningful comparisons between the memberships in A

and B. The property ranking issue is necessary for handling many uses to which fuzzy sets might be put. It is also somewhat of an Achilles' heel, because it is difficult to establish empirically meaningful relationships between sets unless the membership of objects in the sets can be compared.

To illustrate our points, we pursue two concepts that have broad implications throughout psychology. The first is the concept of utility (Shafir and LeBoeuf 2002) seen through the lens of fuzzy set theory. Models for choice behavior based on utility (or related concepts) have a long and voluminous tradition in psychology and economics, and are widely used across the behavioral sciences. Utility theory has been extensively formalized. For this reason, utility represents a useful case in which findings are quite clear about what is and is not likely to be possible for the membership function. The second concept we examine is that of depression "Major Depressive Disorder, or MDD". This is a widely experienced psychopathological phenomenon for which a strong definition does not yet exist and for which, it is becoming clear, standard statistical measurement models do not seem to be particularly appropriate. Fuzzy set-type reasoning may be able to shed light on this important problem.

6.2 Membership Assignment

To make use of fuzzy set theory, it is necessary to address some intertwined but nevertheless distinct issues. We organize these into three basic parts. First is the *interpretation of membership* and the purpose for which the scale will be used. What does partial membership in a set mean? There is no unique answer to this question, as it depends strongly on study context. The second involves the *measurement level* question. At what level of measurement do the membership values lie? That is, what sort of numerical properties can we consider a membership level to have? While these issues are distinct, decisions about one have implications for the other. Different interpretations are associated with different scaling procedures, that in turn result in different levels of measurement. Conversely, which measurement level is necessary depends on the use to which the memberships will be put, and what computations are to be performed on them. Third and finally, several methods of *assigning membership* will be considered. The interpretation of membership and the associated measurement level will be discussed for several examples of these assignment methods. More discussion of membership interpretation and assignment can be found in (Bilgiç and Türkçen 1999; Verkuilen 2005; Smithson and Verkuilen 2006).

6.2.1 Interpretation of Membership

To determine what an appropriate measurement of membership is, it is important to consider the interpretation of membership the investigator intends. Here we lay out the different ways that have been proposed in the past, though it should be noted that there might be others. Five different views of membership have been identified. These views were neatly exemplified by Bilgiç and Türkşen (1999 p. 196, with slight edits to make notation consistent):

The vague predicate "John (*x*) is tall (*T*)" is represented by a number *t* in the unit interval [0, 1]. There are several possible answers to the question "What does it mean to say $t = 0.7$?":

1. Likelihood view: 70% of a given population agreed with the statement that John is tall.
2. Random set view: when asked to provide an interval in height that corresponds to "tall", 70% of a given population provided an interval that included John's height in centimeters.
3. Similarity view: John's height is away from the prototypical object which is truly "tall" to the degree 0.3 (a normalized distance).
4. Utility view: 0.7 is the utility of asserting that John is tall.
5. Measurement view: When compared to others, John is taller than some and this fact can be encoded as 0.7 on some scale.

The first two interpretations have probabilistic underpinnings. The likelihood view essentially asserts that fuzziness equals probability in the sense that the membership here is a sample proportion from a given population. Under this interpretation, there is no vagueness within a single judge, just across judges, with probabilistic uncertainty appearing due to sampling variability. The random set view of fuzzy set theory identifies level sets as the crucial quantity. It focuses on the interval of the support of the fuzzy set, identifying membership with level-cuts of a particular width of the support. The level-cut defines the proportion of the population that asserts that the given object falls into the given interval. These two concepts of fuzziness seem rather distant from likely theoretical applications in the behavioral sciences.

The next three views are much closer in spirit to concepts in the psychological literature. The similarity view identifies membership of 1 with a prototypical object in the set of objects under consideration and considers membership as a normalized distance from this ideal. (This use of fuzzy set theory in psychology was proposed by Eleanor Rosch, who discusses her contribution in Chapter 4 of this volume.) For instance, given object x, d_x a distance of x from the ideal object in set T, and f some non-negative, usually weakly monotone, function, membership might be assigned as:

$$t_x = \frac{1}{1 + f(d_x)}.$$

There is a long tradition in psychological research regarding similarity-based theories, and it is expected that the similarity view will be the most natural to most psychologists. Although the problem of defining an appropriate distance is by no means a trivial one, it has been the subject of voluminous and sophisticated research (Takane and Shibayama 1992, and references therein). The utility view considers membership to be the utility of asserting that the object is in a given set. In a sense it is a variant of the similarity to prototype view, if one is willing to assume there exists at least one object for which this utility would be maximal and another for which this utility would be minimal. (The originally proposed utility view made some rather extreme claims, e.g., regarding the lack of context-specificity of membership, but we believe that there is no reason to maintain these specific claims.) The measurement view considers membership through the lens of representational measurement theory, that formally describes if, and to what extent, relations between numbers (scales) can be used to represent certain qualitative relations between empirical objects. From the measurement perspective, degree of membership is considered a property with specific measurement characteristics. These characteristics are important when statistical operations are performed on the membership values in order to conclude something about this underlying property. The measurement can be considered more of a methodological approach than an independent view on the ontological status of membership values.

The different views described above are not mutually exclusive. For instance, elicited binary judgments to similarity to prototype may be used to define the empirical relations which are in turn considered by way of representational measurement axioms. As it stands, this typology is an undifferentiated "laundry list." A better typology of interpretations of the membership function with formal backing would be a welcome addition to the literature.

6.2.2 The Measurement Level Question

It is important to be clear about the *measurement level* associated with membership in a fuzzy set. By this we mean, what properties of numbers do membership functions possess? Considering the measurement level of membership forces us to think about membership structure and assignment. Also, when the purpose is to draw conclusions about membership based on statistical operations, the level of measurement determines whether conclusions will depend on the arbitrarily chosen scale or not. For example, strictly speaking, operations such as the arithmetic mean require numerical structure at the level of an interval scale, whereas the median or min-max operators require only ordinal measurement To appreciate why measurement level is important in this respect, the development of measurement theory will be discussed below. For a more extensive overview see (Michell 1999) or (Hand 2004).

