CLINICIANS' CONCEPTUAL USE OF COMORBIDITY

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

_____
Jared Wayne Keeley

Certificate of Approval:

_____
F. Dudley McGlynn
Professor
Psychology

_____
Roger K. Blashfield, Chair
Professor
Psychology

_____
Jeffrey S. Katz
Alumni Associate Professor
Psychology

_____
Alejandro A. Lazarte
Assistant Professor
Psychology

_____
George T. Flowers
Dean
Graduate School

CLINICIANS' CONCEPTUAL USE OF COMORBIDITY

Jared Wayne Keeley

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
August 10, 2009

CLINICIANS' CONCEPTUAL USE OF COMORBIDITY

Jared Wayne Keeley

Permission is granted to Auburn University to make copies of this dissertation at its discretion, upon request of individuals or institutions and at their expense. The author reserves all publication rights.

_____
Signature of Author

_____
Date of Graduation

DISSERTATION ABSTRACT

CLINICIANS' CONCEPTUAL USE OF COMORBIDITY

Jared Wayne Keeley

Doctor of Philosophy, August 10, 2009
(M.S., Auburn University, 2005)
(B.A., Knox College, 2002)

115 Typed Pages

Directed by Roger K. Blashfield

Diagnostic overlap, termed comorbidity, is a common occurrence in psychopathology. However, the current classification system for mental disorders (*DSM-IV-TR*; APA, 2000) does not explicitly address how diagnostic concepts should be combined. The *DSM* assumes an additive model where the study of how humans combine concepts in general would predict some degree of multiplicative combination. The study described herein used the methods of cognitive psychology to examine practicing clinicians' conceptualizations of comorbid cases. Two samples of clinicians drawn from the Association of Behavioral and Cognitive Therapies (ABCT; n = 48) and licensed psychologists in the state of Florida (n = 25) were asked to describe three disorders and their combinations using a predetermined list of symptoms. The primary evidence of a multiplicative model of combination was the presence of overextensions, or symptoms which occurred in a combination that were not a part of either constituent disorder.

Results indicated that clinicians included overextensions in non-zero amounts demonstrating use of some level of a multiplicative model in contrast to the additive model assumed in the *DSM*. These findings question the clinical utility of the current diagnostic system and suggest that the current paradigm of descriptive psychopathology may not be congruent with clinicians' conceptualizations.

ACKNOWLEDGMENTS

The author would like to thank Roger Blashfield for his inspiration and support throughout the doctoral process. Special thanks are due to Kristin Raley, Danny Burgess, Carrie Weaver, and Shannon Reynolds for their input and help in developing and implementing this project. Thanks to Pete Zachar for going out of his way to offer advice and expertise to make the project what it was. Last but not least, copious thanks are due to the individual clinicians who participated in this study for taking time out of their busy schedules to engage in the scientific process.

Style manual used	APA Publication Manual (5th edition)
Computer software used	Microsoft Word 2003

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES x

I. INTRODUCTION 1

II. METHOD 20

III. RESULTS: FIRST PHASE 27

IV. RESULTS: SECOND PHASE 36

V. DISCUSSION 40

REFERENCES 58

APPENDIX A 77

APPENDIX B 90

APPENDIX C 92

LIST OF TABLES

1. The mean (and *SD*) percent of overextensions for each disorder combination 68

2. Pairwise comparisons of percent of overextension across disorder combinations 69

3. Correlations between the percent of overextended symptoms and the size of constituent concepts.....70

4. Means (and *SD*s) for the percent of each constituent concept included in a combination.....71

5. Normative (>75%) symptom patterns across the single disorders and combinations 72

6. The mean (and *SD*) percent of recalculated overextensions for each disorder combination.....74

LIST OF FIGURES

1. Distribution of the percent of overextensions for the combination of GAD-APD.....76

INTRODUCTION

In recent years, diagnostic overlap has been a serious theoretical consideration for the classification of mental disorders. Epidemiological work has shown that for those individuals in the community who meet criteria for a mental disorder, half meet criteria for two or more disorders (Kessler, Chiu, Demler, & Walters, 2005). This diagnostic overlap, which has been termed comorbidity, is notably widespread. However, the problem is that the implicit theoretical model of the current diagnostic system predicts that there should be very little overlap. I will begin by providing a brief historical overview of modern diagnostic systems and the respective theoretical implications for diagnostic overlap. I will then turn to recent attempts to understand and define the concept of comorbidity. Although it is possible to study comorbidity as a means of understanding the nature of psychopathology, the studies described herein will focus on the pragmatic nature of how practicing clinicians implicitly work with the concept of comorbidity. It is not only important to understand the phenomenology of psychopathology, but also how clinicians understand that phenomenology and use such information in assigning diagnoses. Within cognitive psychology, a paradigm exists for studying how a person combines two concepts. I will outline a brief review of this literature to set the stage for how the current studies will translate these paradigms into the study of how clinicians combine diagnostic concepts to understand their patients.

The DSMs and Comorbidity

The classification of mental disorders in the United States has a long history; however, only part of that history is relevant to understanding the way in which current diagnostic systems incorporate comorbidity. The first two *Diagnostic and Statistical Manuals of Mental Disorders* (*DSM*) were responses to the nosological confusion present in this country. Prior to their creation, there were at least three major diagnostic systems in use, and idiosyncratic systems were present in nearly every hospital and asylum (Houts, 2000). The *DSM-I* and *DSM-II* were created to unify the diagnostic concepts used across the nation so that large scale epidemiological data could be gathered on the prevalence of mental disorders, but these two systems defined mental disorders in a very imprecise fashion. They provided prose descriptions that characterized the disorders, but were vague as to how a clinician decided if the diagnosis was present or not in any given patient. In the *DSM-I* and *II*, diagnoses were assigned in a hierarchical fashion; therefore, if you qualified for one diagnosis, you were ruled out of having any of a subset of other diagnoses. Thus, it was very unlikely that a patient would be given multiple diagnoses, and comorbidity was not considered an issue. The major concern regarding these two diagnostic systems was their relative unreliability. Clinicians were highly inconsistent across time and across raters when assigning diagnoses (Cooper et al., 1972; Kendell, Cooper, Gourlay, Sharpe, & Gurland, 1971).

Thus, the major impetus for the creation of the third edition of the *DSM* was to increase the reliability of diagnosis. This goal was achieved by defining mental disorder categories through lists of diagnostic criteria, of which a defined number must be present

to qualify for a particular diagnosis. Structured clinical interviews were developed based on these criteria, greatly enhancing the reliability of diagnosis (e.g., First, Spitzer, Gibbon, & Williams, 1995). Unlike the first two editions, with the third and later editions of the *DSM*, diagnoses were given in a bottom-up fashion rather than the top-down hierarchical decision tree of earlier editions. As a result, it became possible for a person to meet criteria for multiple disorders. In fact, early work examining the possibility of diagnostic overlap found that many people did meet criteria for multiple disorders when exclusionary criteria were not employed (Boyd et al., 1984; Kessler et al., 1994).

Modern Conceptions of Comorbidity

Given the diagnostic structure of the *DSMs*, diagnostic overlap can be considered, to some degree, an artifact of the classification system. Because multiple diagnoses are possible, they will occur. If overlap occurs at a rate expected by chance, then there is no meaning to the overlap and it is not theoretically important. On the other hand, if diagnoses co-occur more often than base-rate expectations, the phenomenon has implications for the classification system.

The issue of diagnostic overlap has been termed "comorbidity;" however, there are those that contend this term is a misnomer (Lilienfeld, Waldman, & Israel, 1994). As a term, comorbidity was introduced by Feinstein (1970) in the context of medical epidemiology. Later, Lilienfeld (2003) reported that Feinstein never intended for the term comorbidity to refer to all cases of diagnostic overlap, which seems to be its current usage in the literature. There are several ways to think about diagnostic overlap, and it would be useful to examine each in turn.

First, one could think of diagnostic co-occurrence in the context of a patient being seen in a clinical setting. In that case, the patient presents with one problem (the index disorder) but meets criteria for another disorder (the comorbid condition). That second condition may be clinically relevant to the first or completely independent. For example, if a patient presents with a case of panic disorder and is concurrently depressed, symptoms of depression like lack of motivation may interfere with the treatment of panic. On the other hand, a patient presenting with panic disorder may have a comorbid diagnosis of insomnia, which could have little to do with the treatment of panic. At this level of understanding, comorbidity may directly impact pragmatic treatment decisions.

A related issue concerns the causality of comorbidity. In the cases above, it could be that depression and panic are meaningfully correlated, e.g., the depression may cause the panic, vice versa, a third variable could cause both, or some alternative complex causal model could account for the relationship. However, when two co-occurring conditions are not meaningfully related but simply overlap by chance, as could be the case with panic and insomnia, that sort of comorbidity is of decidedly less theoretical interest. Given the base rates of disorders in the population, one would expect a certain number of cases to present with overlap just by chance.

As a part of the Epidemiologic Catchment Area study, Boyd et al. (1984) conducted the first study to examine if disorders co-occurred at greater than chance levels. Boyd et al. calculated odds ratios on the presence or absence of two disorders, such as panic and depression. An odds ratio is calculated by determining the number of people in a sample that have (a) both disorders, (b) only one of the disorders, or (c)

neither disorder. The frequency of dual diagnosis is multiplied by the frequency of those with neither disorder and divided by the product of the frequency of each single diagnosis. They found an odds ratio of 18.8 for panic and depression, which indicates that if a person has one disorder, the chance of also having the other is 18.8 times greater than one would expect given the base rates of the disorders. In other words, they found that panic disorder occurred at a rate of 0.5% in their sample, and depression occurred at a rate of 2.4%. One would expect the two disorders to overlap at a rate of 0.01% (.05 x .24). However, the disorders actually overlapped at a chance of 2.2%. Thus, there seems to be a significant correlation between the occurrence of these two disorders. Boyd et al. (1984) found significant co-occurrence between many of the *DSM-III* mental disorder categories. These findings have been replicated in more recent, large scale epidemiological studies of comorbidity (Kessler et al., 2005; Kessler et al., 1994).

Despite the evidence provided by Boyd et al. (1984) and the subsequent replications, there are possible artifactual reasons that disorders could co-occur at greater than chance levels. For example, Berksonian bias (Berkson, 1946) states that a person with comorbid disorders has twice the chance of seeking treatment as does a person with a single disorder, and so has a greater chance of being included in studies. In addition, comorbidity rates may be overestimated because of clinical selection bias, i.e., the idea that individuals with multiple disorders may be more impaired, and therefore overrepresented, in treatment studies (du Fort, Newman, & Bland, 1993; Lilienfeld et al., 1994). However, significant rates of comorbidity occur in community based samples (e.g., Kessler et al., 2005). Finally, it has been demonstrated that structured interviews

generate much higher rates of diagnosis overall, and higher rates of comorbidity specifically than unstructured interviews (Verheul & Widiger, 2004; Zimmerman et al., 2005). On the other hand, differences in diagnostic method may not be artifactual differences, but true differences. Structured interviews ensure that the breadth of psychopathology is assessed whereas unstructured interviews only address salient points and legitimate disorders may be overlooked.

Comorbidity cannot be solely explained by chance overlap or selection bias. Instead, there seems to be legitimate correlation between various sets of disorders indicating that the pathology may be meaningfully related. The *DSM*s have followed a categorical model, i.e., the diagnoses included as mental disorders in the manual are conceived as separate categories. Only one statement in the current edition of the *DSM* (APA, 2000) addresses disorders' potential overlap: "In *DSM-IV*, there is no assumption that each category of mental disorder is a completely discrete entity with absolute boundaries dividing it from other mental disorders" (p. xxxi). The authors of the *DSM* recognize that the categorical approach to classification is an assumption and has limitations. Indeed, the majority of data suggest that the current mental disorder categories are not discrete. Although the manual remains neutral regarding the meaning of comorbidity, as a diagnostic system, the *DSM* must have a means of recording comorbidity. When a person meets criteria for more than one disorder, all disorders are diagnosed. However, there are at least two alternatives for the ways in which a clinician could be combining the concepts of two disorders when listing a multiple diagnosis.

The first alternative is an additive model. In this model, the symptoms of the first disorder are simply added to those of the second when they co-occur. Disorders are described in the *DSM-IV-TR* as polythetic criteria sets. To meet criteria for a diagnosis of Major Depressive Disorder, for example, a person must have five of nine symptoms listed. If a person were diagnosed as having the comorbid condition of Panic Disorder, an additive model would add the symptoms of panic to those of depression to complete this person's clinical symptom picture. The *DSM*s implicitly follow an additive model of combination.

However, it could be that clinicians combine concepts in a multiplicative fashion. In this case, the combination of two disorders creates an emergent symptom picture. The clinical presentation of two comorbid conditions is not considered the simple addition of one symptom set to another. Instead, it could be that certain symptoms of one disorder are much more likely to occur when in combination with another disorder. For example, a change in sleeping pattern (either insomnia or hypersomnia) is a common symptom of depression. However, because of the polythetic nature of the concept, this symptom is not required for a diagnosis. Therefore, it is only present in a portion of those who meet criteria for the disorder. A comorbid condition might change the likelihood of a person having that symptom or change the nature of that symptom. For instance, also meeting criteria for Panic Disorder might make it more likely that the person experiences insomnia. In that example, the likelihood of sleep disturbance is increased and the nature of the symptom is specified towards the direction of insomnia rather than hypersomnia. Further, a multiplicative combination of disorders might result in emergent symptoms, or

symptoms which are not considered a part of either single concept. For example, clinicians might envision a person with both panic and depression to be submissive interpersonally, which is not a symptom of either single disorder.

The science of psychopathology has begun to respond to the theoretical implications of comorbidity. One example is the concept of "Negative Affect Syndrome" which is a reaction to the high degree of overlap seen between affective (Major Depressive Disorder, Dysthymia) and anxiety (Generalized Anxiety Disorder, Panic Disorder, Social Phobia, Simple Phobia, etc.) disorders (Brown, Chorpita, & Barlow, 1998; Clark & Watson, 1991; Mineka, Watson, & Clark, 1998). These authors have proposed that the set of affective and anxiety disorders are different expressions of a single syndrome, which they propose is characterized by increased negative affect and varying degrees of positive affect. This idea accounts for the overlap seen among these disorders by claiming they are part of the same pathological process. Using structural equation models (SEM), Brown et al. tested the plausibility of the proposed factors of positive and negative affect, along with autonomic arousal, for accounting for the symptom patterns of individual patients. Across this and other studies, the model has held considerable promise (Brown, 2007; Olino, Klein, Lewinsohn, Rohde, & Seeley, 2008; Sellborn, Ben-Porath, & Bagby, 2008; Tackett, Quilty, Sellborn, Rector, & Bagby, 2008; Watson, 2005; Weinstock & Whisman, 2006). Similar spectrum concepts have also been proposed for the psychotic (Olsen & Rosenbaum, 2006; Siever & Davis, 2004) and autistic (Nylander, Lugnegard, & Hallerback, 2008; Szatmari, 1992; Wolf-Schein, 1996) disorders.

