

# The Sweet Spot of Clinical Intuitions: Predictors of the Effects of Context on Impressions of Conduct Disorder Symptoms

Jessecae K. Marsh and Christopher T. Burke  
Lehigh University

Andres De Los Reyes  
University of Maryland at College Park

How people interpret a mental disorder symptom has been shown to depend on the contextual life factors surrounding its presentation. Specifically, people are more likely to judge a symptom as clinically relevant if that symptom presents in a high-risk environment (e.g., child associates with deviant peers) relative to a low-risk environment (e.g., child associates with normative peer group). Importantly, not all symptoms are influenced by context to the same extent, and there is low agreement across people as to how this influence manifests. In this paper, we explore what factors predict the extent to which clinicians and laypeople interpret mental disorder symptoms as a function of diagnosis-congruent versus incongruent contextual information. We tested the impact of 2 statistical factors (prevalence and diagnosticity) and 2 more intuitive factors (diagnostic importance and abnormality) on the degree to which a symptom is interpreted differently in different contexts. Clinicians' impressions of the diagnosticity and importance of a symptom evidenced a curvilinear relationship with the use of context, with extremely important and unimportant as well as extremely diagnostic and nondiagnostic symptoms being less influenced by context. Laypeople showed a similar curvilinear relation between diagnosticity judgments and context effects. Additionally, clinicians showed a linear relationship between abnormality judgments and context use, with extremely abnormal symptoms being influenced less by context, whereas laypeople showed a curvilinear relationship between symptom abnormality and context use, with extremely abnormal and normal symptoms being influenced the most by context. We discuss implications of these findings for clinical diagnosis.

**Keywords:** clinical decision making, context, conduct disorder, diagnostic judgments, multilevel modeling, curvilinearity

A core component of mental health care involves identifying when a person is experiencing symptoms of a mental health concern (Hunsley & Mash, 2007). The identification of individual mental health symptoms is no easy task given that they do not present in a vacuum; rather, mental health symptoms present in the rich and complicated context of a person's life. For example, finding out that a child has been truant from school can be interpreted differently depending on whether the child is bullied at school, has friends who are also truant, or if the school is in a dangerous neighborhood. Each of these aspects of the surrounding context of a person's life may be used as evidence for or against making a mental disorder diagnosis (e.g., Kirk, Wakefield, Hsieh, & Pottick, 1999; Wakefield & First, 2012). The *Diagnostic and Statistical Manual of Mental Disorders (DSM;* American Psychiatric Association, 2013) explicitly endorses the use of context in thinking about mental disorders, such as in the diagnostic criteria

for conduct disorder that stipulate clinicians should refrain from making such a diagnosis if the patients' symptoms manifest in reaction to life circumstances (e.g., engaging in physical fights for the purpose of self-defense). However, recent work indicates that how clinicians use contextual life factors may depend heavily on the symptom in question (De Los Reyes & Marsh, 2011; Marsh, De Los Reyes, & Wallerstein, 2014). That is, when presented with identical information about the environmental circumstances (i.e., home, school, and peer environment) surrounding a patient, one clinician may use this contextual information to interpret the patient's symptoms as a sign of disorder, whereas another clinician may not. In this paper we explore factors that may predict this flexible use of contextual life information; that is, we investigate the beliefs about disorder symptoms that predict the use of contextual factors in forming impressions of disorder.

## The Influence of Context on Mental Disorder Identification

As stated above, the *DSM* stipulates that the context in which mental health concerns present should be accounted for in the assessment of mental disorder diagnoses (for a review of such provisions, see Wakefield & First, 2012). Specifically, if problematic behaviors or emotions could be described as a natural reaction to the situation in which the person expressing those behaviors is found, then a mental disorder diagnosis is not warranted. Prior work finds that people of various backgrounds (e.g., trained clini-

---

Jessecae K. Marsh and Christopher T. Burke, Department of Psychology, Lehigh University; Andres De Los Reyes, Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland at College Park.

Correspondence concerning this article should be addressed to Jessecae K. Marsh, Department of Psychology, Lehigh University, 17 Memorial Drive East, Bethlehem, PA 18015. E-mail: jessecae.marsh@lehigh.edu