Theories of Measurement

During the first part of the 20th century, the status of psychological measurement was challenged by physicists like Campbell (1920). It was thought that true measurement requires the demonstration of additive structure in the property of interest. The term *fundamental measurement* was coined to refer to measurement of properties that show additivity directly (length, mass and elapsed time) or by derivation of directly additive properties (density and electrical resistance). There are three basic types of scales obeying these axioms. The *absolute scale* has no degree of freedom for transformation. An example would be a count or normalized count such as a probability. A *ratio scale* such as mass, distance or elapsed time from a defined starting point has only an arbitrary unit (e.g., meters or feet for distance are both valid units) but has a well-defined zero. In this case both ratios and differences of a ratio scale are meaningful. An *interval scale* is one that allows for transformation of a unit and also translation of the origin. Common examples include Celsius temperature or calendar time. Only differences of interval scales are meaningful. The *log-interval scale* is a variant of the interval scale on which ratios are meaningful but not differences. Examples include measures such as fuel usage as measured in miles per gallon. *Ordinal measurement*, where only ordering relations are meaningful, does not satisfy the additivity requirements for fundamental measurement.

Obviously few psychological variables are amenable to fundamental measurement. One of the few examples is the von Neumann-Morgenstern utility function. Even when physical measurements such as reaction times or evoked potential are available, it is unclear how these relate to psychological properties such as mental capacity or emotion, thought to underlie them. Townsend (1992) discusses this issue in the context of reaction time, which is commonly used in cognitive psychological studies to determine properties of underlying mental representation of tasks. While reaction time is on a ratio scale, there are a number of subtleties involved in maintaining this interpretation. However, as Townsend notes, the cost of losing the ratio measurement properties of reaction time destroys most of the interesting empirical results derived from studying it.

Stevens (1946) proposed that the definition of measurement should follow the empirical structure of properties. Measurement was redefined as the assignment of numbers to objects according to a rule. The rule corresponds to the type of relation between objects that is captured, determining the level of measurement. If the rule is 'assign different numbers to distinguish different types of objects,' representing only category information, we can speak of measurement at the nominal level. The ordinal measurement level entails ordering of the objects on the property of interest, the interval level entails the comparability of differences and the ratio level entails the direct comparability of quantities.

Associated with these levels are types of transformations that leave intact the type of information captured by the data. A statistic, or the inference based on it, should remain invariant under permissible transformations that represent the underlying property equally well. Strictly speaking, statistics that make use of interval information, such as parametric *t*- and *F*-tests should therefore not be used with ordinal data. The proposed permissibility of statistics evoked a fierce debate (Lord 1953; Gaito 1980; Townsend and Ashby 1984; Zand Scholten and Borsboom 2009). This debate was never truly settled, in part because Stevens' theory was subsumed into a more sophisticated theory, which emphasizes the legitimacy of inferences, not statistics.

This theory, known as representational measurement theory, axiomatizes whether consistent assignment of numbers to objects is possible for some property, and what the associated measurement level is (Krantz, Luce, Suppes and Tversky 1971). In order to claim a measure is of the interval level, the objects under study are required to obey a number of axioms. Additive Conjoint Measurement (ACM) is a highly relevant representational structure for psychological properties, because it allows demonstration of additivity in an indirect fashion. The most important ACM axiom is *double cancellation*, a complex of simultaneous order restrictions, which ensures transitivity requirements hold. Not many psychological variables will obey such axioms perfectly, although they may do so approximately.

One option is to acquiesce, accept that the ordinal level is the best we can do for now and focus on methods that ensure unambiguous inference. As noted above with regards to reaction time, this often requires giving up too much of what can be studied empirically. Another approach is to adopt a pragmatic definition of measurement, one which entails a sizable amount of interpretation.

Measurement by fiat refers to numerical assignment based primarily on theory and intuition of researchers, not on tested axioms. The term was coined by Torgerson (1958), who rightly noted that there is nothing inherently wrong with measurement by fiat, particularly in the early days of inquiry. For the behavioral sciences, measurement by fiat is the only workable definition in the foreseeable future. In nearly any practical measurement task, there will be some combination of representational and pragmatic aspects (Hand 2004). One hopes that, as a field of inquiry develops, better validated measures are developed.

What Level Is Needed?

What measurement level is needed depends on the use to which one wants to put the membership values. If categorization into a trichotomy—not a member, partial member, full member—is all that is needed, the ordinal level of measurement is sufficient. Furthermore, as discussed by Smithson (1987) and Smithson and Verkuilen (2006), much of fuzzy set theory is essentially ordinal, so it is possible to get a lot done with just ordinal membership assignment. As mentioned before, the min-max operators only rely on ordinal comparisons. Even then, inter-set relations still depend on property ranking. For instance, Smithson (2005) shows it is possible to develop an ordinal measure of fuzzy inclusion.

Often, however, a higher measurement level is needed. For instance, anyone intending to use fuzzy set modifiers based on power transformations such as

Zadeh's linguistic hedges needs numerical membership values. Exactly what level of measurement required is somewhat unclear. Wakker (2008) explores the measurement properties of power transformations in great detail. He shows that power transformations require interval measurement, but higher levels have been claimed. More generally, if the membership values are to be subjected to statistical operations that make use of quantitative properties, then an interval level is needed. Measures such as fuzzy cardinality require even higher measurement. Although only very few fuzzy sets will hold up under the harsh scrutiny of axiomatic measurement theory, this does not mean the entire enterprise is hopeless. There are indeed examples of fuzzy set membership assignment methods that can be argued to show quantitative structure.

6.2.3 Assignment Methods

As one might expect, there are many ways of obtaining membership values. These methods have their strengths and weaknesses.