The *DSM* explicitly states in its Introduction that "Our highest priority has been to provide a helpful guide to clinical practice. We hoped to make *DSM-IV* practical and useful for clinicians" (APA, 2000, p. xxiii). Thus, how useful the manual is for clinicians is an important empirical question. If the *DSM* follows a model that contradicts the way in which clinicians conceptualize comorbid conditions, then the model needs to be revised. It is the aim of the studies described herein to begin to examine the model by which clinicians combine diagnostic concepts. If the model of the *DSM* matches that of clinicians, it is more likely that it is achieving its aim of being a helpful, practical, and useful guide. If it does not, then perhaps alternative models should be considered in future editions. Before exploring specifically how clinicians combine diagnostic concepts, it would be useful to understand how people combine concepts in general, which has been an area of much study in the cognitive psychology literature.

Conceptual Combination

A substantial literature in the domain of cognitive psychology has examined the ways in which humans combine concepts. When discussing the issue of diagnostic comorbidity, one could argue that what is occurring could be described as an instance of conceptual combination—a clinician takes one diagnostic concept and tries to incorporate it with another in the context of a single patient. Thus, what is known about conceptual combination generally could provide a framework for understanding how clinicians deal with comorbidity specifically.

Conceptual combination became an area of study for two main reasons. First, it is a meaningful window into natural language processes. Words are combined in everyday

language all the time to express new meanings like *web page*, *cellular phone*, and *flying saucer*. Studying these sorts of combinations provides insight into meaning construal and how language is formed. Second, the study of conceptual combination provides interesting insight into the understanding of concepts in general. The way in which concepts are combined can place constraints and suggest revisions of current theories of concept structure, formation, and usage.

There are many types of concepts, and many ways in which those concepts could be combined. The literature on conceptual combination has focused primarily on noun phrase combinations like *brown apple*, *elephant tie*, or *pet that is also a bird*. These phrases come in several forms. The first is the pairing of an adjective with a noun, as in *brown apple*. In English, the convention is that the first word in such a phrase modifies the second (although this is not always the case; see Costello & Keane (2001) for evidence of so called focus reversals). Thus, the first word in the phrase is usually referred to as the modifier, and the second is called the head noun. These phrases can also occur as noun-noun combinations, as in the case of *elephant tie*. Noun-noun combinations have received much more attention than adjective noun pairs. In adjective-noun phrases, the components are different parts of speech and so may operate under different combinatorial rules than concepts on the same level. These combinatorial phrases can be either predicating or non-predicating. A predicating phrase is one in which the modifier can meaningfully be turned into a sentence predicate, as in the case of *an apple that is brown*. Relative clauses like *pet that is also a bird* simply make that relation explicit. However, not all combinations are predicating in that way, as a *tie that is an*

elephant does not make sense. Non-predicating combinations are bound by some other type of relation besides an identity relation.

A major focus of the work on conceptual combinations has been how people combine two concepts. For example, the combination *elephant tie* could have several interpretations, e.g., a tie made for an elephant, a tie with a picture of an elephant on it, or a very large tie. Theoretically, there are numerous potential interpretations of a novel phrase. Wisniewski (1996) outlined three basic strategies that have been investigated for combining concepts: property mapping, hybrid combinations, and relation linking. Each strategy refers to the ways in which the features of the constituent concepts are combined to create a conjunction.

Property mapping leads to a combination in which a property of the modifier is used to describe the head noun. For example, one interpretation of *elephant tie* states it is a very large tie. In this case, the property of the elephant's size, since it is a salient feature of the elephant, is transplanted to describe the size of the tie. Wisniewski (1996) refers to this as an alignable difference. Both constituents have a property of size, and the combination highlights the difference in their size. He argues that for property mapping to occur, there must be an alignable difference.

The second way of combining concepts is referred to as a hybrid. In a way, it is the simplest method of combining two concepts. A hybrid concept simply takes the features of two concepts and blends them together. For example, a hybrid interpretation of *buffalo cow* might be "a large animal with the properties of both a buffalo and a cow." A hybrid can be thought of as an extreme example of property mapping in which multiple

properties are mapped between the two concepts. Both property mapping and hybridization are examples of an additive method of combining concepts, in which the features of one are simply added to the features of the other.

Relation linking refers to when two concepts are joined through a specific relation. For example, the interpretation of *elephant tie* that says it is a tie made for an elephant relates the concepts of tie and elephant with a *made for* relation. Relative clause combinations of the sort *X that is also a Y* make that relation explicit. Given that the identity relation is perhaps the simplest relation, early work in conceptual combination focused on understanding how concepts combine when the relation joining them is simply that the phrase instantiates something that is a member of one concept and also a member of another. In general, when two concepts are combined in that fashion, one concept dominates the features of the combination (Chater, Lyon, & Myers, 1990; Hampton, 1988; Storms, De Boeck, Van Mechelen, & Ruts, 1996). In other words, the combination resembles one of its constituent members more than the other.

More complex, non-predicating relations also follow a general pattern of combination. Words can be joined by any number of relations, but certain words tend to use some relations more than others. Gagne and Shoben (1997) determined the frequency of thematic relations for a large set of head nouns and modifiers. If a certain relation accounted for 60% or more of the combinations of which the word was a part, then that relation was considered high-frequency. For example, when *mountain* was used as a modifier, its most frequent relation was *in* (e.g., *mountain bird* interpreted as meaning a bird in the mountains). They then compared how sensible various combinations of words

were across high versus low frequency relations. They concluded that the manner in which a word is usually combined (i.e., the frequency of its thematic relations) is a large determinant of how easily the combination is comprehended. One interesting finding was that only the modifier played a large role in determining the combination. The frequency of the head noun's relations made little difference.

The results of Gagne and Shoben's (1997) study have several problems, as pointed out by Wisniewski and Murphy (2005). First, Gagne and Shoben's stimuli, in trying to be novel combinations, failed to account for the pattern of natural language whereby some modifiers are associated with certain types of nouns more than others. Thus, determining the relation frequency of words in a random set of words may miscalculate the frequency at which words use certain relation types in natural language. Second, some combinations are more plausible than others. Wisniewski and Murphy (2005) cite the difference between the phrases *cooking hole* and *sports tension*. It may be relatively difficult to interpret *cooking hole* as a hole for cooking compared to interpreting *sports tension* as tension caused by sports because the latter is more plausible or likely. This disparity may have confounded the differences they found in their study. Further, some phrases are more familiar or lexicalized than others. For instance, Gagne and Shoben included the phrases *financial crisis* and *plastic crisis* as stimuli in their study. *Financial crisis* is a combination that already has a familiar meaning and usage in language, whereas *plastic crisis* may be a truly novel combination for most participants. When Wisniewski and Murphy reanalyzed Gagne and Shoben's data, and included

familiarity and plausibility ratings, they found that these two factors could have potentially accounted for the effect Gagne and Shoben originally found.

Most models of conceptual combination have focused on the role of relation linking, and have considered property mapping and hybrid combinations to be rare or unimportant. However, Wisniewski (1996) and Wisniewski and Love (1998) examined the frequency with which the three types of combinations occur. They found evidence that, although relation linking strategies may be common, property mapping strategies occur in non-trivial amounts. In certain contexts, property mapping strategies are even more common than relation linking. Hybrid combinations do occur, but are relatively rare and only seem to occur in certain contexts. Thus, since both property mapping and relation strategies occur at high rates, it is interesting to determine what sorts of factors lead a person to use one strategy or another.

Wisniewski (1996) examined 64 noun-noun combinations. Half of the pairs contained highly similar constituents (e.g., shark piranha), the other half were highly dissimilar (e.g., pineapple piranha). Participants generated a description of the term, which was then coded as an example of relation linking, property mapping, hybrid combination, or other. The coding occurred according to the definitions of the terms given above. For example, a response was coded as relation linking if the noun pair was described by the relation of the two objects, e.g., "box that holds ladders for *ladder box*" (p. 438). The author stated that only a very few number of responses were judged to belong to multiple categories and that raters agreed on 90% of the codes. Relation linking was much more common in dissimilar pairs (52%) than in similar pairs (7%). Property

mapping occurred more frequently when the word pair was similar (72%) versus dissimilar (48%). Finally, hybrid combinations occurred only in similar pairs, accounting for 20% of all the similar pairings.

Wisniewski (1996) interpreted these findings as meaning that the similarity of the combination plays an important role in how the concepts are combined. When concepts are very similar, it is easy to take the feature of one component concept and map it onto a corresponding feature of the other. He referred to this as an alignable difference. In other words, when two concepts are very similar, they have corresponding features. A shark and a piranha are both predatory, live in water, and viewed as vicious. However, a piranha and a pineapple have no such features in common. In the case when the two concepts are similar, like the shark and piranha, both have similar feature domains but the specifics of the feature may be different. Thus, they have a difference that is alignable on some feature. When the concepts are dissimilar, there are no features that the concepts share, and so there are no parts of the concept which can directly map onto the other. Another way of thinking about alignable differences is that they are similar to the different levels of an independent variable. They are different aspects of the same feature. However, it is not possible to compare a level of one independent variable to a level of a different independent variable in the same way, as in the case of a dissimilar pair. Thus, Wisniewski stated that property mapping seems to occur when two concepts are similar, but some other process must occur when the concepts are dissimilar.

When the concepts are dissimilar, relation mapping becomes an alternative strategy. Notice, in Wisniewski's (1996) data, participants combined dissimilar pairs

using a property mapping strategy about half of the time and a relation linking strategy the other half. This suggests that when concepts are dissimilar, it is still possible to map some properties directly. Still, sometimes participants did not map a property but explained how the dissimilar concepts were combined through explaining their relation. Wisniewski hypothesized that the relation linking strategy may be a way that participants cognitively reconcile why two disparate things are joined.

An interesting phenomenon occurs when two disparate categories are combined. Sometimes, when two concepts are combined, a feature of the combination emerges that was not a feature of either constituent concept, or at least was not as salient a feature. These emergent properties are termed "overextensions" (Hampton, 1988). Specifically, sometimes a category will have greater similarity to a conjunction than it will to either constituent concept (Osherson & Smith, 1981). For example, goldfish might be considered very typical of the combined concept *pet fish*, but not particularly typical of either *pets* or *fish* in general. Such overextensions proved difficult for traditional concept theories to explain. Set theory would predict that the combination of a concept would be the intersection of those two concepts, i.e., things that were both pets and fish. Hampton (1988) studied the conjunction of *sports that are also games*. In that example, participants sometimes rated things that were not *sports* as a *sport that is also a game* (e.g. chess) or something that is not a *game* as a *sport that is also a game* (e.g. fencing). Chater, Lyons, and Myers (1990) replicated Hampton's findings of overextension while controlling for a variety of potential confounds in the methodology.

In the studies above, overextensions were the inclusion of particular exemplars in a combined concept. Overextensions also occur at the level of the features of a concept. For example, Springer and Murphy (1992) found that certain features of a combined concept were judged to be true of the concept much faster than features of the head noun. For instance, the sentence "Peeled apples are white" is an example of an emergent property of the combined concept peeled and apple. Peeled things are generally not white, nor are apples, but peeled apples are white. Thus, "white" is an overextended feature of the combination. The sentence "Peeled apples are white" was judged to be true faster than a similar sentence in which the feature listed was true of one of the constituent concepts, specifically "Peeled apples are round." Apples in general are round and this feature does not change in virtue of the apple being peeled. Springer and Murphy interpret this finding to mean that combined concepts like *peeled apples* are interpreted holistically rather than by the features of their component concepts. If they were interpreting the meaning of each individual concept first and then combining the concepts, the features of the head noun alone would be accessed first, and the sentence "Peeled apples are round" would be more quickly recognized as true.

The current study

To return to the issue of comorbidity, there are several lessons that can be gleaned from the literature reviewed above. First, both additive and multiplicative models of combination have been observed in various contexts. The choice of an additive or multiplicative model is largely, but not entirely, determined by the similarity of the concepts being combined. It follows that the ways in which comorbid disorders might be

combined could also depend upon the perceived similarity of the diagnostic concepts. Further, "overextensions" are primary evidence that concepts are combined in a multiplicative fashion. Evidence of overextended symptoms in the combination of mental disorders would suggest that the disorders are being combined following a multiplicative model.

Given that the similarity of stimuli is the best predictor of how concepts will be combined, this study will employ stimuli that represent what we consider to be two very similar disorders (Major Depressive Disorder and Generalized Anxiety Disorder) and a disorder that is dissimilar to both (Antisocial Personality Disorder). Major Depressive Disorder (MDD) and Generalized Anxiety Disorder (GAD) are commonly comorbid conditions (Kessler et al., 2005; Kessler et al., 1994). Further, recent theoretical and empirical work has claimed that they are two expressions of the same negative affect syndrome, and are not separate psychopathological conditions (Brown, Chorpita, & Barlow, 1998; Clark & Watson, 1991; Mineka, Watson, & Clark, 1998). While MDD and GAD are conceptualized as internalizing disorders, or symptoms involving an inward expression, Antisocial Personality Disorder (APD) is seen as an externalizing disorder, i.e., symptoms are expressed outward (Krueger, Markon, Patrick, & Iacono, 2005). Therefore, the findings cited above (Hampton, 1988; Springer & Murphy, 1992, Wisniewski, 1996) would predict that MDD and GAD would be combined using additive, property mapping strategies, and so overextensions would be less likely, where the combinations of the two with APD would be more likely to follow multiplicative, relation linking strategies and overextensions may emerge.

The methodology employed in this study will be able to answer a set of questions. First, the study will examine the prevalence of multiplicative models of concept combination among practicing clinicians as operationalized by symptom overextension. Given that the current diagnostic model is additive, the presence of any multiplicative combinations is theoretically interesting. Second, the study will examine factors that predict overextensions. The literature cited above has demonstrated that similarity is a common factor in determining the type of combination (Wisniewski, 1996). That finding may or may not replicate in the current context, and other factors may emerge as predictors of overextensions (e.g., years of experience). Third, some of the literature on conceptual combination has found that one concept may dominate the combination (Chater et al., 1990; Hampton, 1988; Storms et al., 1996). The data provided by this study can address if a particular disorder is seen as primary to other disorders as evidenced by its dominating the combination. Fourth, the data may provide a normative pattern of the symptoms that are seen as particularly indicative of individual disorders and those disorder combinations, thereby setting the stage for future studies that might inform later revisions of the diagnostic manual.

METHOD

Participants

Participants for this study came from two samples: clinicians licensed in the state of Florida and professionals belonging to the Association of Behavioral and Cognitive Therapies (ABCT). The Florida state licensing board provides a list of all licensed practicing psychologists in the state. Of the 4,028 clinicians licensed in the state, 500 were randomly selected to receive information about the study. The second sample consists of members of ABCT. The association provides a list of all members mailing addresses. At the time of data collection, there were 2423 members, not including graduate students or associated members (e.g. vendors). The 69 members who listed their home state as Florida were also excluded so that they did not have a double chance of being selected. Five-hundred members of ABCT were randomly selected to receive information about the study.