cians and laypeople) can be sensitive to this role of context when making judgments of clinically relevant psychological phenomena (e.g., mental disorders and risk factors for mental health concerns). For instance, when provided with a description of a child expressing antisocial behavior, clinicians (Hsieh & Kirk, 2003; Kirk & Hsieh, 2004; Kirk & Hsieh, 2009; Pottick, Kirk, Hsieh, & Tian, 2007), clinical trainees (Kirk et al., 1999; Wakefield, Pottick, & Kirk, 2002), and people with no training in mental health (Wakefield, Kirk, Pottick, Hsieh, & Tian, 2006) were less likely to judge a mental disorder diagnosis as warranted if an environmental explanation for the symptoms was provided, such as a child's antisocial behavior as a reaction to living in a dangerous neighborhood. Similarly, clinicians rated a diagnosis of major depression as less likely if the symptoms stemmed from an environmental event such as job loss or the death of a loved one than if no such explanation was provided (Kim, Paulus, Nguyen, & Gonzalez, 2012; see also Carta et al., 2013). Further, experimental (De Los Reyes, Ehrlich et al., 2013), field (Hartley, Zakriski, & Wright, 2011), and controlled laboratory observation research (De Los Reyes, Bunnell, & Beidel, 2013; De Los Reyes, Henry, Tolan, & Wakschlag, 2009) indicates that when informants provide reports about the psychosocial functioning of a target participant (e.g., patient), disagreements between reports (e.g., parent reports disruptive behavior in a child patient that the teacher does not) occur, in part, because informants vary in the contexts in which they observe the behaviors being assessed (e.g., parent at home vs. teacher at school; see also De Los Reyes, Thomas, Goodman, & Kundery, 2013; De Los Reyes et al., 2015). Overall, a constellation of behaviors that could warrant a disorder diagnosis can be reinterpreted as problematic or not depending on contextual life factors that could serve as explanatory forces (Wakefield, 2013). Importantly, these contextual factors can influence decisions about treatment (Kim & LoSavio, 2009; Wakefield, Kirk, Pottick, & Hsieh, 1999) and the believed efficacy of treatment (Mulvey & Reppucci, 1988).

This previous research focused on how contextual life information is used after people accumulate enough information to suggest a single diagnosis. Contextual factors can also influence the interpretation of single symptoms of disorder earlier in the information collection process. De Los Reyes and Marsh (2011) provided mental health clinicians with descriptions of hypothetical children that contained one symptom of conduct disorder and three other pieces of information about the child's life. These life factors described characteristics that did not have a formal role in diagnosis, but either were behaviors that seemed likely of a child with conduct disorder, which the authors termed consistent context factors (e.g., his friends' parents tend not to like him, his parents have a hard time finding a babysitter because people refuse to sit him more than once), or inconsistent context factors that were matched in general content but were not associated with a diagnosis of conduct disorder (e.g., his friends' parents tend to like him, his parents have a hard time finding a babysitter because they need sitters at inconvenient times of day). Using this design, the authors tested whether an individual symptom of disorder was seen as more or less indicative of disorder depending on the nature of the surrounding contextual life factors.

De Los Reyes and Marsh (2011) found that clinicians judged the likelihood of conduct disorder to be higher when a disorder symptom was presented with consistent context factors than when

presented alongside inconsistent context factors. However, this effect of contextual factors did not affect symptoms equally. Rather, diagnostic likelihood judgments for some symptoms were stable regardless of the nature of the surrounding contextual information, whereas judgments of other symptoms varied greatly depending on this contextual information. Little agreement was observed among participants about which symptoms were more or less sensitive to contextual factors. Stated another way, no given symptom was agreed upon as being a symptom that signaled disorder the same way regardless of context, nor was any symptom agreed upon as always being interpreted differently in different contexts. This pattern has been replicated in lay participants' judgments of conduct disorder symptoms, as well as symptoms of another behavior disorder (i.e., attention-deficit/hyperactivity disorder; Marsh, De Los Reyes, & Wallerstein, 2014).

The previous work on the interpretation of single symptoms suggests that contextual life factors influence clinicians' and laypeople's interpretations of individual mental health symptoms in ways that seem idiosyncratic to or dependent on the interpreter. That is, the exact same symptom may be a strong or a poor indicator of disorder depending on what context it presents in and the extent to which the evaluating judge infers that the meaning of that symptom depends on contextual factors. Such variation in the evaluation of symptoms across clinicians may help explain why, for instance, field trials testing the reliability of *DSM* mental disorder diagnoses often yield relatively low interrater reliability in applying diagnostic criteria to assessing patients (Freedman et al., 2013). Consequently, a key aim of this study is to examine what beliefs predict the use of contextual information when making clinical judgments about mental disorder symptoms.