Direct Scaling

A number of direct scaling procedures have been developed over the years. These rely on a judge (broadly defined) providing a direct assignment of a membership value to an object. What can be done with ratings so acquired varies greatly, ranging from taking them at face value to subjecting them to verification of the axioms of representational measurement theory. A substantial number of direct scaling methods exist (Bilgiç and Türkşen 1999). All are variants of well-known methods of data gathering used in the behavioral sciences. *Polling* interprets sample proportions gathered from many judges as membership. In the terminology of scaling theory this is referred to as single stimulus response. One weakness is the fact that this strategy is unlikely to be workable when one wishes to consider an individual judge's membership values. *Direct rating* elicits membership values from one or more judges by presenting objects to them and gathering some sort of numerical response, possibly subject to subsequent transformation. Graded response formats can, in principle, provide more information than binary ones, especially about discriminable objects, but at the cost of stronger assumptions and more potential for inter-judge heterogeneity. Polling and direct rating are simply variants of each other, with polling providing judges with a binary response format and the gradation for membership being derived from aggregates of these binary responses, whereas direct rating elicits this gradation immediately. Two related strategies are mentioned by Bilgiç and Türkşen (1999). *Reverse rating* reverses the process of direct rating by having judges assign objects to given values of membership. The method of *interval estimation* naturally matches the random set view, by asking a judge to assign objects to intervals of the support set that match a given membership level. A generally useful reference on direct scaling is (Torgerson 1958).

Indirect Scaling

A number of scaling models have been developed that could be used to build membership functions up from more elemental responses provided by one or more judges. We consider one of the simplest and best studied examples, the von Neumann-Morgenstern utility function, in some detail. This provides a scaling procedure that can generate interval scales based on choices made over a sequence of gambles (Luce and Raiffa 1957). Much like Additive Conjoint Measurement, the von Neumann-Morgenstern procedure comes down to testing crucial assumptions about transitivity. Consider a hypothetical decision maker, Rose, who has to decide between three alternative vacations, each lasting a week: Stay home (S), trip to Paris (P), and trip to London (L). Assume further that it has been established for Rose that she prefers Paris to London, Paris to staying home and London to staying home. Without loss of generality, we may assign $u(S) = 0$ and $u(P) = 1$. By providing Rose gambles SpL, SqP and LrP, where XyZ denotes "receiving option X with probability y, else receiving option Z" it is possible to assign a numeric utility for the trip to London such that $0 < u(L) < 1$, assuming her choices are transitive. (The procedure can, of course, be generalized to more than three objects.) It should be noted that a given decision maker's choices may not satisfy the von Neumann-Morgenstern axioms and thus not be scalable. This would happen if transitivity conditions were violated, for instance if she preferred Paris to London, London to staying home, but staying home to

Paris. To assess membership this way, we identify utility with membership. Researchers in behavioral decision making have found a number of systematic anomalies to von Neumann-Morgenstern utility theory, including a relatively greater aversion to loss than preference for gain, but it still represents a useful first approximation and scaling method.

A number of other scaling models provide similar insight. For instance, the well- known paired comparisons procedures allow the generation of an interval scale of utility, assuming that choices made by judges satisfy stochastic transitivity conditions, which are probabilistic generalizations of the von Neumann-Morgenstern transitivity conditions. Böckenholt (2006) presents a substantial review of the probabilistic utility scaling literature, which he connects to the behavioral decision making literature. Another approach can be found in the well-known Rasch model. The Rasch model is an item response theory model that relates latent ability to the probability of answering an item correctly via a logistic function. The model can be used to estimate latent values based on ordinal item scores. The Rasch model can be seen as an instantiation of ACM (Perline, Wright and Wainer 1979) and thus provides a method for generating an interval scale. If data fit this model, then the underlying property can be said to show display? quantitative structure, at least according to some (Embretson and Reise 2000; Bond and Fox 2007). The Rasch model is very strict however, and in many cases only a small set of items will show adequate fit. Even if good fit is achieved, the latent ability scale in the model varies over the entire real line. There are bounded representational structures that address this problem, but the Rasch model as a form of ACM is not one of these. One way to establish relatively natural endpoints however, could be to assign a 0 to respondents who answered all items incorrectly, a 1 to respondents who answered all items correctly and to rescale the latent ability values of all other respondents accordingly. This sits well with the fact that these extreme scores are not used in the estimation of the model since they result in an infinitely small or large latent value when conditional maximum likelihood estimation is used.

Unfortunately, while an interval scale is useful for many purposes, it lacks the defined endpoints desired in fuzzy set theory. While the example above has $u(S) = 0$ and $u(P) = 1$, any other two numbers would do just as well because utility is an interval scale. Something extra is necessary to identify these anchor points. Often this can be done by considering other aspects of the choice set. For instance, option P may represent for this decision maker the ideal week's vacation, better than all other possibilities. In this case it makes sense to identify it as Rose's ideal point, which means that it becomes reasonable to assume $u(X) < 1$ for all other choices X less preferred than P, without loss of generality. Similarly, she could definitely want to take some trip, in which case $u(X) > 0$ for all other choices X preferred to S . A different person, Colin, might have a different preference ordering. Böckenholt (2004) showed how to identify, given some judicious addition of information and reasonable assumptions, the origin of the scale in a paired comparisons scaling context that provides natural statistical tests for consistency. He provides a number of examples and code to fit his models. Lootsma (1999) also discusses using reasonable anchor points such as the status quo policy or an ideal policy to help set the origin in a useful way.