The rationale for the inclusion of these two samples is that they represent very different sampling biases. For the first, the Florida clinicians are mostly local practitioners of varying theoretical orientation; however, they are all present in a limited geographical range. The ABCT sample includes a wider variety of professionals, including practitioners, faculty, and researchers from all geographic regions of the United States, but they are limited to behavioral and cognitive-behavioral orientations. If a different pattern of results emerges from these two samples, those results would indicate

that there might be interesting theoretical or geographical influences upon comorbidity conceptualization. However, if the two samples display a similar pattern, these results would speak strongly to the generality of the phenomenon.

The average participant was 46 years old (*SD* = 11.33) with 20 years of experience (*SD* = 10.45). Most participants were female (63%). The majority of participants had experience working with adults (88%), many with adolescents (66%), and 48% of the sample had experience with children. Forty-five percent of participants consulted the *DSM* once a week, 34% once a month, 15% consult it rarely, and 6% consult it daily. The largest portion of participants listed their theoretical orientation as cognitive-behavioral (63%), with an additional 19% describing themselves as either integrative or eclectic and 6% as strictly behavioral. The remaining 12% were single representations of orientations including psychodynamic, interpersonal, applied developmental, humanistic, etc. Participants worked in a variety of settings, including private practice (41%), faculty position (16%), hospital/medical center (14%), community mental health/outpatient clinic (11%), Veterans Administration (5%), college counseling center (5%), and others. The two samples did not differ significantly on any of these demographic variables except for orientation ($\chi^2(3) = 11.73$, $p < .01$). As expected, the ABCT sample had a higher proportion of cognitive-behavioral or behavioral clinicians, where a Florida clinician was more likely to be eclectic/integrative or in the 12% of "other" orientations.

Participants were initially mailed an information letter about the study and a postcard to return if they were interested in participating. Those who returned postcards

were mailed the first phase of the study. Those who completed and returned the materials were then mailed the follow-up phase of the study. The return rate for the first mailing of postcards was 9.70% (40 from Florida, 57 from ABCT). Of those who returned postcards, 75.26% returned the first phase of the study (25 from Florida, 48 from ABCT). Of the 73 individuals who returned packets, 50.68% (12 from Florida, 25 from ABCT) returned materials from the follow-up portion of the study. Overall, participants in the ABCT sample had a better initial response rate and better follow-up than Florida participants.

Materials

The stimuli for the study were three common psychiatric disorders, all possible pairwise combinations of the disorders, and the three-way combination. The disorders included were Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), and Antisocial Personality Disorder (APD). In an effort to limit the difficulty and time commitment of the task, only three disorders were used.

Pilot work tested multiple methodologies for examining conceptual combinations with these stimuli. For more information about the pilot work, please reference Appendix A. The result of the pilot study was the use of a forced-choice methodology, where participants would circle symptoms from a predetermined list that they felt were expressions of the disorder or disorder combination. In the pilot study, Panic Disorder was used rather than GAD. The current study shifted to using GAD as this disorder is a better parallel to MDD (a general expression of depression versus a general expression of

anxiety). Further, in the work on negative affect syndrome cited above, GAD is more closely related to MDD than is Panic Disorder.

Forced choice task. Stimuli were presented each on a separate page in the same format. At the top of the page "An individual with …" was presented in bold type, where the ellipsis was replaced with the name of the disorder or disorder combination. This phrase was included to direct the participant to think of exemplars of the category rather than some abstracted prototype of the category. Participants received the following instructions:

> *On the following pages, you will find the name of a disorder or multiple disorders. Imagine what it would be like to see a person with those diagnoses in therapy. On the same page you will find a list of descriptions. Based on your clinical experience, circle as many* ***symptoms/expressions*** *as you feel fit to describe the disorder or group of disorders. Your goal is not to reproduce the criteria in the DSM, but to list the clinically relevant features of each disorder based on your experience. You have as much time as you need to work. This is not a test, and there are no "right" answers.*

Under the name of the disorder on each page was a list of 120 symptoms. These symptoms were chosen to be representative of the entire domain of psychopathology and personality functioning, as they are drawn from a major psychopathology assessment instrument (the *Personality Assessment Inventory* [*PAI*], Morey, 1991) and the currently most investigated theory of personality functioning that has been applied to the personality disorders, the Big Five personality factors (Costa & Widiger, 1994). Of the 120 symptoms included, 60 symptoms came from the *PAI* and 60 came from the Big Five factors. For the *PAI*, the 60 symptoms were taken from the clinical and treatment scales, excluding items on the validity and interpersonal scales. The 15 clinical and treatment

scales are represented by 343 items. The 60 items selected for inclusion were those from each scale that had the highest loadings on their respective scale based on factor analytic studies. An equal number of items were selected from each scale, assuming that the items were not directly redundant. Items were rephrased to be one to three words in length. The remaining 60 items came from the Big Five personality factors. Each of the five factors has six facets. A representative descriptor for each tail of each facet was included (2 tails x 6 facets x 5 factors = 60 items). The 120 descriptions were presented in alphabetical order. They are included for reference in Appendix B.

Follow-up task. The follow-up task used the same forced-choice list of symptoms, but altered the method of presentation for the stimuli. The rationale for the follow-up task was threefold. First, it provided a measure of test-retest reliability which can assess the degree to which the pattern of a participant's responses is due to chance. Second, the first phase of the study does not address the directionality of influence in the combination. For instance, a clinician might consider MDD with comorbid GAD to have a different symptom picture than GAD with comorbid MDD. Third, in the first phase, participants are not directly told to add or omit symptoms from the single disorders when combining comorbid conditions. Therefore, overextensions and omissions from the lists of original symptoms for a single disorder may simply be due to chance inclusion or exclusion when circling symptoms in a combination.

To address these concerns, the follow-up task presented each single disorder in the same manner as the first study. However, following each single disorder, participants were asked what would change if that person also had one of the other two remaining

disorders, both presented in turn. They were instructed to circle descriptions they would add and to cross out descriptions they would omit. Each of the three single disorders were presented, followed by each of the possible additions, so that all possible two way permutations were represented. The order of presentation was randomized across participants. The directions for the task were as follows:

> *On the following pages, you will find the name of a disorder. Imagine what it would be like to see a person with those diagnoses in therapy. On the same page you will find a list of descriptions. Based on your clinical experience, circle as many* ***symptoms/expressions*** *as you feel fit to describe the disorder. Your goal is not to reproduce the criteria in the DSM, but to list the clinically relevant features of each disorder based on your experience. You have as much time as you need to work. This is not a test, and there are no "right" answers. Once you have completed the task for a disorder, you will be asked what you would expect to change if that same person also had another disorder. You may either add symptoms (circle added symptoms) or take symptoms away (cross out deleted symptoms). You will be asked to complete this process for three different initial diagnoses.*

Demographic questionnaire. Participants also completed a demographic questionnaire, assessing their age, sex, years of clinical experience, domain of experience (adults, adolescents, children), frequency with which they consult the *DSM*, theoretical orientation, and primary work setting. In addition, participants were asked to rate their familiarity with each of the disorders included in the study on a 5-point Likert-type scale ranging from *not at all* to *very*. Participants also rated the perceived similarity of the possible disorder pairs on a Likert-type scale ranging from -3 (*dissimilar*) to 3 (*similar*) with 0 as a neutral midpoint.

Procedure

Participants were given as much time to work as they needed. However, they were asked to complete the materials in one sitting if possible. The packet of materials for the first phase was organized such that participants completed the demographic questionnaire first, then they saw the instruction sheet for the task, followed by the task itself. The three single disorders were presented first in random order, followed by the three two-way combinations and one three-way combination in random order. This arrangement ensured equal exposure to the stimuli prior to encountering the combinations while controlling for possible order effects. The organization of the follow-up packet included the instruction sheet first, followed by the presentation of a single disorder and its two possible combinations. After that set, a participant was presented with a second single disorder and its two combinations, followed by the final single disorder and combinations. The order of the single disorders and the order of the possible combinations were randomized.

RESULTS: FIRST PHASE

Missing Data

Three participants did not complete all the materials by omitting descriptions of the combination of GAD and APD (all three) or MDD and APD (two of the three). For example, one participant wrote on the materials, "[I] have never seen nor can [I] imagine seeing someone with this combination." As many of the analyses conducted below are based on multivariate statistics and these analyses require complete data points, the three participants that left one or more pages of the study blank were excluded from all analyses, for a final $n = 70$.

Prevalence of overextensions

The primary evidence of a multiplicative model of combination is the presence of overextensions. In the case of these data, an overextension is a symptom that is endorsed in a combination that is not present in the symptom lists of either constituent disorder. For the initial phase of the study, the number of overextensions for each participant for each possible combination was calculated. However, the raw number of overextensions may be misleading, because participants included different numbers of symptoms in their descriptions. Therefore, the raw number of overextensions was divided by the total number of symptoms circled and multiplied by 100 to create a percent. Table 1 presents the average percent of symptoms in a combination that were overextensions across both samples.

As seen in Table 1, total amounts of overextension occurred at a rate of over 10% across the possible combinations (differences between the samples will be discussed in a later section). To test if this rate is significantly different from zero, all four combinations were entered into a MANOVA with a default test matrix value of zero. The overall MANOVA was significant; Wilks' λ (4,66) = 0.26, $p < .001$. The individual tests for each combination were also significant: for MDD-GAD, $F(1,69) = 160.10$; for MDD-APD, $F(1,69) = 119.99$; for GAD-APD, $F(1,69) = 116.19$; and for MDD-GAD-APD, $F(1,69) = 117.10$, all *p*s < .001.

While there were significant rates of overextension, the variability across participants was large. The range of values present for each of the four combinations went from 0% to approximately 50%. As an example, Figure 1 depicts the distribution of percent of overextension for the GAD-APD combination. Thus, there seem to be factors that contribute to an individual participant's propensity to overextend symptoms or not.

Factors that predict overextension

The previous literature on conceptual combination noted that the similarity of two concepts was the best predictor of how the concepts would be combined (Wisniewski, 1996). Concepts that are more similar should evidence fewer overextensions. The diagnostic concepts of MDD and GAD are more similar to each other than they are to APD. Thus, combinations involving APD should have a higher rate of overextension.

Within the same MANOVA conducted above, a number of custom hypothesis tests examined if APD combinations evidenced higher rates of overextension. Table 2 presents all the possible pairwise comparisons across the disorder combinations. In short,

participants had a higher rate of overextension when combining GAD and APD than any other combination, the rest being equal. Combinations of MDD and APD were not higher, as might be expected.

The above test is based on the manipulation of disorders assumed to be similar or dissimilar by previous research. However, it could be that participants did not consider this pattern of similarity to be true. Participants completed similarity ratings on a Likert-type scale of -3 (not at all similar) to 3 (very similar) for each of the disorders included in the study. The mean perceived similarity rating for the MDD-GAD combination was 0.73 (*SD* = 1.49), indicating moderate similarity. Participants indicated that the disorders in the combinations of MDD-APD (*M* = -1.63, *SD* = 1.37) and GAD-APD (*M* = -1.89, *SD* = 1.34) were dissimilar. Thus, participants endorsed the same pattern of similarity between these disorders predicted by the literature.

While the overall pattern of responses only evidence higher rates of overextension in the combination of GAD and APD, each individual's similarity rating may predict that person's rate of overextension. Separate regressions were conducted predicting the percent of overextensions for each combination using the respective similarity ratings as predictors. Participants' perceived similarity ratings were not significant predictors of their rates of overextension for any disorder combination.

However, there are many other factors that may predict an individual's rate of overextension. An individual's familiarity with the diagnosis may influence the number of overextensions. For example, a person who is not familiar with a given diagnosis may be less likely to include symptoms beyond the basic descriptions in the *DSM*. Similarly, a

clinician's years of experience or use of the *DSM* may influence their conceptualizations of disorders and their combinations. Separate regressions were conducted for each of the combinations, entering the factors of familiarity, years of experience, and frequency of consulting the *DSM* as independent variables. None of these variables were significant predictors of overextension rate.

The effect may be due to differences in theoretical orientation. Participants' orientations were coded into one of four types: cognitive-behavioral ($n = 43$), eclectic/integrative ($n = 14$), behavioral ($n = 4$), and "other" including single instances of psychoanalytic, humanistic, existential, etc. ($n = 9$). Because of the small *n*s for the behavioral and "other" orientations, purely behavioral clinicians were included with cognitive-behavioral and "other" orientations were lumped with the eclectic/integrative clinicians. The overextension rates for these two groups were not different for any combination.

A possible hypothesis is that the size of the concept is a determinant for the number of overextensions produced. Some participants circled high numbers of symptoms when describing the individual disorders. Thus, when describing the combinations, there are few symptoms left that could be potential overextensions. Table 3 presents the correlations between the percent of overextensions for a concept and the size of the overall concept and the size of its constituent concepts. All of the correlations are negative, indicating that when the constituent concepts are large, there are fewer overextensions. The one exception to the pattern was the size of the GAD-APD combination and the percent of overextensions generated for that combination. The lack

of a correlation demonstrates that even when there were many symptoms included, there was still a high number of overextensions which is consistent with this combination having a higher overall percentage of overextensions than the others.

In addition, participants were consistent in the number of symptoms they included across the study. Correlations among the sizes of the concepts (both single disorders and combinations) ranged from .57 to .88, with the mean correlation being .70 (*SD* = .09). In other words, if a participant circled a high number of symptoms for one description, that person was likely to circle a high number of symptoms across all stimuli.

Given the level of correlation between the size of the concepts and the percent of overextensions that occur, one would expect size to significantly predict the percent of overextensions. However, it may be that one of the constituents is a better predictor than the others. Thus, for each overextension variable, all of its relevant concept sizes were entered into stepwise regressions to determine if they add unique variability beyond the others. Stepwise regression selects the variable with the strongest relationship to the dependent measure and enters it into the model first and then determines if the other variables contribute unique variance to the prediction. For the percent of overextensions for the MDD-GAD combination, the model was significant, $F(1,68) = 14.95$, $p < .001$, $R^2 = .18$. The size of the original MDD concept was entered first (B = -.003, β = -.43, SE = .001) but neither the size of GAD nor the MDD-GAD combination added a significant amount of unique variance. For the MDD-APD combination, the model was significant, $F(1,68) = 25.37$, $p < .001$, $R^2 = .27$. The size of the original APD concept was the only variable selected (B = -.005, β = -.52, SE = .001). For the GAD-APD combination, the

overall model was again significant, $F(1,68) = 15.72$, $p < .001$, $R^2 = .19$, and only the size of the original APD concept was selected (B = -.006, β = -.43, SE = .001). Finally, for the three-way combination, the overall model was significant, $F(1,68) = 30.34$, $p < .001$, $R^2 = .31$. Even with an extra variable, the size of the original APD concept was selected as the only predictor (B = -.005, β = -.56, SE = .001). The results of these stepwise regressions indicate that the variability in the size of the constituent concepts is largely redundant, consistent with the high correlations across concept sizes. In other words, if a participant circled a lot of symptoms for one disorder, they did so for all disorders and disorder combinations. Nonetheless, when APD was included in the combination, the size of that concept was the best predictor of the percent of overextensions present.