### The Influence of Beliefs on Clinical Judgments

People's personal beliefs and theories about mental disorders greatly influence the way they make clinical judgments. Clinicians and laypeople form personalized theories of how the symptoms of mental disorders are related that include highly specific causal or statistical relationships between symptoms (Ahn, Levin, & Marsh, 2005; Flores, Cobos, López, Godoy, & González-Martín, 2014; Kim & Ahn, 2002; Krol, De Bruyn, & Van Den Bercken, 1995; Witteman, Harries, Bekker, & Van Aarle, 2007). Importantly, little agreement exists within clinicians or laypeople in the relationships between symptoms and the roles individual symptoms play, with agreement varying greatly depending on the mental disorder in question (de Kwaadsteniet, Hagmayer, Krol, & Witteman, 2010; Kim & Ahn, 2002). Despite a lack of consensus on the nature of actual theories, people still use the beliefs or theories they have formed in similar ways. For example, the symptoms a judge perceives as the most centrally important to creating other features of a mental disorder also tend to be those that the judge perceives as most central to making a disorder diagnosis (e.g., Kim & Ahn, 2002). This finding generalizes across different types of clinical judgments, in that how a symptom fits into a person's theory of a mental disorder may be more important in deciding clinical judgment than the exact nature of the symptom (Berens, Witteman, & van de Ven, 2011; de Kwaadsteniet et al., 2010; Flores, Cobos, López, & Godoy, 2014; Yopchick & Kim, 2009).

The previous research highlights how idiosyncratic theories of mental health symptoms influence clinical judgment. In this paper

we test the possibility that people's personalized beliefs about individual symptoms of mental disorder predict how those symptoms are influenced by context. Specifically, we propose that the lack of agreement in the extent to which judges use contextual information to form their impressions of individual symptoms (De Los Reyes & Marsh, 2011; Marsh et al., 2014) is influenced by a lack of agreement among participants in their beliefs about those symptoms and their relation to the disorder diagnosis. Furthermore, we predict that symptoms that are perceived to play a similar role in a disorder should be similarly influenced by context.

There are many possible beliefs that could predict the use of contextual information in interpreting individual symptoms of mental disorder. First, research has impressed upon clinicians the importance of making clinical judgments based on statistical measures or probabilities (Dawes, Faust, & Meehl, 1989; Youngstrom, 2013). A clinician, in thinking about how to interpret a single symptom of disorder, may therefore reflect on the symptom's statistical properties within the disorder category. One way a clinician could do this would be to think about how often a given symptom has appeared across all of the cases of a particular disorder she has observed. This estimation of how common a symptom is in a disorder is what we will refer to as an estimation of a symptom's *prevalence*. Thinking about the prevalence of symptoms may help identify a common sign of disorder. Alternatively, clinicians may reflect on all of the cases they have seen with a given symptom and then estimate how many of those cases possessed a certain disorder. This estimation of the likelihood of a disorder given a specific symptom is a measure of the *diagnosticity* of a symptom. Thinking about diagnosticity of a symptom may help identify what behaviors differentiate disordered from normative behavior. The measures of prevalence and diagnosticity map on to the measures of category validity and cue validity, respectively, used in the categorization literature (Ahn et al., 2005; Corter & Gluck, 1992; Sloman, Love, & Ahn, 1998). Our conceptions of prevalence and diagnosticity bear resemblance to concepts in the diagnostic testing literature as well. How common a symptom is in a disorder (i.e., prevalence) can be thought of as relating to the concept of sensitivity, or the proportion of diagnoses that are correctly identified by a test. Likewise, the likelihood of a disorder given a symptom (i.e., diagnosticity) could be likened to the concept of specificity, or the ability of a test to identify people who do not have a diagnosis (see Altman & Bland, 1994). In these comparisons the test can be thought of as the presence of a given symptom. Clinicians' impressions of prevalence and diagnosticity can be predictive of how central a symptom is to a disorder representation (Ahn et al., 2005). As such, it is possible that clinicians' perceptions of these statistical properties will predict symptoms' imperviousness to contextual factors. A key element to note here is that we are specifically interested in clinicians' perceptions of these statistical properties. We do not predict nor is it important to our exploration that clinicians' estimates of these properties are accurate as established through epidemiological studies of mental disorder. Rather, we are interested in whether the beliefs clinicians hold about the prevalence or diagnosticity of symptoms, regardless of their accuracy, are predictive of the use of context.

Symptoms that are seen as extremely prevalent or as extremely diagnostic may be impervious to context. That is, a symptom could be so omnipresent in a disorder or such a clear sign of disordered

behavior, that no matter what context it appears in it is taken as an indicator of the specific disorder. In such a case, we would predict that disorder likelihood ratings would be high for symptoms rated as high in prevalence or diagnosticity in any context. A linear relationship between prevalence or diagnosticity and imperviousness to context may exist, such that as a symptom is seen as less and less prevalent or diagnostic it is likewise seen as more influenced by context. However, symptoms perceived as very low in prevalence or as possessed by a small percentage of people with the disorder in question may be seen as lacking utility when making a diagnosis. In this way, such symptoms would also be interpreted the same regardless of contextual life factors because they never seem indicative of a diagnosis. This suggests that prevalence and diagnosticity could influence judgments in a curvilinear way, with symptoms interpreted as existing in this sweet spot of moderate prevalence or moderate diagnosticity being more influenced by contextual information than those either high or low in prevalence or diagnosticity. Discovering whether linear or curvilinear relationships best describe how beliefs predict the use of context therefore has important implications for understanding which symptoms—those low on our measures of interest and tangentially related to diagnosis or those symptoms falling in the moderate range of our measures—are most likely to be interpreted differently depending on context.