However, a fair degree of confusion remains in the scaling literature. For example, one popular method of indirect scaling that has been employed in the fuzzy set literature is the Analytic Hierarchy Process (Saaty, 1977). AHP is itself a reinvention of techniques found in the scaling literature of the 1950s (Torgerson, 1958) and is related to magnitude ("ratio") scaling common in psychophysics. Ratio scaling of the psychophysical variety should not be thought of as generating a ratio scale at all. That is, ratio scaling does not yield a scale that is invariant up to multiplication by a constant and for which both ratios and differences are meaningful. Birnbaum has shown that ratio judgments are not consistent in the way that would be required for them to provide a ratio scale. See (Hardin and Birnbaum 1990) for a summary of this line of research. Instead, ratio scaling works on log-interval scales, which are simply exponentiated interval scales for which there should be no expectation that both differences and ratios be consistent as there would be for a true ratio scale. 'However'?Lootsma (1999) criticizes the analytic hierarchy process (AHP), which has been widely applied in fuzzy set applications on similar grounds. Another example of this confusion involves the Fuzzy Logic Model of Perception (e.g., Friedman and Massaro 1998), which models binary choice and rating data using a transformation of membership functions. Crowther et al. (1995) show that it is equivalent to the Rasch model.

Transformation/Measurement by Fiat

Frequently, values for an existing variable can be obtained, and the investigator has some idea—based on theory, face validity or intuition—as to how this existing variable aligns with the concept embodied by the fuzzy set. For instance, a pre-existing interval scale with identifiable endpoints can, without loss of generality, be linearly transformed into the unit interval. Also, other nonlinear transformations might be chosen to represent the sorts of qualitative features believed to underlie the concept of membership.

Hybrid Strategies

In practice it is often necessary to adopt a hybrid of the above methods. For instance, values generated from indirect scaling may satisfy the requirements of an interval scale but measurement by fiat may be necessary to set the maximal and minimal values. This was exactly what was done to set the endpoints of the membership function in the utility scaling example given previously.

Validation

Validation of membership assignment is a necessary but difficult task. Scaling models and axiomatic measurement theory provide routes for verifying internal properties of scales. For instance, the double cancellation axiom and other axioms of conjoint measurement verify that an additive representation for the given empirical system is possible and a number of different models, e.g., the Rasch model, obey these axioms. It is important to note that verification of measurement level assumptions via Rasch modeling requires an acceptable model fit. Such fit is often achieved only after step-wise exclusion of items that show misfit. This procedure can lead to an artificial unidimensional scale. Validation therefore requires replication of results using identical but also different tests. However, internal verification is not the only task one might want to undertake. A great deal of the power of fuzzy sets comes from the use of connectives such as intersection or union or operations such as negation, which allow sets to be derived from more basic sets. However, it is not clear a priori whether, for fuzzy sets A, B if derived and directly obtained membership agree. In this case it seems quite reasonable to elicit membership values a, b, $a \cap b$, $a \cup b$, etc., over a set of objects X and see whether these agree with those derived from the chosen connectives and operations. Research in behavioral decision making suggests that it what? is unlikely to hold, given the well-documented inability of judges to correctly guess probabilities of compound events (e.g., Wolfe and Reyna 2010). This is clearly an area for further research.

6.3 Property Ranking

We have already discussed this point to some degree. To continue our exploration of insights from utility theory we examine the problem of interpersonal comparison of utility. It is widely held that a decision maker may have a utility function, assuming her decisions are sufficiently consistent to satisfy the von Neumann-Morgenstern axioms, and probabilistic utility theories have relaxed the requirements to allow a certain level of inconsistency likely to be inherent in real choices. In a sense it seems that interpersonal comparison of utility ought to be possible—a pithy quote illustrating this taken from (Sen 1970, p. 199) is "it should not be difficult to say that Emperor Nero's gain from burning Rome did not outweigh the loss of the rest of the Romans"—but formalizing this comparison has proven to be elusive.

Luce and Raiffa (1957, Chapter 6) provide a thought-provoking example of the property ranking problem in the context of a two-person game first proposed by philosopher R. B. Braithwaite. The "story" of this game involves two neighbors, our hypothetical decision makers, Rose and Colin, who live in adjoining apartments and have to allocate time playing musical instruments. Braithwaite asked whether it would be possible to generate a fair allocation of time, taking into account the strength of preference. Each player can choose to play (P) or not play (NP). Utility functions are obtained for both and are normalized by assigning a value of 0 to each player's least preferred policy and 1 to each player's most preferred policy, permissible for an interval scale. That is, the best option is assigned one utile, the worst option zero utiles and all others fractionally between. These utiles are player-specific. The purpose is to determine if it is possible to equate them across players. The resulting normal form game with normalized utilities for each player is given here. For instance, if both Rose and Colin are playing, she receives 0 utiles and he receives 0.11 utiles.

	Colin Plays	**Colin Doesn't Play**
Rose Plays	(0,0.11)	(1,0.22)
Rose Doesn't Play	(0.50,1)	(0.33,0)

Luce and Raiffa examine two related solutions to the interpersonal comparison of utility. The simpler of the two takes Nash's bargaining solution and works backwards to convert to units of the original measurement scales, while the other elaborates this procedure slightly. Doing this shows that Rose is indifferent between the arbitrated solution and a lottery which weights her most preferred alternative with probability 0.652 and her least preferred alternative with probability 0.348. Similarly Colin achieves 0.763 utiles. To achieve this result, Colin should play while Rose remains silent 16 out of 23 nights while Rose should play on the remainder. This method is shown to satisfy all axioms of the

Nash equilibrium except the independence of irrelevant alternatives (IIA), which states that the addition of objects to the choice set should not change relationships among existing objects. Unfortunately, the normalization employed above depends on the objects given here, even if the choices added are dominated or otherwise irrelevant alternatives. While the set of alternatives might be able to be pruned only to the set of meaningful choices, in general we cannot assume so.