Composition of the combinations

The conceptual combination literature has noted that when concepts are combined, sometimes one of the constituents will dominate the features of the combination. To examine if this result occurred within this study, the number of symptoms included in a combination from one of its constituents was divided by the total number of symptoms for the combination minus the number of overextensions. This procedure calculated the percent of symptoms in the combination that were included from the constituent concept while controlling for overextensions which were not a part of either. These values are presented in Table 4. Note that the combined percentage of symptoms from the constituents is greater than 100% because some symptoms overlapped across constituent concepts. For example, a participant may have included 6 symptoms in the descriptions of MDD and GAD that occurred in both. The pattern

revealed in Table 4 indicates that GAD was dominated by both MDD and APD in combinations, but MDD and APD exerted equal weight.

Normative symptom patterns

Because each individual symptom description was recorded, it is possible to examine the symptom pattern endorsed across clinicians. The symptoms that over 75% of clinicians endorsed are listed in Table 5 for each of the disorders and their combinations. Of note, clinicians selected noticeably fewer symptoms for the combination of GAD-APD compared to the consensus reached for other disorder combinations. This finding again reflects the heterogeneity of responses to this concept. Also, overextensions occurred even in this context. Some symptoms emerged as being endorsed by greater than 75% of the sample in the combinations that were not similarly endorsed in the constituent concepts.

Comparisons between ABCT and Florida

The two samples of ABCT and Florida clinicians were chosen to represent different possible sampling biases. ABCT clinicians are geographically diverse, but theoretically homogenous, whereby Florida clinicians are located within a particular region of the United States, but represent a broader range of theoretical orientations. If there are systematic differences between the samples in the way in which the participants combine concepts, then there may be geographic or theoretical effects.

The mean percent of overextensions for each disorder combination across samples are presented in Table 1. The percent of overextensions for each combination were entered into a MANOVA as dependent measures with sample as a between-subjects

factor. The two samples were not different in the percent of overextensions they produced for any of the combinations: MDD-GAD $F(1,68) = 0.89$, *ns*; MDD-APD $F(1,68) = 0.17$, *ns*; GAD-APD $F(1,68) = 0.04$, *ns*; and MDD-GAD-APD $F(1,68) = 1.52$, *ns*. All other analyses presented above produced identical results when the two samples were analyzed separately.

Reanalysis of overextensions

A limit of the forced-choice methodology is that participants are more likely to include symptoms by chance than by free recall. For instance, in reading the list of 120 symptoms provided, an inclusion that would not have otherwise occurred to a participant might seem like a feasible selection. One concern of the results presented previously is that overextensions could be selections that did not occur to a participant when describing an individual disorder but are nonetheless part of that individual's concept for the disorder. The participant omitted the symptom initially, but through repeated exposure to the list of 120, includes the symptom as an "overextension" in a combination. By this logic, the number of overextensions in the analyses above may be an overestimate.

Given that a major goal of the study is to determine if clinicians employ a multiplicative model of diagnostic combination, I recalculated the number of overextensions for each participant using a stringent correction for chance omission from the original concept. For each symptom, I calculated the percent of the total sample that included that symptom in their descriptions of each single disorder. If a symptom was included by 34% or more of participants, it was considered sufficiently likely. Thence, if a symptom that was overextended in a combination was included by more than one-third

of the participants in either of the constituent concepts, the overextension was considered to be a chance omission, as many participants considered it to be a part of the single disorder concept. It is important to note that this cutoff is arbitrary, and its adjustment would alter the results described below. This cutoff was chosen as a stringent test such that any overextensions that remained would be relatively unlikely members of either constituent.

The recalculated average percent of overextensions for each combination are included in Table 6. These values are substantially reduced. Nonetheless, when the percentages for each combination were entered into a MANOVA with a test matrix of zero, the overall Wilks' λ (4,66) = 0.30, $p < .001$. Individual tests for each variable demonstrated that the mean for each combination was not equal to zero: for MDD-GAD, $F(1,69) = 74.70$; for MDD-APD, $F(1,69) = 51.07$; for GAD-APD, $F(1,69) = 90.90$; and for the three-way combination, $F(1,69) = 72.87$, all *p*s < .001. Therefore, despite stringent exclusion criteria, a non-zero amount of overextension still occurred across each combination.

RESULTS: SECOND PHASE

Overall analyses

There was an unexpected ambiguity in the instruction set for the second phase task. Clinicians were instructed to rate a single disorder the same as for the first task. They were then told to add or subtract symptoms from that original list to reflect how the description would change by adding a comorbid condition. The ambiguity resulted from the addition of the final combination. To illustrate, consider the following example. A clinician would first select symptoms to describe MDD. That person would then add or subtract symptoms from that list to reflect the comorbid condition of MDD-GAD. The final addition would be changes for including APD. However, some clinicians interpreted this instruction to mean the combination of MDD-APD. Others considered it to mean MDD-GAD-APD. The specific instructions are quoted verbatim in the method section, and a careful review will note that there is no direction given to participants on this issue. There is some evidence in clinicians' responses to indicate which approach they followed. For example, in the third iteration a clinician may subtract a symptom that was added in the second iteration, indicating a three way approach. However, for most participants, their responses do not clearly indicate one way or the other which approach they may have followed. Therefore, the results are not easily interpretable and will not be analyzed here. Instead, a discussion of the limited interpretation of these data is included in Appendix C.

Test-retest reliability

Nonetheless, one aspect of the second phase of the study was unaffected by the ambiguity presented above. Participants' descriptions of the single disorders were not affected and are still directly comparable to their descriptions from the first phase. Thus, their responses can provide a measure of test-retest reliability for the task. There are many ways to estimate and calculate reliability. The kappa statistic provides the advantage of taking base rate information into consideration as participants varied in the number of symptoms they selected. An example of calculating kappa may best illustrate this point.

Assume a participant endorsed 21 symptoms for MDD in the first phase and 16 in the second phase. Some of these symptoms would be the same, some would be endorsed only at time 1, and some only at time 2. The distribution might be as follows:

		Phase 1	
		included	not included
Phase 2	included	15	1
	not included	6	98

The percent of overall agreement (P_O) between the two ratings is the number of symptoms included both times added to the number of symptoms included neither time divided by the total number of symptoms [(15 + 98)/120 = .94]. However, this number is an overestimation of the amount of agreement between the two times because it

capitalizes on chance agreement, i.e., the high number of symptoms that were not endorsed either time. Kappa accounts for the percent of chance agreement (P_C) by calculating the marginal probabilities of including a symptom or not as follows:

		Phase 1		
		included	not included	
Phase 2	included	15	1	16/120 = .133
	not included	6	98	104/120 = .877
		21/120 = .175	99/120 = .825	

$$P_C = (.175 * .133) + (.825 * .877) = .7468$$

In other words, given the distribution of symptoms included at the two times, there is nearly a 75% chance of agreement. Kappa is the ratio of the amount of observed agreement that exceeds chance and the amount of possible remaining agreement. The formula for calculated kappa is presented below:

$$\kappa = \frac{P_O - P_C}{1 - P_C}$$

Therefore, kappa represents the percent of agreement that remains once chance overlap has been removed. For this example:

$$\kappa = \frac{.94 - .75}{1 - .75} = .76$$

The median length of time between a participant's completion of the first and second phase was 24 days (IQR = 15.00 – 39.50). Thirty-seven participants completed the second phase of the study. The average kappa coefficient for MDD was .67 (*SD* =

0.13, range .33 - .88). For GAD, the average was .57 (*SD* = 0.13, range .33 - .82) and for APD it was .60 (*SD* = 0.14, range .25 - .82). These kappa coefficients are moderate and indicate fair consistency in participants' responses but also non-trivial variability.

DISCUSSION

There were several major findings in this study. First and foremost, overextensions occurred at non-zero rates, indicating that some degree of a multiplicative model is present in how clinicians conceptualize comorbid conditions. Second, there was a high degree of variability across clinicians in terms of the amount of overextension that occurred, indicating individual differences in the application of a multiplicative model. Third, as predicted by the literature on conceptual combination (Wisniewski, 1996), the similarity of the concepts did have an effect on how they were combined, although the effect was present only in the combination of GAD and APD. Fourth, also as predicted by that literature (Chater et al., 1990; Hampton, 1988; Storms et al., 1996), one concept (GAD) tended to be dominated such that not as many of its features appeared in combinations as did the other two disorders. Fifth, in the normative symptom pattern across all participants overextensions still occurred and the combination of GAD and APD stood out from the rest. Sixth, there was no difference in overextension rate between the two samples. Finally, the test-retest reliability of the task was moderate. Each of these findings will be discussed in turn.

Overall overextension rate

The main purpose of the study was to examine possible models of how clinicians combine comorbid conditions conceptually. As stated in the introduction of this paper, the current edition of the *DSM* implies an additive model whereby the combination of

disorders is equivalent to the addition of the symptoms of each constituent disorder. An alternative would be a multiplicative model, where the combination of two disorders is more than the simple addition of its parts. There are several ways a multiplicative model could be expressed. First, the severity of the pathology could magnify in combination. Second, the nature of the particular symptoms could be differentially expressed or influenced by a combination. For example, a change in sleep pattern is a symptom of Major Depressive Disorder. That change could be hypersomnia or insomnia. If combined with a certain type of disorder, like an anxiety disorder, that symptom may become more likely to be expressed in the direction of insomnia. A third possible expression of a multiplicative model is the presence of a feature in a combination that does not belong to either of its components, or an overextension. The current study explicitly examined the presence of overextensions as the most obvious indication of a multiplicative model of combination, although the other forms would be equally interesting. Given that an additive model would predict no overextensions, any degree of overextension is theoretically interesting.

In this study, clinicians overextended symptoms across combinations at a rate of approximately 14%. While the majority of the features of a combination are additive in nature, nonetheless some features emerge in a multiplicative fashion. This finding is directly related to previous research of conceptual combination where individuals use both property mapping and relation-linking strategies (Wisniewski, 1996; Wisniewski & Love, 1998). While the current study is analogous, it did not directly assess the combinatorial strategy of participants. Nonetheless, the findings of the current study

indicate that property mapping is likely a common strategy, as many symptoms in combinations were directly present in the constituent disorders. However, it would be interesting to assess participants' strategies directly, as the use of the terms "additive" and "multiplicative" are not synonymous with "property mapping" and "relation-linking," respectively. The first set of terms refers to only the outcome of the description, without assuming anything about the process, where the second set of terms explicitly assess the process of combination. Future studies may examine both the outcome and the process of clinicians' conceptual combination of mental disorders, further elucidating the phenomenon of observed overextensions.

The current diagnostic system does not have a means of accounting for the overextended portion of how clinicians conceptualize cases. Given the current push towards "clinical utility" in the *DSM-V* (First, 2005; First & Westen, 2007; First et al., 2004; Samuel & Widiger, 2006; Verheul, 2006), this lack has several implications for the basic functions of a diagnostic system. A major goal of any classification system is communication between professionals (Blashfield, Keeley, & Burgess, 2009). Take the example of a clinician who is transferring care of a client to another professional. Ideally, a diagnosis should summarize many, but not all, of the relevant features of a case. The relatively brief communication of "This patient has a diagnosis of X" is loaded with information about symptom expression, prognosis, associated factors, and possible treatment directions. However, if such communications are systematically omitting information when comorbidity is present, the diagnostic system is failing to achieve this basic purpose.

Another major goal of a classification system is description (Blashfield, Keeley, & Burgess, 2009). A classification system should be designed to capture the relevant features of the objects in its domain. If the system systematically ignores certain aspects of its domain, then it does not adequately describe the phenomenology of the world. It is important to note that this study was an examination of clinicians' perceptions of psychopathology, not of the psychopathology manifested by patients with comorbid conditions. Ideally, clinicians' perceptions should reflect the symptomotology that patients actually experience, although that may not necessarily be the case. A growing body of literature has documented changes in pathology that occur with comorbidity.

In a search of the PsycInfo database, 431 articles addressed the key words of comorbidity and psychopathology since January 2006. Among these were studies of every area of psychopathology, with particular emphasis on the childhood disorders and personality disorders. A few examples merit mention, as each address different ways in which a multiplicative model of combination could be expressed. First, Philipsen and colleagues (2008) found that individuals with Borderline Personality Disorder who had a history of childhood ADHD had increased difficulty controlling anger and more stress related dissociative or paranoid symptoms than Borderline individuals without a history of ADHD. This study demonstrates an increased likelihood of two particular symptoms of a disorder in the presence of a comorbid condition. Second, data collected as part of the Collaborative Longitudinal Personality Disorders Study demonstrated that Major Depressive Disorder was less likely to remit when the individual also had a comorbid personality disorder compared to depression alone (Markowitz et al., 2007). In this

example, the comorbid condition altered the course of the disorder. Third, Hurtig et al. (2007) found that children with ADHD and comorbid substance abuse, Conduct Disorder, or Oppositional Defiant Disorder exhibited more ADHD symptoms than a comparable group of children with ADHD alone. Here, comorbid conditions increased the average number of symptoms. Fourth, when Obsessive-compulsive Disorder (OCD) is accompanied by Major Depressive Disorder, the severity of the obsessions, compulsions, and depressive symptoms are worse than with uncomplicated OCD, and the nature of the obsessions are more likely to be aggressive (Besiroglu, Uguz, Saglam, Agargun, & Cilli, 2007). This example demonstrates a change in both the severity and the nature of symptoms in the presence of comorbid conditions.

A way in which the current study can contribute to this literature is through demonstrating that clinicians conceptualize comorbidity in a way that is not captured by many of the current paradigms for studying psychopathology. For example, structured interviews are a standard in the field for assessing psychopathology and establishing a diagnosis in studies. Structured interviews are valuable for the dramatic increase in diagnostic reliability across clinicians that they provide, which is a necessary component for drawing conclusions in a study. Without reliability in diagnosis, there can be no faith that the study addressed what it proposed to address. However, the structure that creates the increase in reliability also systematically excludes the possibility of documenting and accounting for overextended symptoms. Thus, the use of structured interviews, which was a necessary step for scientific advancement in the study of psychopathology, also limits the directions in which that study can progress. Studies that explicitly are designed

to study psychopathology in the context of comorbidity should consider other methods that may capture dimensions of pathology that would escape notice with structured interviews.

This point is highlighted by the fact that the studies of comorbid psychopathology cited above address the severity and nature of the symptoms present, but they do not address the possibility of overextended symptoms. This deficit is not a mere oversight in the literature, as at present the investigators of these studies do not employ methods capable of capturing the presence of overextended symptoms. The results of this study suggest that if they were to include extra means of assessing outside psychopathology, the investigators might find symptoms beyond those expected.

Variability Across Clinicians

While the data show that overextensions occurred across clinicians, the variability in clinicians' overextension rates was high. Some clinicians followed a truly additive model with no overextensions. For example, one clinician completed the task in a "perfect" additive model, where all symptoms that were included in a single disorder were present in any of its combinations, and no additional symptoms emerged. However, another included 57% overextensions in his description of GAD-APD. Participants widely varied in how much they overextended, and they were not necessarily consistent in that rate across different combinations. Clearly, there are some factors that contribute to why some clinicians included more overextensions than others.