This lines up with prior empirical research in utility theory. For example, Wallsten et al. (1986) provide a thorough, systematic study that aimed to formalize common linguistic quantifiers. The goal of their study was to determine if subjects (economics and social science graduate students at a major research university) could provide membership functions for probability words such as "doubtful," "probable" or "likely." Membership assignment was done by way of a conjoint measurement method based on paired comparisons of the degree to which each word applied to a pair of gambles presented. Membership values were assigned as estimated parameters from the scaling models based on axioms of ratio measurement. These models were shown to have acceptable fit. In particular, they showed that subjects appear to be internally consistent in the sense that their understanding of the probability words satisfactorily predicted relationships among different words predicted by the measurement model (up to sampling variability), which amount to satisfying ordering conditions among the comparisons. However, membership functions between subjects were not very consistent. Similarly, subjects in experiment performed by Böckenholt (2005) seem to be self-consistently additive, but there appear to be non-negligible individual differences, very much in line with those found by Wallsten et al. More advanced procedures such as have been explored in the literature on probabilistic scaling with hierarchical nonlinear models might improve this situation, at least in the sense of being able to better quantify the amount of between-subject heterogeneity; see (Verkuilen and Smithson in press) for examples.

How does this relate to fuzzy set theory? The solution to interpersonal comparison of utility (or lack thereof) is essentially the same problem as property ranking. A number of proposed fuzzy set based techniques are essentially premised on property ranking or consequences of it. For instance, the measurement of fuzzy inclusion proposed in (Ragin 2000) strongly depends on the margins of the distribution of the two fuzzy sets being compared. This in turn depends on comparability across sets, i.e., property ranking. Sometimes this seems plausible given the content of the sets to be compared, but by no means is it always so, and depends strongly on the membership assignment in a manner that a different measure of inclusion would not. For instance, Ragin's proposed procedure will declare two stochastically independent variables to be fuzzy subsets if the margins are sufficiently skewed.

6.4 Example: Major Depressive Disorder

We now consider an example in some detail. Here, we show how fuzzy set theory can be of relevance for the assessment of psychopathological disorders such as depression. Most formalizations of psychiatric diagnosis schemes have an implicit method of membership assignment that does not concur with the subtlety of actual diagnosis by practitioners. Take for instance an archetypical psychopathological disorder, major depressive disorder (MDD). Conventional clinical assessment of MDD, following the guidelines of DSM-IV, takes on the following form. There are 9 major symptoms of clinical depression; depressed mood, diminished interest and pleasure in daily activities, weight problems (loss or gain), psychomotor agitation or retardation nearly every day, sleep problems (insomnia or hypersomnia), fatigue or loss of energy, concentration problems, feelings of worthlessness and suicidal ideation. The diagnosis of MDD according to the American Psychiatric Association (DSM-IV-TR) criteria can be considered the assignment of people to the crisp set depressed (1) or not depressed (0) on the basis of the following membership assignment: if a person displays 5 or more of these symptoms for a period of at least two weeks, he or she is considered to be depressed, i.e., to display MDD. This membership assignment scheme (not referred to as such in the DSM) is the topic of much discussion: although most clinicians agree that these 9 symptoms are of central importance, the exact criterion (5 or more during 2 weeks) immediately leads to numerous examples that defy intuition and clinical practice: people who either display a maximum of four symptoms for any period of time (even if it includes suicidal ideation) will not, according to these criteria, be considered depressed. Also, as all symptoms are created equal, we get an implicit rank-ordering of patients that does not align well with decisions of clinical practice or common sense. Consider the following example: Person A suffers from the following four, intuitively severe, MDD symptoms: A depressed mood, diminished interest and pleasure in daily activities, feelings of worthlessness and suicidal ideation. Now consider person B, who suffers from the following five symptoms: weight problems, psychomotor agitation, fatigue, sleep problems and concentration problems. Strictly speaking, not only is person B more depressed, person A should not be considered depressed at all. Although we do not in any way imply that the latter symptoms are trivial or irrelevant, it seems a stretch to argue that all MDD symptoms are equal in terms of psychological impact. If this is something clinicians and academic psychologists would agree on, it is worth attempting to explicate and formalize this fact.

Several recent developments (e.g., Borsboom 2008; Cramer et al. 2010; Zachar and Kendler 2007) have focused on shortcomings of this classical approach and have considered dynamical systems or causal modeling approach instead. Others have extended the conventional assessment of depression by means of IRT analysis; for an overview, see (Reise and Waller 2009). It seems clear that the classic assessment scheme fails to line up with clinical practice and suffers from serious conceptual and psychometric problems. We propose that the application of fuzzy set theory to the psychological concept of depression provides a novel approach to the problem of MDD diagnosis, and illustrates the benefits and flexibility of fuzzy set theoretical applications.

For the purpose of this illustration, we will focus on a Major Depressive Disorder (MDD) and the symptoms to focus on the details of membership functions. The proposed framework can be extended to include larger networks of symptoms and disorders. The core is in line with conventional fuzzy set theory: instead of considering an individual depressed or not depressed, we will consider any individual to occupy a position on a fuzzy set scale of depression that ranges from 0 (no depression symptoms) to 1 (all depression symptoms). Note that 0 does not necessarily imply happiness, merely the absence of depression symptoms. To align with the requirement of duration, we will assume the symptoms as considered below are present for at least two weeks.

For the concept of depression, the fuzzy set membership values between 0 and 1 should be considered conceptually meaningful, as opposed to applications that consider fuzzy sets a quantification of uncertainty or measurement error. That is, a person who is assigned a value of 0.4 on a fuzzy set depression scale is in a certain psychologically and clinically meaningful state, that may yield predictive information about future psychopathology such as the extent to which certain types of therapy or intervention by means of medication will be successful.

In our example below, we will consider a hypothetical population of 10,000 members that display a representative selection of the MDD symptoms as described above. In the appendix online we provide R code to simulate a representative population of subjects with realistic symptom constellations, and to implement three of the described membership functions. The probabilities of symptom patterns are based on (Aggen et al. 2005), but may be adapted based on relevant empirical data. A given person may display any constellation of these nine symptoms, although they are not equally likely (a person who 'only' suffers from sleep problems is much more likely than a person who 'only' suffers from suicidal ideation, but none of the other symptoms). The following membership function implementations are examples, and may be applied to any dataset. In line with Smithson and Verkuilen (2006) and Verkuilen (2005), we will consider various ways in which we can assign membership values to individuals in this population, based on various membership function procedures. Conceptually, an individual can take on a value between 0 and 1, where 1 represents fully/very depressed, and 0 represents not depressed at all. We can translate the DSM-IV MDD diagnosis as the following (crisp) membership function: If a person displays 5 or more symptoms during two weeks, then a person is depressed (1), otherwise he/she is not depressed (0). For reasons mentioned above, this leads to undesirable consequences. We will now examine various perspectives that move away from the traditional, crisp set perspective to possible fuzzy membership functions that yield values between 0 and 1, based on differing membership functions.

6.4.1 Direct Assignment

This is a simple and traditional implementation of fuzzy set theory. A group of expert clinicians rates a person on a depression scale between 0 and 1, based on, for instance a diagnostic interview or the subjective assessment of the severity of symptoms. A severely depressed person may be assigned a 0.9, a person who displays some symptoms of depression may be assigned a 0.2, and a (control) person with no symptoms has a 0. Given these assignments and the symptoms people display (already known), we can use simple regression to deduce a weighting scheme for symptoms, such that the membership functions assignment corresponds most closely to the subjective ratings of the expert clinicians. The benefit of the estimates so derived is that they reflect the *implicit* weighting scheme clinicians apply. Secondly, it gives important diagnostic information about the error terms of symptoms: the regression yields information about the amount of information about MDD the presence of a symptom contains. For instance, it seems quite possible for a non-depressed person to have trouble sleeping, whereas a non-depressed person with suicidal ideation seems, at least intuitively, less likely. If true, this will yield larger confidence intervals for beta weights, which may represent important diagnostic information.

6.4.2 Indirect Scaling

With this method, instead of weighing overall depression severity, experts can rate the relative severity of symptoms. Judging on the basis of criteria (such as the influence on daily life, the extent to which displaying a symptom is a mark of severe depression), experts such as psychiatrists can assign relative weights to the 9 symptoms of depression on the basis of theoretical considerations, making sure the symptom weights sum to 1. For instance, lack of sleep is presumably weighed less severely than suicidal ideation. Together, the nine symptoms represent the full spectrum from 0 (absence of symptoms) to 1 (all symptoms). However, unlike the traditional DSM classification, not all symptoms are created equal: If a person suffers from suicidal ideation and depressed mood, he or she will (presumably) be considered more depressed than a person who has problems sleeping and weight gains, although both persons display, or suffer from, two symptoms of depression. Doing so yields fuzzy set values between 0 and 1 for any subject depending on the number and severity of the symptoms they display.

6.4.3 Disorder Prevalence

One of the core benefits of fuzzy set theory is that is better aligns with the manner in which (linguistic) concepts are used and interpreted. In fact, it is possible to utilize such interpretations as the basis of a membership function. One way to accommodate intuitions about psychological constructs into membership functions is to ensure that the distribution of values over a populations fits in with our common sense perspective on the prevalence of the disorder (or psychological construct). For instance, whatever membership function we choose, if it yields membership values such that 80 percent of the population scores, say, higher than 0.9 (which would roughly align with the linguistic term very depressed), most would agree that this does not fit well with conventional intuition. This intuition can be formalized within the fuzzy set framework. It is possible to decide, by fiat, the approximate distribution of fuzzy set values for a population and deduce from this distribution the weights of symptoms in such a way that the desired distribution holds for the population under study. For instance, imagine that a group of experts decides that for a meaningful and useful interpretation of fuzzy set values, the following distributional characteristics should hold: 50 percent of a population should score 0 (be completely MDD symptom free), 30 percent scores 0.3 or higher, 10 percent scores 0.7 or higher, and 1 percent scores 0.9 or higher. This can be represented as an approximate density function of fuzzy set values for a population. Given the multivariate distribution of symptoms, we can assign weights to each of these symptoms in such a way that the fuzzy set values so generated conform to the distribution decided on by a panel of experts. In this way, the fuzzy set values generated will conform to a natural interpretation understanding of the scale between 0 and 1. This strategy therefore may allow us to formalize and quantify natural intuitions concerning the prevalence of disorder severity in a population.

6.4.4 Benefits

The above are some examples of possible membership functions for MDD that formalize the common intuition that people can 'be' in a psychological state somewhere between 'not depressed at all' to 'fully or severely depressed'. There are several reasons to consider fuzzy set theoretical applications for psychological concepts such as MDD. Classifying people as either depressed or not-depressed, with no shades of gray, does not align with clinical practice or common intuition, and limits the scientific study of the concept. Fuzzy set theory is much better in line with conventional thinking and therapeutic assessment, and its membership value criteria can be adjusted such as to best fit. Although practitioners rightly argue that the DSM does not reflect actual clinical practice, it is still the authoritative source for clinical diagnosis. If what practitioners actually do is more subtle (and this seems likely), it is surely worth attempting to formalize these practices in models that can accommodate more sophisticated strategies.

Fuzzy set theory also has several benefits over other diagnostic tools that have been applied to psychiatric disorders, such as Item Response Theory (IRT). See Reise and Waller (2009) for an excellent overview. Firstly, IRT models have a latent variable scale that runs from minus to plus infinity. Although such scaling procedures are sensible for ability scales, for concepts such as MDD this is less appropriate. For instance, MDD is characterized by 9 symptoms, that are either present or absent. This suggests (in addition to fuzzy boundaries), two natural endpoints: displaying no MDD symptoms (0) up to displaying all symptoms (1). This is less restrictive than one might suppose, as the manifest properties of the data do not identify the latent space in an IRT model, so it is possible to have a model that makes exactly the same predictions as the Rasch model that generates scale values in [0,1] (Rossi et al. 2002). Secondly, due to the nature of depression symptoms, IRT models of psychiatric disorders commonly encounter problems, such as the fact that the discrimination values of items are not distributed evenly over the ability scale, leading to problems in model fitting and item design. Reise and Waller note that common IRT models do not distinguish genuine absence of a trait from very low values.