However, many of the face-valid factors assessed in this study did not account for the differences in participants' responses. Variables like years of experience, familiarity

with the diagnosis, frequency of consulting the *DSM*, and theoretical orientation were not related to participants' overextension rate. The lack of a relationship is surprising, given that the majority of the literature on case conceptualization has found that experience (Brammer, 1997; Ladany, Marotta, & Muse-Burke, 2001; Lee & Tracey, 2008; Mayfield, Kardash, Kivlighan, 1999), use of *DSM* diagnosis (Falvey, 2001), and theoretical orientation (Constantine, 2001) lead to differences in the complexity of the conceptualization. It may be that there are qualitative differences in the cognitive processes that occur between single case conceptualization and comorbid conceptualization. Future studies should examine other possible factors that may explain the difference, like individual differences in creativity or personality.

Nonetheless, one factor did predict overextension rate. The number of symptoms a participant included in the combination and the single disorders significantly predicted overextension rate, such that participants that included higher numbers of symptoms included fewer overextensions. This effect appears to be an artifact of the methodology. If a participant includes a high number of symptoms initially, there are fewer possible alternatives to select as overextensions. Participants seemed to follow particular response sets, as they were consistent across the task in how many symptoms they included in their descriptions as evidenced by the average correlation of .70 for the number of symptoms included across descriptions. This finding may best be described as a ceiling effect, due to the fixed number of symptoms that could be used as descriptions. Thus, the overextension rates observed in this study may be underestimates, as a different, free-

response methodology may provide opportunity for an unrestricted number of overextensions to emerge.

One disorder combination remained immune to the methodology effect: GAD-APD. For most combinations, a relationship held such that including a high number of symptoms led to fewer overextensions. There was no relationship between the number of symptoms included for GAD-APD and the percent of overextensions, such that participants who included a high number of descriptions still maintained a high rate of overextension. As will be discussed more in the next section, this finding provides more evidence that the dissimilarity of this combination may have forced participants to change their strategy when describing it.

The Effect of Similarity

One original hypothesis of the study, based on the findings of Wisniewski (1996), was that similar concepts would be more likely to evidence additive models of combination, where participants would find alternative methods of explaining the relation between two dissimilar concepts, possibly through overextensions. Thus, the stimuli for the study included two similar concepts, MDD and GAD, and a concept that was dissimilar to both, APD. The overextension rate for the combination of GAD-APD was higher than for the other combinations.

There are a number of theoretical differences between the disorder concepts of GAD and APD. First, one is considered a longstanding, stable pattern of personality where the other is an acute, temporary condition, although emerging research has challenged this preconception (Clark, 2007; Grilo et al., 2004; Shea et al., 2002; Zanarini

et al., 2005). Second, GAD is considered an internalizing disorder, where symptoms are expressed inward and cognitive styles tend to blame oneself for misfortunes (Watson, 2005; Weiss, Susser, & Catron, 1998). On the other hand, APD is an externalizing disorder with the majority of symptoms being expressed externally and a cognitive style that blames others for problems (Kreuger et al., 2005; Maccoon & Newman, 2006). While this is not an exhaustive list, one line of research would consider the two disorders to be theoretical antitheses. Specifically, GAD is associated with moderate levels of autonomic arousal, which are interpreted and expressed as anxiety across a variety of situations (Andor, Gerlach, & Rist, 2008; Brown, Marten, & Barlow, 1995). Contrary, APD is associated with a low level of autonomic arousal, which has been proposed to lead to a variety of phenomena including stimulation seeking behavior and a lack of response to punishment (Farrington, 1997; Raine, 2002; Raine, Reynolds, Venables, & Mednick, 1997; Raine, Reynolds, Venables, & Mednick, & Farrington, 1998).

Three participants refused to describe the combination of GAD-APD, stating in essence that the two concepts were too dissimilar to combine. In a way, these participants were speaking to the similarity effect seen in other participants with higher rates of overextension. The highest individual rates of overextension seen in the study occurred in this combination. As stated above, the overextension rate of this combination was not affected by the number of descriptions included, further indicating that participants had to alter their strategy to account for the combination.

However, there was not an increased rate of overextension for the combination of MDD-APD or the three way combination. The lack of an effect does not seem to be due

to participants failing to see MDD and APD as dissimilar, as their ratings of the perceived similarity of the two followed the expected pattern. The lack of an effect must be due to some other factor. Perhaps, MDD and APD are more commonly seen in clinical practice as comorbid conditions than are the pair of GAD and APD. There is no evidence that MDD and APD differ in as fundamental a fashion as the arousal differences seen between GAD and APD. However, more study is necessary to determine what factor, if any, led to the difference seen in this study.

Concept Domination

The literature on conceptual combination has consistently found that when two concepts are combined, often the features of one concept dominate the combination (Chater et al., 1990; Hampton, 1988; Storms et al., 1996). In this study, both MDD and APD dominated GAD in their combinations. Participants consistently included fewer features of GAD in the combination compared to those coming from MDD or APD. This finding may be indicative of a general cognitive process, but it might also describe something particular about clinical cognition. In clinical work, often clinicians identify targets for treatment in a hierarchical fashion. For example, in the treatment of MDD and GAD, a clinician would likely focus on the symptoms of depression as they are potentially life threatening due to the risk of suicide, while symptoms of anxiety would receive secondary attention. The same process may be mirrored in this study, where the more “severe” disorders of MDD and APD were given consideration before GAD.

Also, each iteration of the *DSM* has had various exclusionary rules such that if one disorder was present, it would exclude the diagnosis of another. To use the example

of the *DSM-I* (APA, 1952), the structure of the manual followed a hierarchical decision tree that functioned as a set of exclusionary rules. A clinician started at the top of the hierarchy, and once a diagnosis was reached, the clinician did not continue through the remainder of the decision tree, thereby functionally excluding all diagnoses below that point. The *DSM-IV-TR* (APA, 2000) continues to have exclusionary criteria, but they are more to the effect of excluding diagnoses that may be accounted exclusively by the presence of another condition. For example, a diagnosis of schizophrenia is not given if psychotic symptoms occur exclusively in a period of substance intoxication or as a result of a general medical condition. While there are no explicit exclusionary criteria ruling out the possibility of using the three diagnostic categories included in this study, clinicians could be mirroring the process by excluding symptoms of GAD in favor of the other two disorders.

Normative Symptom Pattern

While this study was not an examination of psychopathology, it does provide some comment on clinicians' views of psychopathology. Specifically, due to the nature of the data, it was possible to examine clinicians' consensus of symptom patterns for the disorders and combinations used in the study. However, before discussing these results, it is important to note that the definition of consensus (i.e. 75% agreement) was arbitrary and the results would change with a different cut point. The cut point of 75% allowed for a majority while respecting the high degree of variability present in the data. Higher values dropped the number of agreed upon descriptions significantly, while lower values

may be considered over inclusive to chance effects and personal idiosyncrasies of clinicians.

The first feature of the descriptions to note in Table 5 is that there are few surprises in the descriptions of the single disorder concepts. For MDD, 12 of the 17 descriptions listed directly occur in the diagnostic criteria of the *DSM-IV-TR* (APA, 2000). The others, while not directly mentioned, are easily associated with the features included (easily discouraged, feel empty, lonely, pessimistic, and withdrawn). All criteria for a Major Depressive Episode were included in the consensus list of clinicians. For APD, 10 of the 13 symptoms were diagnostic criteria. Two of the remaining were alcohol abuse and substance use, which in the *DSM* system would describe commonly comorbid conditions of alcohol or substance dependence or abuse. The remaining symptom of interpersonal difficulty, while not directly included as a criterion, is a possible (but not necessary) consequence of many of the other criteria like deceitfulness, irresponsibility for social obligations, and lack of remorse for harming others. Interestingly, the only remaining *DSM* criterion which was not included by clinicians was impulsivity, which missed the threshold by being endorsed by only 70% of the sample. The consensus description of GAD did not conform as neatly to the *DSM* description as for the other two disorders. For GAD, 6 of the 10 symptoms included were DSM diagnostic criteria. The other four (cautious, heart palpitations, jittery, and performance anxiety) could reasonably be included as general descriptions of anxiety, although heart palpitations are usually associated with Panic Disorder rather than GAD. Three diagnostic criteria were not captured by clinicians' consensus descriptions: lack of control, lack of energy, and

irritability. While participants did not indicate that they were subjectively any less familiar with GAD than the other two disorders, the consensus data show that they did not capture the concept of GAD as well as MDD or APD as a group.

The next interesting feature of the normative disorder descriptions is that overextensions occurred even in this context. While admittedly few, a number of symptoms were included by over 75% of clinicians as a member of a combination that were not agreed upon as symptoms of the constituent single disorders. This result is an indication not only that overextensions occur on an individual basis, but that clinicians as a group employ a multiplicative model of combination.

Finally, once again the concept of GAD-APD does not follow the pattern of the other concepts. Clinicians agreed upon only 5 symptoms for the combination of GAD-APD, where the number of features included in all the other concepts was between 10 and 17. This small number reflects the increased heterogeneity of clinicians' descriptions of this particular combination. With other combinations, the majority of clinicians included a set of key features with some individual variability around those. However, with the combination of GAD-APD, clinicians had much more variability as to the symptoms they included, and there was not a core set of "key" features that emerged. In other words, participants' descriptions were much more idiosyncratic, perhaps indicating that participants struggled with trying to describe the two disparate concepts, and in so doing picked features on a more non-systematic basis.

Sample Differences

The two sample strategy used in this study was designed to counterbalance possible biases that may have affected the results. Specifically, the ABCT sample was theoretically homogenous. If only this sample was employed, overextension rate could feasibly be attributed to the effect of theoretical orientation. Similarly, if only the sample of Florida clinicians was used, the overextension rate could have been due to geographical differences in clinicians' training. However, the two samples were not different in their rate of overextension, speaking to the generalizability of the phenomenon. However, future studies should continue to use a variety of sampling strategies and populations to look for systematic differences in the ways clinicians conceptualize cases.

Reliability

The test-rest aspect of the follow-up portion of the study allowed for an estimate of participants' reliability in performing the task. By using the kappa statistic, it was possible to account for participants' base rates of including or excluding descriptions while factoring out chance agreement. The average kappa statistics indicate that participants' responses were moderately reliable across time, but also indicated non-trivial variability. This finding calls into question the consistency of participants' responses in the first portion of the study. In particular, the method of calculating overextensions relied upon the assumption that participants are mostly reliable in their descriptions of disorders, at least across the approximate hour it took to complete the first portion of the study. Part of the lower reliability coefficients is due to the effect of time,

with a median of almost a month between administrations. It is possible that another factor in the reliability is inconsistency in the descriptions of disorders, i.e. chance factors in selecting symptoms that could affect the results of a single administration. In other words, participants may have by chance omitted symptoms in the description of a single disorder, only to include them in the repeated administration of a combination. This chance omission would then appear to be an overextension.

Limitations

As just noted in the section on reliability, the greatest limitation of the results of this study is the possibility that some degree of the overextensions observed were chance omissions. If so, then the rate of overextension noted in the results would be an overestimate due to an artifact of the methodology. The structure of the task through sequential presentation of single disorders and combinations capitalizes on possible chance omissions. There are two possible means of correcting for this artifact: statistical and methodological.

One possible statistical correction was presented in the final section of the results of the first phase of the study. I recalculated the number of overextensions by excluding any overextensions that may have been likely chance omissions. An overextension was considered to be a chance omission if a third or more of participants included it as a description of either single disorder. Despite the variability in the symptoms clinicians' included, there were relatively few symptoms that were not included by at least a third of clinicians. Therefore, this redefinition is a stringent test of overextensions, as there are relatively few ways to overextend a symptom. Even in this context, some overextensions

remained, and participants included them at non-zero rates across each combination. Thus, by this test, overextensions are not solely a function of chance inclusion or omission.

It is also possible to examine chance omissions through altering the methodology. The follow-up study was designed to methodologically correct for these chance factors by having participants describe a single disorder and then explicitly alter that list by adding and subtracting symptoms when considering a comorbid condition. However, due to an unanticipated problem with the directions, the results of this portion of the study are limited. For a more complete description of these results, refer to Appendix C. Nonetheless, this methodology, with altered directions, provides promise for examining a more informed estimate of overextension rate.

A second major limitation of the study is the issue of order effects. While the stimuli in the first phase were randomized, the stimuli themselves did not represent all possible orders. For example, the combination of MDD-GAD appeared in different sequence relative to the other combinations, but the stimulus always presented the instruction to "Describe an individual with Major Depressive Disorder and Generalized Anxiety Disorder." The reverse order of the two disorders was not presented. It is reasonable to assume that the combination of the concepts may not be perfectly transitive, i.e. MDD with comorbid GAD is not necessarily the same as GAD with comorbid MDD. Indeed, some studies in the realm of conceptual combination have found that combinations are not transitive (Hampton, 1988).

The follow-up portion of the study was also designed to examine the transitivity of combination, as both possible orders were included. However, because the directions were unclear about the addition of disorders, some participants treated the addition of the third disorder as a three-way combination (e.g., MDD-GAD-APD) where others treated it as the second of two possible two-way combinations (MDD-APD, with GAD omitted). Again, these results are discussed in Appendix C, and with an adjustment in the directions the method may prove useful for illuminating the question of transitivity in combination.

Conclusions and Implications

This study was the first to examine clinicians' conceptualizations of comorbid conditions. While there are problems with the methodology (e.g. likely ceiling effects, moderate reliability), it nonetheless provides a useful start for examining a potentially fruitful area of research. In particular, the study examined a fundamental assumption of the classification system of mental disorders, i.e., that comorbid conditions are combined in an additive fashion. Informed by the literature on conceptual combination, this study operationalized an alternative, multiplicative model of combination through the presence of overextended symptoms. These symptoms occurred at non-trivial rates, implying that clinicians potentially utilize a multiplicative model of combination in conceptualizing comorbid cases. The current science of psychopathology has yet to search for overextended symptoms in the presence of comorbidity. Given the results of this study, it is an interesting and potentially paradigm shifting direction for future descriptive studies of psychopathology. Regardless, the current classification system does not always match

the way in which clinicians conceptualized comorbid cases in this study. As such, the clinical utility of the *DSM* is potentially compromised in any cases beyond single disorder descriptions. Much work is yet to be done before any concrete recommendations can be made for revising the structure of the classification system. This study is a first step towards challenging the conventional wisdom of the field and moving towards a more accurate agreement between clinicians' conceptualizations and diagnostic descriptions.

REFERENCES

American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th edition, text revision). Washington, DC: Author.

Andor, T., Gerlach, A., & Rist, F. (2008). Superior perception of phasic physiological arousal and the detrimental consequences of the conviction to be aroused on worrying and metacognitions in GAD. *Journal of Abnormal Psychology*, *117*, 193-205.

Berkson, J. (1946). Limitations of the application of the four-fold table analysis to hospital data. *Biometrics Bulletin*, *2*, 47-53.