Finally, fuzzy set membership has several interesting applications and extensions. We mention three such possibilities. For instance, by assigning fuzzy membership between 0 and 1, it is possible to assess fuzzy cardinality, i.e. to sum partial values of a construct over a population. Within the classical DSM framework, this is not possible. Take for instance two groups of individuals. In the first group, all individuals display 4 MDD symptoms over an extended period of time. In the other group, all members display no MDD symptoms at all. From a traditional perspective, they are considered 'equally depressed'. Having a formalized quantification of degrees of depression gives the mathematical tools to study phenomena that may be too subtle to be measurable by traditional means. Secondly, subtle temporal dynamics of psychiatric diagnoses such as seasonal affective disorder, the effect of (natural) disasters on populations or the increase in depression due to economic adversity may be studied by fuzzy cardinality. That is, despite the fact that subtle changes in a population occur at a sub-threshold level, it may be possible to study the change in (weighted) symptom distributions over time, and quantify such phenomena more accurately. Thirdly, fuzzy set membership functions may allow one to model an archetypical feature of psychiatric diagnoses: the fact that symptoms of, for instance, MDD rarely come alone. This phenomenon, of co-occuring symptoms and psychopathology is commonly referred to as comorbidity. Fuzzy set membership assignments offers potentially valuable diagnostic information for comorbidity. For instance, a person might be diagnosed as having an MDD based on traditional DSM criteria, but may also display several symptoms of other disorders at a sub-threshold level. These other, sub-threshold symptoms (such as symptoms from generalized anxiety disorder), would formally not be considered in the clinical assessment of a person. This is especially troubling giving the fact that comorbidity of different clinical disorders (or symptoms) is very prevalent (Cramer et al. 2010). The phenomenon of sub-threshold symptoms can be incorporated more naturally in fuzzy set theory, where the assessment of an individual can be made on the basis of fuzzy cardinality (i.e. the summation of partial memberships of various disorders). A weighted summation over the totality of symptoms may allow for more fine-grained assessment: a person who suffers from various symptoms that, in isolation, may not warrant a traditional diagnosis of, for instance, generalized anxiety disorder (GAD), can still be diagnosed based on appropriately weighted symptoms, that can be more accurately monitored over time than plain presence or absence of a disorder. Smithson and Verkuilen (2006) show some applications of fuzzy set theory for comorbidity, illustrating some commonalities between covariances and the sum and product operators for fuzzy union and intersection. A number of examples were provided, illustrating the comorbidity of dislike, disgust and fear, and also several psychiatric symptoms in two samples of children, those referred to clinic or not.

These are but some of the possible extensions of the current proposition, namely that fuzzy set theory allows for a tractable formalization of the intuition that psychopathology is not a clear and crisp affair, but a matter of degree, and should therefore be treated as such.

6.5 Discussion

Fuzzy set concepts enjoyed an initial surge of popularity in psychology in the 1970s, followed by a long period of skepticism. A more modern approach has been to avoid programmatic declarations and consider what problems could be solved using the concepts of fuzzy set theory. To appreciate this one first has to have an idea of what membership is and how membership to a degree is determined. Interpretations and methods can vary and are combined in many cases.

In this chapter we have focused less on relationships between fuzzy sets (although obviously those are important) and primarily on properties internal to the set. In particular we highlight the notion of degree-vagueness and recognizable endpoints as being essential in determining whether a fuzzy set based analysis is appropriate. We have focused on defining membership in a meaningful way. Three crucial elements in this respect are the interpretation of membership degree, the assignment method and the measurement level associated with this method and the property of interest.

Several different interpretations have been proposed; which makes sense in a given case will depend on the context of the problem. More than one interpretation can be appropriate in some cases. The similarity and utility interpretation of membership degree seem to provide the most appealing view for social scientists. Assignment methods can also vary greatly and can be used in combination or in sequence in some cases. Indirect scaling such as paired comparison seems to best connect with existing methods, but other methods offer interesting new approaches for applied researchers. The measurement level associated with membership assignment is hard to determine. Often the best we can do is the ordinal level. In some cases however, methods such as paired comparison or Rasch modeling can be employed to at least investigate whether there is some indication of quantitative structure to the degree of membership. More work needs to be done on lesser known representational structures for bounded empirical structures. Many of these structures are not amenable to standard measurement procedures, but could prove useful in a fuzzy set context.

The application of fuzzy set theory to preference rating and decision making using the von Neumann-Morgenstern axioms and the diagnosis of depression shows that fuzzy set theory can serve as an important tool to further our understanding of complex psychological properties and processes. Undoubtedly, more areas could benefit from the application of fuzzy set theory, especially areas where categorization problems are unsuccessfully approached with either continuous or categorical models, such as in the depression example.

References

Ackerman, F. (1994). Roots and consequences of vagueness. *Philosophical Perspectives*, 8: 129-136.

Aggen, S., Neale, M. and Kendler, K. 2005. DSM criteria for major depression: Evaluating symptoms patterns using latent-trait item response models. *Psychological Medicine* 35: 475-487.

Bilgiç, T. and Türkşen, I. B. 1999. Measurement of membership functions:

Theoretical and empirical work. In *Fundamentals of fuzzy sets*, ed. D. Dubois and H. Prade, 195-232. Boston and Dordrecht: Kluwer.

Böckenholt, U. 2004. Comparative judgements as an alternative to ratings: identifying the scale origin. *Psychological Methods* 9: 453-465.

Böckenholt, U. 2006. Thurstonian-based analysis: Past, present, and future utilities. *Psychometrika* 71: 615-629.

Bond, T. and Fox, C. 2007. *Applying the Rasch model: Fundamental Measurement in the Human Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Borsboom, D. 2008. Psychometric perspectives on diagnostic systems. *J. of Clinical Psychology* 64: 1089-1108.

Brehm, J. and Gates, S. 1993. Donut shops or speed traps: Evaluating models of supervision on police behavior. *American J. of Political Science* 37: 555-581.

Campbell, N. 1920. *Physics: The Elements*. Cambridge: Cambridge University Press.

Cramer, A., Waldorp, L., Maas, H. van der and Borsboom, D. 2010. Comorbidity: A network perspective. *Behavioral and Brain Sciences* 33: 137-150.

Crowther, C., Batchelder, W. and Hu, X. 1995. A measurement-theoretic analysis of the fuzzy logic model of perception. *Psychological Review* 102: 396-408.

Embretson, S. and Reise, S. 2000. *Item response theory for psychologists*. Hillsdale, NJ: Lawrence Erlbaum.

Friedman, D. and Massaro, D. 1998. Understanding variability in binary and continuous choice. *Psychometric Bulletin & Review* 5: 370-389.

Gaito, J. 1980. Measurement scales and statistics: resurgence of an old misconception. *Psychological Bulletin* 87: 564-567.

Hand, D. 2004. *Measurement theory and practice: The world through quantification*. New York: John Wiley.

Hardin, C. and Birnbaum, M. 1990. Malleability of "ratio" judgments of occupational prestige. *American J. of Psychology* 103: 1-20.

Healy, M. and Goldstein, H. 1976. An approach to the scaling of categorized attributes. *Biometrika* 63: 219-229.

Krantz, D. H., Luce, R. D., Suppes, P. and Tversky, A. 1971. *Foundations of measurement, vol. I*. New York: Academic Press.

Lakoff, G. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *J. of Philosophical Logic* 2: 458-508.

Lootsma, F. 1999. *Multi-criteria Decision Analysis via Ratio and Difference Judgement*. Boston and Dordrecht: Kluwer.

Lord, F. 1953. On the statistical treatment of football numbers. *The American Psychologist* 8: 750-751.

Luce, R. and Raiffa, H. 1957. *Games and decisions: Introduction and critical survey*. New York: John Wiley.

Michell, J. 1999. *Measurement in Psychology: A Critical History of a Methodological Concept*. New York: Cambridge University Press.

Osherson, D. and Smith, E. 1981. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9:35–58.Perline, R., Wright, B. and Wainer, H. 1979. The Rasch model as additive conjoint measurement. *Applied Psychological Measurement* 3: 237-255.

Ragin, C. 2000. *Fuzzy-set social science*. Chicago: University of Chicago Press.

Reise, S. and Waller, N. 2009. Item response theory and clinical measurement. *Annual Review of Clinical Psychology* 5: 27-48.

Reiss, A. 1971. *The police and the public*. New Haven, CT: Yale University Press.

Rossi, N., Wang, X. and Ramsey, J. 2002. Nonparametric item response estimates with the EM algorithm. *J. of Educational and Behavioral Statistics* 27: 291-317.

Saaty, T. 1977. A scaling method for priorities in hierarchical structures. *J. of Mathematical Psychology* 15: 234-281.

Sen, A. 1970. *Collective choice and social welfare*. San Francisco: Holden-Day.

Shafir, E. and LeBoeuf, R. A. 2002. Rationality. *Annual Review of Psychology* 53: 491-517.

Smithson, M. 1987. *Fuzzy set analysis for behavioral and social sciences*. New York: Springer.

Smithson, M. 2005. Fuzzy set inclusion: Linking fuzzy set methods with mainstream techniques. *Sociological Methods & Research* 33: 431-461.

Smithson, M. and Oden, G. 1999. Fuzzy set theory and applications in psychology. In *Practical applications of fuzzy technologies*, ed. H. Zimmermann, 557-585. Boston and Dordrecht: Kluwer.

Smithson, M. and Verkuilen, J. 2006. *Fuzzy set theory: Applications in social sciences*. Thousand Oaks, CA: Sage.

Stevens, S. S. 1946. On the theory of scales of measurement. *Science* 103: 677-680.

Takane, Y. and Shibayama, T. 1992. Structures in stimulus identification data. In *Probabilistic multidimensional models of perception and cognition*, ed. F. Ashby, 335-362. Hillsdale, NJ: Lawrence Erlbaum.

Torgerson, W. 1958. *Theory and methods of scaling*. New York: John Wiley.

Townsend, J. 1992. On the proper scale for reaction time. In *Cognition, information processing and psychophysics: Basic issues*, ed. H. Geissler, S. Link and J. Townsend, 105-120. Hillsdale, NJ: Lawrence Erlbaum.

Townsend, J. and Ashby, F. 1984. Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin* 96: 394-401.

UN DP 2009. *Human Development Report* 2009. New York: Palgrave Macmillan.

Verkuilen, J. 2005. Assigning membership in a fuzzy set analysis. *Sociological Methods & Research* 33: 462-496.

Verkuilen, J. and Smithson, M. in press. Mixed and mixture regression models for continuous bounded responses using the beta distribution. *J. of Educational* *and Behavioral Statistics*.

Wakker, P. 2008. Explaining the characteristics of the power CRRA utility family. *Health Economics* 17: 1329-1334.

Wallsten, T., Budescu, D., Rappoport, A., Zwick, R. and Forsythe, B. 1986. Measuring the vague meaning of probability terms. *J. of Experimental Psychology* 115: 348-365.

Wolfe, C. and Reyna, V. 2010. Semantic coherence and fallacies in estimating joint probabilities. *J. of Behavioral Decision Making* 23: 203-223.

Zachar, P. and Kendler, K. 2007. Psychiatric disorders: A conceptual taxonomy. *American J. of Psychiatry* 164: 557-565.

Zadeh, L. 1965. Fuzzy sets. *Information and Control* 8: 338-353.

Zand Scholten, A. and Borsboom, D. 2009. A reanalysis of Lord's statistical treatment of football numbers. *J. of Mathematical Psychology* 53: 69-75.