Besiroglu, L., Uguz, F., Saglam, M., Agargun, M., & Cilli, A. (2007). Factors associated with major depressive disorder occurring after the onset of obsessive-compulsive disorder. *Journal of Affective Disorders*, *102*, 73-79.

Blashfield, R., Keeley, J., & Burgess, D. (2009). Classification. In P. H. Blaney & T. Millon (Eds.), *Oxford textbook of psychopathology* (pp. 35-57). New York: Oxford University Press.

Boyd, J. H., Burke, J. D., Gruenberg, E., Holzer, C. E., Rae, D. S., George, L. K., et al. (1984). Exclusion criteria of DSM-III: A study of co-occurrence of hierarchy-free syndromes. *Archives of General Psychiatry*, *41*, 983-989.

Brammer, R. (1997). Case conceptualization strategies: The relationship between psychologists' experience levels, academic training, and mode of clinical inquiry. *Educational Psychology Review*, *9*, 333-351.

Brown, T. A. (2007). Temporal course and structural relationships among dimensions of temperament and *DSM-IV* anxiety and mood disorder constructs. *Journal of Abnormal Psychology*, *116*, 313-328.

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationship among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, *107*, 179-192.

Brown, T A., Marten, P. A., & Barlow, D. H. (1995). Discriminant validity of the symptoms constituting the *DSM-III-R* and *DSM-IV* associated symptom criterion of generalized anxiety disorder. *Journal of Anxiety Disorders*, *9*, 317-328.

Chater, N., Lyon, K., Myers, T. (1990). Why are conjunctive categories overextended? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 497-508.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology*, *58*, 227-257.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*, 316-336.

Constantine, M. (2001). Multicultural training, theoretical orientation, empathy, and

multicultural case conceptualization ability in counselors. *Journal of Mental Health Counseling*, *23*, 357-372.

Cooper, J. E., Kendell, R. E., Gurland, B. J., Sharpe, L., Copeland, J. R. M., & Simon, R. (1972). *Psychiatric diagnosis in New York and London*. Maudsley Monograph No. 20, London: Oxford University Press.

Costa, P. T., & Widiger, T. A. (Eds.). (1994). *Personality disorders and the five-factor model of personality*. Washington, DC: American Psychological Association.

Costello, F. J., & Keane, M. T. (2001). Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 255-271.

du Fort, G. G., Newman, S. C., & Bland, R. C. (1993). Psychiatric comorbidity and treatment seeking: Sources of selection bias in the study of clinical populations. *Journal of Nervous and Mental Disease*, *181*, 467-474.

Falvey, J. (2001). Clinical judgment in case conceptualization and treatment planning across mental health disciplines. *Journal of Counseling and Development*, *79*, 292-303.

Farrington, D.P. (1997). The relationship between low resting heart rate and violence. In A. Raine, P.A. Brennan, D.P. Farrington, & S.A. Mednick (Eds.), *Biosocial bases of violence* (pp. 89–106). New York: Plenum.

Feinstein, A. R. (1970). The pre-therapeutic classification of co-morbidity in chronic disease. *Journal of Chronic Diseases*, *23*, 455-468.

First, M. B. (2005). Clinical utility: A prerequisite for the adoption of a dimensional approach in *DSM*. *Journal of Abnormal Psychology, 114*, 560-564.

First, M. B., Pincus, H. A., Levine, J. B., Williams, J. B. W., Ustun, B., & Peele, R. (2004). Clinical Utility as a Criterion for Revising Psychiatric Diagnoses. *American Journal of Psychiatry, 161*(6), 946-954.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1995). *Structured clinical interview for DSM-IV Axis I disorders*. New York: Biometrics Research Department.

First, M. B., & Westen, D. (2007). Classification for clinical practice: How to make *ICD* and *DSM* better able to serve clinicians. *International Review of Psychiatry*, *19*, 473-481.

Gagne, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 71-87.

Grilo, C., Sanislow, C., Gunderson, J., Pagano, M., Yen, S., et al. (2004). Two-year stability and change of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *Journal of Consulting and Clinical Psychology*, *72*, 767-775.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 12-32.

Houts, A. (2000). Fifty years of psychiatric nomenclature: Reflections on the 1943 War Department Technical Bulletin, Medical 203. *Journal of Clinical Psychology*, *56*, 935-967.

Hurtig, T., Ebeling, H., Taanila, A., Miettunen, J., Smalley, S., McGough, J., et al. (2007). ADHD and comorbid disorders in relation to family environment and symptom severity. *European Journal of Child and Adolescent Psychiatry*, *16*, 362-369.

Kendell, R. E., Cooper, J. E., Gourlay, A. J., Sharpe, L., & Gurland, B. J. (1971). Diagnostic criteria of American and British psychiatrist. *Archives of General Psychiatry*, *25*, 123-130.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 617-627.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eschlman, S., et al., (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry*, *51*, 8-19.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, *2*, 111-133.

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for *DSM-V*. *Journal of Abnormal Psychology*, *114*, 537-550.

Ladany, N., Marotta, S., & Muse-Burke, J. (2001). Counselor experience related to complexity of case conceptualization and supervision preference. *Counselor Education and Supervision*, *40*, 203-219.

Lee, D., & Tracey, T. (2008). General and multicultural case conceptualization skills: A cross-sectional analysis of psychotherapy trainees. *Psychotherapy Theory, Research, Practice, Training*, *45*, 507-522.

Lilienfeld, S. O. (2003). Comorbidity between and within childhood externalizing and internalizing disorders: Reflections and directions. *Journal of Abnormal Child Psychology*, *31*, 285-291.

Lilienfeld, S. O., Waldman, I. D., & Israel, A. C. (1994). A critical examination of the use of the term and concept of comorbidity in psychopathology research. *Clinical Psychology: Science and Practice*, *1*, 71-83.

Maccoon, D., & Newman, J. (2006). Content meets process: Using attributions and standards to inform cognitive vulnerability in psychopathy, antisocial personality disorder, and depression. *Journal of Social and Clinical Psychology*, *25*, 802-824.

Markowitz, J., Skodol, A., Petkova, E., Cheng, J., Sanislow, C., Grilo, C., et al. (2007). Longitudinal effects of personality disorders on psychosocial functioning of patients with major depressive disorder. *Journal of Clinical Psychiatry*, *68*, 186-193.

Mayfield, W., Kardash, C., & Kivlighan, D. (1999). Differences in experienced and novice counselors' knowledge structures about clients: Implications for case conceptualization. *Journal of Counseling Psychology*, *46*, 504-514.

Mineka, S., Watson, D., & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, *49*, 377-412.

Morey, L. C. (1991). *Personality Assessment Inventory Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Nylander, L., Lugnegard, T., & Hallerback, M. (2008). Autism spectrum disorders and schizophrenia spectrum disorders in adults: Is there a connection? A literature review and some suggestions for future clinical research. *Clinical Neuropsychiatry: Journal of Treatment Evaluation*, *5*, 43-54.

Olino, T., Klein, D., Lewinsohn, P., Rohde, P., & Seeley, J. (2008). Longitudinal associations between depressive and anxiety disorders: A comparison of two trait models. *Psychological Medicine*, *38*, 353-363.

Olsen, K., & Rosenbaum, B. (2006). Prospective investigations of the prodromal state of schizophrenia: Review of studies. Acta *Psychiatrica Scandinavica*, *113*, 247-272.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *11*, 35-58.

Philipsen, A., Limberger, M., Lieb, K., Feige, B., Kleindienst, N., Ebner-Priemer, U., et al. (2008). Attention-deficit hyperactivity disorder as a potentially aggrevating factor in borderline personality disorder. *The British Journal of Psychiatry*, *192*, 118-123.

Raine, A. (2002). Annotation: The role of prefrontal deficits, low autonomic arousal and early health factors in the development of antisocial and aggressive behavior in children. *Journal of Child Psychology and Psychiatry*, *43*, 417-434.

Raine, A., Reynolds, C., Venables, P.H., & Mednick, S.A. (1997). Resting heart rate, skin conductance orienting, and physique. In A. Raine, P.A. Brennan, D.P. Farrington, & S.A. Mednick (Eds.), *Biosocial bases of violence* (pp. 107–126). New York: Plenum.

Raine, A., Reynolds, C., Venables, P.H., & Mednick, S.A., & Farrington, D.P. (1998). Fearlessness, stimulation-seeking, and large body size at age 3 years as early predispositions to childhood aggression at age 11 years. *Archives of General Psychiatry*, *55*, 745–751.

Samuel, D., & Widiger, T. (2006). Clinicians' judgments of clinical utility: A comparison of the *DSM-IV* and five-factor models. *Journal of Abnormal Psychology*, *115*, 298-308.

Sellborn, M., Ben-Porath, Y., & Bagby, R. (2008). On the hierarchical structure of mood and anxiety disorders: Confirmatory evidence and elaboration of a model of temperament markers. *Journal of Abnormal Psychology*, *117*, 576-590.

Shea, M., Stout, R., Gunderson, J., Morey, L., Grilo, C., et al. (2002). Short-term diagnostic stability of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *American Journal of Psychiatry*, *159*, 2036-2041.

Siever, L., & Davis, K. (2004). The pathophysiology of schizophrenia disorders: Perspectives from the spectrum. *American Journal of Psychiatry*, *161*, 398-413.

Springer, K., & Murphy, G. L. (1992). Feature availability in conceptual combination. *Psychological Science*, *3*, 111-117.

Storms, G., De Boeck, P., Van Mechelen, I., & Ruts, W. (1996). The dominance effect in concept conjunctions: generality and interaction aspects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1266-1280.

Szatmari, P. (1992). The validity of autistic spectrum disorders: A literature review. *Journal of Autism and Developmental Disorders*, *22*, 583-600.

Tackett, J., Quilty, L., Sellborn, M., Rector, N., & Bagby, R. (2008). Additional evidence for a quantitative hierarchical model of mood and anxiety disorders for *DSM-V*. The context of personality structure. *Journal of Abnormal Psychology*, *117*, 812-825.

Verheul, R., (2006). Clinical utility of dimensional models for personality pathology. In T. Widiger, E. Simonsen, P. Sirovatka, & D. Regier (Eds.). *Dimensional models of personality disorders: Refining the research agenda for* DSM-V. Washington, DC: American Psychiatric Association.

Verheul, R., & Widiger, T. A. (2004). A meta-analysis of the prevalence and usage of the personality disorder not otherwise specified (PDNOS) diagnosis. *Journal of Personality Disorders*, *18*, 309-319.

Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for *DSM-V*. *Journal of Abnormal Psychology*, *114*, 522-536.

Weinstock, L, & Whisman, M. (2006). Neuroticism as a common feature of the depressive and anxiety disorders: A test of the revised integrative hierarchical model in a national sample. *Journal of Abnormal Psychology*, *115*, 68-74.

Weiss, B., Susser, K., & Catron, T. (1998). Common and specific features of childhood psychopathology. *Journal of Abnormal Psychology*, *107*, 118-127.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language*, *35*, 434-453.

Wisniewski, E. J., & Love, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language*, *38*, 177-202.

Wisniewski, E. J., & Murphy, G. L. (2005). Frequency of relation type as a determinant of conceptual combination: A reanalysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 169-174.

Wolf-Schein, E. (1996). The autistic spectrum disorder: A current review. *Developmental Disabilities Bulletin*, *24*, 33-55.

Zanarini, M., Frankenburg, F., Hennen, J., Reich, D., & Silk, K. (2005). The McLean Study of Adult Development (MSAD): Overview and implications of the first six years of prospective follow-up. *Journal of Personality Disorders*, *19*, 505-523.

Zimmerman, M., Rothschild, L., & Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, *162*, 1911-1918.

Table 1

The mean (and *SD*) percent of overextensions for each disorder combination

	Combination			
Sample	MDD-GAD	MDD-APD	GAD-APD	MDD-GAD-APD
ABCT (n = 45)	14.27 (9.21)	14.41 (10.65)	17.49 (11.90)	14.24 (10.72)
Florida (n = 25)	12.17 (8.45)	13.30 (10.98)	18.14 (16.86)	11.13 (8.88)
Total (n = 70)	13.52 (8.94)	14.01 (10.70)	17.72 (13.76)	13.13 (10.15)

Table 2

Pairwise comparisons of percent of overextension across disorder combinations

Comparison*	Test Value	η^2
MG vs MA	$F(1,69) = 0.19$, *ns*	.003
MG vs GA	$F(1,69) = 7.64$, $p < .01$	.10
MG vs MGA	$F(1,69) = 0.17$, *ns*	.002
MA vs GA	$F(1,69) = 8.61$, $p < .01$	.11
MA vs MGA	$F(1,69) = 1.18$, *ns*	.02
GA vs MGA	$F(1,69) = 9.57$, $p < .01$	.12

* MG = MDD-GAD, MA = MDD-APD, GA = GAD-APD, & MGA = MDD-GAD-APD

Table 3

Correlations between the percent of overextended symptoms and the size of constituent concepts

	r	*p*
MDD-GAD overextensions		
size of MDD	-.43	<.001
size of GAD	-.38	<.01
size of MDD-GAD	-.31	<.01
MDD-APD overextensions		
size of MDD	-.37	<.01
size of APD	-.52	<.001
size of MDD-APD	-.28	<.01
GAD-APD overextensions		
size of GAD	-.28	<.01
size of APD	-.43	<.001
size of GAD-APD	-.16	*ns*
MDD-GAD-APD overextensions		
size of MDD	-.46	<.001
size of GAD	-.43	<.001
size of APD	-.56	<.001
size of MDD-GAD-APD	-.40	<.01

Table 4

Means (and *SD*s) for the percent of each constituent concept included in a combination

	MDD	GAD	APD	Test
MDD-GAD	75.02	57.70		$F(1,69) = 55.07$, $p < .001$, $\eta^2 = .44$
	(11.95)	(14.02)		
MDD-APD	60.86		57.31	$F(1,69) = 1.30$, *ns*
	(14.47)		(16.15)	
GAD-APD		48.30	65.40	$F(1,69) = 20.28$, $p < .001$, $\eta^2 = .23$
		(17.59)	(17.16)	
MDD-GAD-APD	53.17	37.67	49.13	$F(2,138) = 18.66$, $p < .001$, $\eta^2 = .21$*
	(14.44)	(14.14)	(16.07)	

* Post-hoc comparisons across the three were MDD to GAD $F(1,69) = 70.87$, $p < .001$, $\eta^2 = .51$; MDD to APD $F(1,69) = 1.58$, *ns*; GAD to APD $F(1,69) = 18.50$, $p < .001$, $\eta^2 = .21$

Table 5

Normative (>75%) symptom patterns across the single disorders and combinations

MDD (17)	GAD (10)	APD (13)
anhedonia	apprehensive	alcohol abuse
apathy	cautious	breaks promises
change in appetite	difficulty relaxing	callous
easily discouraged	heart palpitations	deceptive
feel empty	jittery	interpersonal difficulty
feeling worthless	nervousness	legal trouble
irritable	performance anxiety	lies
lack of energy	tense	manipulative
lack of interest	trouble concentrating	physically aggressive
lonely	trouble sleeping	reckless behavior
pessimistic		self-centered
psychomotor retardation		substance use
sad		verbally aggressive
suicidal ideation		
trouble concentrating		
trouble sleeping		
withdrawn		

MDD-GAD (17)	MDD-APD (11)	GAD-APD (5)
anhedonia	alcohol abuse	alcohol abuse
apprehensive	anhedonia	difficulty relaxing
change in appetite	apathy	*irritable
difficulty relaxing	breaks promises	nervousness
easily discouraged	interpersonal difficulty	substance use
feel empty	irritable	
feeling worthless	legal trouble	
irritable	lies	
lack of energy	manipulative	
nervousness	trouble sleeping	
pessimistic	trouble concentrating	
*ruminative		
sad		
suicidal ideation		
tense		
trouble concentrating		
trouble sleeping		

Table 5 (cont.)

MDD-GAD-APD (17)

anhedonia
apathy
breaks promises
difficulty relaxing
feeling worthless
interpersonal difficulty
irritable
*lack of social support
legal trouble
manipulative
nervousness
*occupational problems
substance use
tense
trouble sleeping
trouble concentrating
*unstable mood

* Denotes a symptom in the combination consensus that was not a part of the single disorder consensus.

Table 6

The mean (and *SD*) percent of recalculated* overextensions for each disorder combination

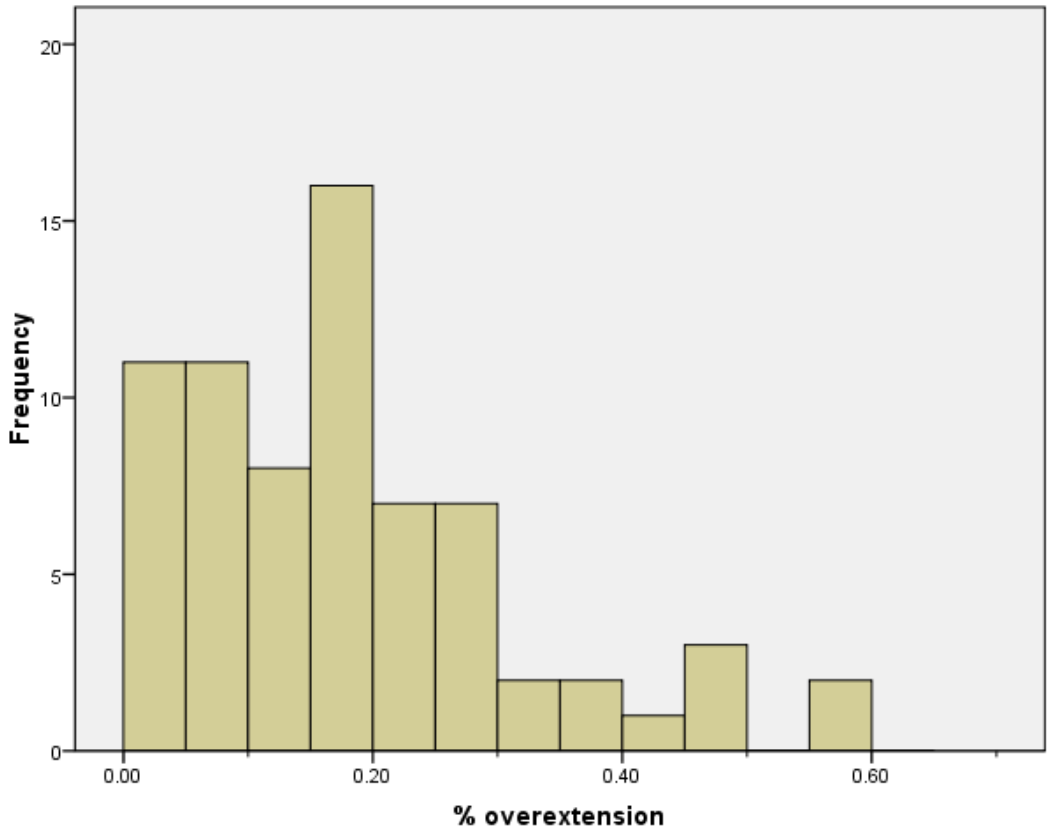
	Combination			
Sample	MDD-GAD	MDD-APD	GAD-APD	MDD-GAD-APD
ABCT (n = 45)	3.82 (3.84)	2.68 (3.23)	5.89 (5.12)	2.48 (2.37)
Florida (n = 25)	4.54 (4.18)	3.33 (3.74)	6.38 (5.76)	2.02 (2.09)
Total (n = 70)	4.08 (3.95)	2.91 (3.41)	6.06 (5.32)	2.31 (2.27)

*Overextensions were recalculated by only including overextensions that were endorsed as symptoms by the total sample in either single disorder less than 34% of the time.

Figure Caption

Figure 1. Distribution of the percent of overextensions for the combination of GAD-APD.

Figure 1



APPENDIX A: METHODOLOGY PILOT STUDY

Prior to the current study, pilot work was conducted to examine the feasibility of two possible methodologies. In previous work, graduate student clinicians were given a set of disorders and disorder combinations and asked to generate descriptions of these stimuli. Participants were allowed to write as many descriptions as they saw fit. However, the translation of these data into a quantifiable metric proved difficult. The coding of overextensions was very unreliable (only 35.70% agreement). Thus, the current pilot study was designed to compare a free response method, identical to that used previously, to a forced choice method in a sample of practicing clinicians.

Method

Participants

Participants were 10 clinicians in attendance at the Alabama Psychological Association convention held in June of 2007. One participant was dropped from the analyses for failing to complete all experimental materials. Of the remaining 9, the clinicians were an average of 52.44 years of age (SD = 11.64) with an average of 27.11 years of experience (SD = 9.51). Most (n = 6) clinicians were female. Two-thirds endorsed working with adults, 89% had experience working with adolescents, and 67% worked with children. Four clinicians responded that they consult the *DSM* on a daily basis, another 4 on a weekly basis, with the last clinician consulting the *DSM* monthly. Participants were compensated by being entered into a $50 drawing.

Materials

The stimuli for the study were three common psychiatric disorders, all possible pairwise combinations of the disorders, and the three-way combination. The disorders included were Major Depressive Disorder (MDD), Panic Disorder (PD), and Antisocial Personality Disorder (APD). The rationale for including these three disorders was twofold. First, all three disorders had been used in our previous work on conceptual combination with graduate students. In an effort to limit the difficulty and time commitment of the task, only three disorders were used. Second, given that the similarity of the stimuli is the best predictor of how the concepts will be combined, these three stimuli represent what we consider to be two very similar disorders (MDD and PD) and a disorder that is dissimilar to both (APD). MDD and PD are commonly comorbid conditions (Kessler et al., 2005; Kessler et al., 1994). Further, recent theoretical and empirical work has claimed that they are two expressions of the same negative affect syndrome, and are not separate psychopathological conditions (Brown, Chorpita, & Barlow, 1998; Clark & Watson, 1991; Mineka, Watson, & Clark, 1998). While MDD and PD are conceptualized as internalizing disorders, or symptoms involving an inward expression, APD is seen as an externalizing disorder, i.e., symptoms are expressed outward (Krueger, Markon, Patrick, & Iacono, 2005).

Free response task. The stimuli were presented in one of two methodologies. The first was a free response task. Each stimulus was presented on a separate page. At the top of the page "An individual with …" was presented in bold type, where the ellipsis was

replaced with the name of the disorder or disorder combination. This phrase was included to direct the participant to think of exemplars of the category rather than some abstracted prototype of the category. The participant received the following instructions:

> *On the following pages, you will find the name of a disorder or multiple disorders. Imagine what it would be like to see a person with those diagnoses in therapy. Based on your clinical experience, you are to list as many* ***symptoms/expressions*** *of the disorder or group of disorders as you can think of. Please limit each symptom description to three words or less. Your goal is not to reproduce the criteria in the DSM, but to list the clinically relevant features of each disorder based on your experience. You have as much time as you need to work. This is not a test, and there are no "right" answers.*

The instructions were followed by a content irrelevant example designed to illustrate that symptoms need not be limited to one modality (e.g. behavioral description, demographics, level of functioning, affect, interpersonal relations).

Forced choice task. Similar to the free response task, stimuli were presented each on a separate page in the same format. However, participants received the following instructions:

> *On the following pages, you will find the name of a disorder or multiple disorders. Imagine what it would be like to see a person with those diagnoses in therapy. On the same page you will find a list of descriptions. Based on your clinical experience, circle as many* ***symptoms/expressions*** *as you feel fit to describe the disorder or group of disorders. Your goal is not to reproduce the criteria in the DSM, but to list the clinically relevant features of each disorder based on your experience. You have as much time as you need to work. This is not a test, and there are no "right" answers.*

Under the name of the disorder on each page was a list of 120 symptoms. These symptoms were chosen to be representative of the entire domain of psychopathology and personality functioning, as they are drawn from a major psychopathology assessment

instrument (the *Personality Assessment Inventory* [*PAI*], Morey, 1991) and the currently most investigated theory of personality functioning that has been applied to the personality disorders, the Big Five personality factors (Costa & Widiger, 1994). Of the 120 symptoms included, 60 symptoms came from the *PAI* and 60 came from the Big Five factors. For the *PAI*, the 60 symptoms were taken from the clinical and treatment scales, excluding items on the validity and interpersonal scales. The 15 clinical and treatment scales are represented by 343 items. The 60 items selected for inclusion were those from each scale that had the highest loadings on their respective scale based on factor analytic studies. An equal number of items were selected from each scale, assuming that the items were not directly redundant. Items were rephrased to be one to three words in length. The remaining 60 items came from the Big Five personality factors. Each of the five factors has six facets. A representative descriptor for each tail of each facet was included (2 tails x 6 facets x 5 factors = 60 items). The 120 descriptions were presented in alphabetical order. They are included for reference in Appendix B.

Demographic questionnaire. Participants also completed a demographic questionnaire, assessing their age, sex, years of clinical experience, domain of experience (adults, adolescents, children), and frequency with which they consult the DSM. In addition, participants were asked to rate their familiarity with each of the disorders included in the study as well as the perceived similarity of the possible pairs.

Procedure

Participants were recruited by the primary investigator at the registration table. The experimenter sat next to a sign advertising the study. After receiving informed

consent information and agreeing to participate, each participant completed the study at his or her leisure after receiving the materials. Participants were instructed to complete the task at one time, rather than completing part and returning to it later. Each participant was randomly assigned to one of the two methodologies. The packet was arranged such that the demographic questionnaire was the first page. The instructions for the relevant task followed. Next, each of the individual disorders were presented in the same order (MDD, PD, APD) in the hope that each disorder would be equally primed or activated in the participants' memory. Then, all the possible pairwise comparisons and the three-way comparison were presented in a randomized order. This design was meant to minimize any potential ordering effects on symptom generation.

Coding

For the forced choice task, no coding was necessary. Each symptom circled for each stimulus was recorded and entered for analysis. For the free response task, the primary investigator coded each response using the forced choice symptom list. The closest possible match to the description given by the participant was recorded. For example, participants often listed "hopeless" as a symptom of MDD. This word does not directly occur in the list of 120 symptoms, but it is similar to "pessimistic." Thus, a response of "hopeless" would be coded as "pessimistic," and other similar decisions were necessary for fitting participants' responses into the coding scheme. For the purpose of the pilot, only one rater coded responses. However, in the previous work mentioned above, when multiple raters were used, reliability for this sort of task was unacceptable. Further, often there was no matching term on the forced choice list. These terms were

recorded and given a number, thereby expanding the coding list for future responses. The 120 symptom list necessarily did not cover all possible expressions of a disorder, and breadth of coverage of all psychopathology was given priority over complete coverage of the *DSM* defined symptoms of each disorder included in the study.

Results

The primary purpose of the pilot study was to determine the feasibility of each methodology along with the utility of the resulting data. Qualitatively, participants in the free response methodology offered much more interesting data. Regarding clinicians' models of conceptual combination, overextensions, or symptoms which do not occur in either constituent concept, are indicative of a multiplicative model. Participants in the free response method offered overextensions in forms that were not easily captured in a direct symptom coding. Often, in the combination, the participant would describe how the nature of the symptom changed relative to its expression in the individual disorder. For example, a participant might include anxiety as a symptom of PD, but in the combination of PD with APD state that the anxiety is related to illegal activity. Strictly speaking, in terms of coding symptoms, the aforementioned example is not an overextension, as both symptoms may have been included in the individual lists. However, the participant's response indicates how the concepts are being combined.

In contrast, the data from the forced choice task are readily quantifiable, and offer a plethora of equally interesting analysis possibilities. Given that this is an exploratory study, the qualitative nature of the free response data may be counterproductive. While the free response data are intriguing, they may be better served once a basic theoretical

context for the phenomena has already been established. Outside of that context, they are suggestive rather than determinate. Therefore, the current study will employ the forced choice methodology and the free response method will be reserved for future studies.

The forced choice data from the five participants in the pilot provide interesting results. First, Table 1 shows the percentage of overextensions each participant provided for each combination. Overall, participants are overextending the combinations to a small degree. However, the variability across participants is large, as is the within-participant variability across combinations. Nonetheless, these data suggest a few potential patterns. First, participants overextended a greater percentage of symptoms for the combination of PD and APD than the other combinations. According to the proposed method of combination proposed by Wisniewski (1996), this effect may be due to the dissimilarity between the two concepts. However, five participants are an insufficient sample size to conduct inferential tests to determine if the number of overextensions in the PD-APD combination are indeed greater than the others. Further, the sample is too small to conduct regression analyses to examine if factors such as perceived similarity predict the rate of overextension.

As noted in the review above, sometimes one concept will dominate a combination. Table 2 presents the percentage of symptoms in the combination that came from each constituent. From an initial examination of the Table, it appears that MDD tends to dominate combinations in which it is included, which is evidenced by a higher percentage of the combination coming from MDD symptomotology. It is important to note that these percentages are computed from raw frequencies and do not take the

relative sizes of the original constituents into account. For example, participants usually included more symptoms in MDD (*M* = 30.80, *SD* = 7.63) than either PD (*M* = 17.40, *SD* = 6.99) or APD (*M* = 22.40, *SD* = 8.65) as single disorders. The increased presence of MDD in the combinations could simply be a result of having more possible symptoms to include. Table 3 takes an opposite approach and displays the percent of symptoms included in the original disorder that were used in the combination. These data are less clear, and instead indicate that each constituent disorder is equally utilized in producing the combinations.

Finally, it is also possible to examine the overall pattern of symptom descriptions across participants. In that sense, we can examine which symptoms clinicians agree belong to each single disorder and each combination. Table 4 presents the symptoms that at least 4 of the clinicians included for each stimulus. Across all clinicians, some symptoms were consistently overextended in the combinations (as denoted by an asterisk).

Appendix A - Table 1

The percentage of overextensions for each participant across each disorder combination

Participant	MDD-PD	MDD-APD	PD-APD	MDD-PD-APD	Mean
1	.22	.20	.00	.13	.14
2	.00	.10	.40	.13	.16
3	.35	.34	.60	.35	.41
4	.12	.05	.05	.05	.07
5	.00	.00	.00	.00	.00
Mean	.14	.14	.21	.13	.16

Appendix A - Table 2

Percentage of symptoms in the combination included in each constituent

P	MDD-PD		MDD-APD		PD-APD		MDD-PD-APD		
	MDD	PD	MDD	APD	PD	APD	MDD	PD	APD
1	.67	.28	.40	.47	.33	.67	.38	.13	.44
2	.75	.63	.81	.10	.40	.24	.54	.33	.27
3	.53	.18	.52	.21	.27	.13	.39	.17	.20
4	.71	.55	.67	.54	.55	.57	.65	.43	.48
5	1.00	.92	1.00	1.00	1.00	1.00	1.00	.86	1.00
Mean	.73	.51	.68	.46	.51	.52	.59	.38	.48

Appendix A - Table 3

Percentage of symptoms from the constituent disorder that are used in the combination

P	MDD-PD		MDD-APD		PD-APD		MDD-PD-APD		
	MDD	PD	MDD	APD	PD	APD	MDD	PD	APD
1	.55	.56	.27	.39	.56	.56	.27	.22	.39
2	.58	.93	.55	.11	.56	.40	.84	.89	.87
3	.72	.39	.52	.32	.44	.21	.62	.44	.47
4	.81	.96	.95	.89	.86	.68	.95	.96	.81
5	.41	.79	.28	.35	.43	.26	.24	.43	.30
Mean	.61	.73	.51	.41	.57	.42	.58	.59	.57

Appendix A - Table 4

Most frequent symptoms for each disorder or combination

MDD (58)
alcohol abuse
anhedonia
apathy
change in appetite
difficulty relaxing
easily discouraged
feel empty
feeling worthless
health problems
lack of energy
lack of interest
lonely
psychomotor retardation
sad
suicidal ideation
trouble concentrating
trouble sleeping
withdrawn

PD (42)
cautious
difficulty relaxing
nervousness
panic
ruminative
tense

APD (52)
callous
deceptive
family trouble
legal trouble
manipulative
occupational problems
self-centered
sensation seeking
stormy relationships

MDD.PD (61)
anhedonia
change in appetite
difficulty relaxing
feeling worthless
health problems
nervousness
*occupational problems
panic
ruminative
*substance use
suicidal ideation
trouble concentrating

MDD.APD (72)
feeling worthless
*interpersonal difficulty
*substance use
trouble sleeping

PD.APD (62)
panic
tense

MDD.PD.APD (78)
alcohol abuse
change in appetite
*interpersonal difficulty
panic
*self-harm
stormy relationship
*substance use
suicidal ideation
trouble concentrating
trouble sleeping

Note: The number in parentheses next to each stimulus is the total number of symptoms endorsed for the disorder; an * denotes an overextended symptom given those included in the single disorder description.

APPENDIX B: FORCED-CHOICE SYMPTOM LIST

accomplished
aimless
alcohol abuse
anhedonia
apathy
apprehensive
art lover
assertive
auditory hallucinations
austere
breaks promises
broad-minded
callous
calm
capable
careless spending
cautious
change in appetite
cheerful
close-minded
compulsions
conceited
contentious
conventional
cooperative
creative
cynicism
daring
deceptive
dedicated
delusions
dependable
difficulty relaxing
disinterested in culture
disorganized
dull
easily discouraged
easygoing
empathetic
energetic
even-tempered
expressive emotionally
family trouble
fear of abandonment
feel empty
feeling worthless
financial problems
generous
glib
grandiosity
gregarious
gullible
headaches
health problems
heart palpitations
humble
impulsive
incompetent
inflexible
innovative
intense affect
interpersonal difficulty
irritable
jittery
lack of energy
lack of friends
lack of interest
lack of social support
legal trouble
leisurely
lies
lonely
manipulative
nausea
nervousness
nightmares
no close relationships
not easily discouraged
occupational problems
organized
overcontrolled
panic
paranoia
performance anxiety
pessimistic
physically aggressive
practical
psychomotor retardation
rapid thoughts
reckless behavior
recurrent memories
reflective
reserved
resilient
restrained affect
ruminative
sad
self-centered
self-conscious
self-harm
sensation seeking
sincere
spontaneous
stormy relationships
submissive
substance use
suicidal ideation
tempermental
tense
thought disorder
trouble concentrating
trouble sleeping
uncontrolled temper
unpredictable

unstable home life	verbally aggressive	warm
unstable mood	vulnerable	withdrawn

APPENDIX C: SECOND PHASE RESULTS AND DISCUSSION

Coding of response style

As stated in the main body of the paper, participants approached the second phase task in different ways. The task was structured such that participants first completed a description of a single disorder (e.g., MDD), followed by the addition of a second disorder (e.g., GAD), and completed with the addition of the third (e.g., APD). Due to the ambiguity present in the instructions, participants could approach the task in two ways: (a) a two-way combination of disorders (MDD-GAD followed by MDD-APD), or (b) a three-way combination of disorders (MDD-GAD followed by MDD-GAD-APD). It is impossible to know with certainty what approach participants used; however, there are some sources of evidence that differentiate the possibilities. The most straightforward source of evidence was a "three-way subtraction" where a participant would subtract a symptom in the third iteration that was not present in the original description but added in the second step. A second source of information was an "over-included" symptom, or a symptom that was added in the second or third iteration which was already included in the original description. Occasionally, a participant would over-include a symptom in the third iteration that was subtracted in the second, demonstrating a three-way approach. On the converse, sometimes a participant would add the same symptom in the second and third iteration, indicating a two-way approach. However, a number of participants did not evidence any of the signs listed above, presenting an ambiguous two- or three-way

approach. Thus, participants' approach to the task could be coded as two-way (n = 12), three-way (n = 14), or ambiguous (n = 11).

Relationship of response style to factors from the first phase

Members from each sample (ABCT n = 25 and Florida n = 12) were equally likely to follow any of the response styles ($\chi^2(2) = 0.21$, *ns*). Furthermore, response style was not associated with a clinician's theoretical orientation, work setting, or frequency of consulting the *DSM*. However, participants who followed a two-way strategy were older (M = 53.08, SD = 7.84) and more experienced (M = 27.27, SD = 10.36) than either those who used a three-way strategy (age M = 41.43, SD = 11.00; experience M = 17.79, SD = 10.25) or were ambiguous (age M = 40.64, SD = 6.92; experience M = 16.64, SD = 7.42); for age $F(2,34) = 7.30$, $p < .01$, for experience $F(2,33) = 4.25$, $p < .05$. Post-hoc comparisons using the Tukey error-correction demonstrate that for both age and experience the two-way participants were significantly different from the three-way and ambiguous participants, who were equal to each other. Participants did not differ on their familiarity with the diagnoses, number of symptoms they included in their first phase descriptions, or first phase overextension rate.

Overextension rate

Just as in the first portion of the study, the percent of overextensions in a combined concept is the primary dependent measure rather than the number of overextensions. However, because of the change in the nature of the task, the percent of overextensions is calculated differently. Participants were asked to add and subtract symptoms for the combination from an original list describing a single disorder.

Therefore, the total number of symptoms representing the combined concept is the number for the original disorder plus the symptoms added in the next step minus those subtracted in the next step. Using this definition necessarily excludes the participants who followed a three-way strategy when completing the task. The analyses below will first present only those participants whom the experimenter is reasonably sure completed the task using a two-way strategy. Second, the participants who were judged to be ambiguous between a two- and three-way strategy will be added to increase power. However, because of the uncertainty as to how participants completed the task, these results can only be interpreted in a very limited fashion and may not be likely to replicate.

The percent of overextensions for each possible combination for the two-way participants ($n = 12$) were entered as dependent measures into a MANOVA. One participant was excluded from the analysis for omitting descriptions of GAD-APD and APD-GAD. That participant also did not complete the description of that combination in the first portion of the study. Table 1 presents the means and standard deviations for the remaining two-way participants across each disorder combination. The overall test approached statistical significance; Wilk's λ (6,5) = 0.15, $p = .055$, $\eta^2 = .85$; but each individual variable was significantly different from a value of zero (see Table 1). The values are less than those seen in the first phase with no error correction, but are higher than those with the conservative recalculation. Similar to the first phase, the standard deviations represent a high degree of variability across participants.

When the ambiguous participants ($n = 11$) are added to the analysis, the overall test becomes significant; Wilk's λ (6,16) = 0.41, $p < .05$, $\eta^2 = .59$. The means, standard

deviations, and individual tests compared to a value of zero are shown in Table 2. Notice that the values do not change drastically except for the combination of APD-GAD. There is one participant in the ambiguous-style group who included 81% overextensions in the combination, pushing the mean value higher. Despite this outlier, a test between the two-way and ambiguous participants does not reveal a statistically significant difference; Wilk's λ (6,15) = 0.70, *ns*; with all comparisons for each variable also non-significant.

Overall, when participants are explicitly instructed to add symptoms in a combination, overextensions still occur. In the first phase, a major limitation was that overextensions could have been chance selections or symptoms that were mistakenly forgotten in an earlier description of a single disorder. The second phase presents a methodological correction for this possibility. In this study, participants still included non-zero rates of overextension. While these rates are not as high as those seen in the first phase generally, they were higher than the conservative correction of including only overextensions that were included by less than 34% of the total sample in describing a disorder. Thus, that correction likely underestimated the true number of overextensions, while the general analyses likely included some unintentional overextensions.

Comparison of overextension rate between phases 1 and 2

It is also possible to examine a direct comparison of participants' overextension rates between the first and second phases using a series of paired-samples *t*-tests. Because there was no significant difference between the overextension rates in phase 2 for the two-way and ambiguous participants, they are combined in this analysis to increase sample size. Further, as the second phase participants completed both possible orders,

only the comparable order was used in the analysis. Table 3 presents the means, standard deviations, and test values for the comparisons. The mean difference across the first phase and the second phase was not significant for MDD-GAD, but it approached significance for MDD-APD (p = .052). However, participants included a higher percentage of overextensions in the first phase for GAD-APD than for the second. This comparison is direct change from phase 1 to 2 for individual participants, rather than an examination of overall mean differences across the two times, which include different samples. Therefore, while the overall percentages dropped from the first to second phase, that change may represent a drop-out bias. In other words, participants who included higher percentages of overextensions, thereby raising the means in the first phase, did not participate in the second phase. The rates may be more comparable with a larger sample size. Nonetheless, there might continue to be a drop in overextension rate that is an artifact of the methodology.

Concept transitivity

The first phase failed to address the transitivity of the combinations. For example, participants might consider MDD with comorbid GAD to be different from GAD with comorbid MDD. By presenting both possibilities, the second phase methodology is able to examine if clinicians considered the different orders to be the same or not. Some literature addressing conceptual combination in general has found that people do not describe transitive combinations the same way (Hampton, 1988). Because the transitivity of concepts can only be assured in the two-way participants, only they will be analyzed here.

Transitivity was measured by creating a total list of symptoms for a combination by taking the list for the original concept and then adding and subtracting symptoms as indicated by the participant. Kappa coefficients were then calculated for each transitive pair (see the main paper for a thorough description and discussion of kappa). Table 4 presents the mean, standard deviation, and range for the kappa coefficients of the 12 two-way participants. The one participant that did not include a description of GAD-APD or APD-GAD was included, but the value for that pairing was omitted. The kappa values indicate that participants did not treat the concepts in a perfectly transitive fashion. Nevertheless, the values are above .5, indicating that participants were more consistent than not. A custom-hypothesis MANOVA comparing the three variables found no difference in their mean values, Wilk's λ (2,9) = 0.76, *ns*. Interestingly, the mean kappa values are comparable to the test-retest values described in the main body of the paper. This finding may indicate that the transitivity values are bound by an upper limit of consistency for the task overall.

Despite equivalent means, the standard deviations and ranges indicate that there was greater variability in kappa coefficients for combinations involving APD than for the pairing of MDD and GAD. Indeed, the range and standard deviation were highest for the pairing of GAD and APD, consistent with findings in the first phase that participants treated this pair of disorders differently from other stimuli.

Conclusions and future directions

The second phase methodology presented some unanticipated challenges for interpreting the data. Nonetheless, the data provide a useful pilot of the methodology and

hint at the applicability of the findings. The overextension rate changed from the first phase methodology to the second, but the change may not represent a statistically significant change for individual participants. If the findings are taken at face value, then the first phase methodology overestimates overextension rate, but overextensions still exist in non-zero amounts in the second phase. Further, the second phase findings indicate that clinicians do not treat the combination of disorders transitively. For example, GAD with comorbid APD is not the same as APD with comorbid GAD. It would be interesting if that result were to replicate with a larger sample. Overall, with a clarification of the directions, the second phase methodology holds promise for elucidating some of the questions raised by the first phase.

Appendix C – Table 1

Means and standard deviations for the percent of overextensions across disorder combination in Study 2 for two-way participants

Combination	Mean	*SD*	Test	η^2
MDD-GAD	8.12	8.02	$F(1,10) = 11.28$, $p < .01$	.53
MDD-APD	8.56	7.19	$F(1,10) = 15.63$, $p < .01$	.61
GAD-MDD	7.70	5.51	$F(1,10) = 21.48$, $p < .001$	.68
GAD-APD	10.55	9.51	$F(1,10) = 13.54$, $p < .01$	.58
APD-MDD	6.40	5.55	$F(1,10) = 14.60$, $p < .01$	.59
APD-GAD	8.53	11.36	$F(1,10) = 6.20$, $p < .05$	.38

Appendix C – Table 2

Means and standard deviations for the percent of overextensions across disorder combination in Study 2 for two-way and ambiguous participants

Combination	Mean	*SD*	Test	η^2
MDD-GAD	9.02	11.57	$F(1,21) = 13.35$, $p < .001$	.39
MDD-APD	8.49	11.75	$F(1,21) = 11.49$, $p < .01$	.35
GAD-MDD	7.30	9.76	$F(1,21) = 12.32$, $p < .01$	.37
GAD-APD	9.19	9.42	$F(1,21) = 20.91$, $p < .001$	.50
APD-MDD	6.73	8.25	$F(1,21) = 14.66$, $p < .001$	.41
APD-GAD	10.09	18.20	$F(1,21) = 6.77$, $p < .05$	.24

Appendix C – Table 3

Comparison of mean overextension rates across Studies 1 and 2

	Mean		SD		Test
	Study 1	Study 2	Study 1	Study 2	
MDD-GAD	10.70	8.63	7.32	11.46	$t(22) = 0.72$, *ns*
MDD-APD	14.51	8.35	11.02	11.50	$t(22) = 2.05$, *ns*
GAD-APD	17.30	9.19	13.39	9.42	$t(21) = 2.98$, $p < .01$

Appendix C – Table 4

Descriptive statistics for kappa coefficients for transitive combinations

	Mean	SD	Minimum	Maximum
MDD & GAD	.67	.13	.50	.98
MDD & APD	.60	.19	.29	.98
GAD & APD	.57	.24	.17	1.00