Clinicians have been encouraged to incorporate statistical measures into their assessments because of the natural tendency clinicians have to trust their own intuitions (Davidsen & Fosgerau, 2014; Dawes et al., 1989; for a review, see Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2014). In thinking about how clinicians' personal theories may interact with contextual factors, these more intuitive, qualitative beliefs may predict how a given symptom is interpreted. For example, another way of thinking about how a given symptom differentiates problematic from nonproblematic behavior is to think about how odd it would be for the average person to display that symptom. We refer to this as the general *abnormality* of a given behavior. Previous research has shown that abnormality judgments are tightly linked to clinicians' diagnostic judgments (Kim et al., 2012) but can help clinicians express psychopathology judgments outside the constraints of the *DSM* (Kim, Paulus, Gonzalez, & Khalife, 2012; Kirk & Hsieh, 2009). As such, abnormality may be a strong indicator of whether a behavior should always or never be indicative of a mental disorder. Even more generally, clinicians may be influenced by a more general sense of how *important* they think a symptom is to their process of diagnosis. If a symptom's presence is perceived as critical for diagnosis, the symptom may be perceived as indicative of diagnosis regardless of context. This idea of diagnostic importance captures an intuitive sense clinicians may have of symptoms they preferentially look for in diagnosis. Importantly, the sense of importance of a symptom in a clinician's view of a disorder may be decided by other beliefs held by the clinician (e.g., for a clinician the most important symptoms may be the most abnormal ones or the most diagnostic ones) or may be its own construct of just an intuitive sense of the importance of that symptom. In this sense, the idea of importance here is taken as a general intuition about how critical a symptom is to diagnosis (for a qualitative description of this intuitive judgment see Davidsen & Fosgerau, 2014).

As with the two discussed statistical measures, these more qualitative measures may hold a linear relationship with the use of

context. Symptoms may become less and less influenced by context the more abnormal and important they are perceived to be. Alternatively, symptoms that are seen as very normal or very unimportant to diagnosis may also be impervious to context because they do not increase the likelihood of disorder judgments in any context. In this case, a curvilinear relationship may exist between these measures and the influence of context, with symptoms viewed as moderately abnormal or important showing greater dependence on context than those viewed as either low or high in abnormality or importance.

We have emphasized that professional mental health clinicians' idiosyncratic beliefs about mental disorders may ultimately predict how they make clinical decisions, regardless of the clinical accuracy of those beliefs (e.g., prevalence estimates for a given symptom). Laypeople are likewise influenced by their own idiosyncratic beliefs in assessing symptoms of mental disorder (e.g., Kim & Ahn, 2002; Marsh & Ahn, 2012). However, due to the fact that experience within a domain changes the way that information about categories in that domain is structured (Chi, Feltovich & Glaser, 1981; Murphy & Wright, 1984; Rottman, Gentner, & Goldwater, 2012; Shafto & Coley, 2003; Tanaka & Taylor, 1991; see Marsh & Ahn, 2012, for a discussion of expertise effects in mental health), we would expect that clinicians possess different theories about the roles of the symptoms of those disorders than laypeople do. As such, we would not necessarily expect laypeople to be influenced by the same types of beliefs or in the same ways by the same beliefs as clinicians. Specific to our measures, laypeople may not have experience with a large set of people with a given mental disorder that would help form impressions of the prevalence of symptoms in different disorders. Consequently, they may not consider symptom prevalence in making diagnostic likelihood judgments. Similarly, laypeople do not engage in formal diagnosis, so they may have no basis for determining what is a more or less important symptom for a diagnosis, keeping it from being a meaningful predictor of the influence of contextual factors. However, laypeople are charged with doing a much more informal version of diagnosis such as assessing whether their loved ones or themselves are experiencing distress that warrants seeking the help of a mental health professional (for a review, see De Los Reyes et al., 2015). As such, they may be sensitive to what they believe is abnormal behavior. Likewise, they may be sensitive to thinking about a single odd behavior and what the likelihood is that someone engaging in that behavior is experiencing a mental disorder (i.e., the diagnosticity of a symptom). If these measures are more intuitive to laypeople, then they may be more predictive of the use of context. In this paper we explore this issue by including a sample of laypeople and assessing what beliefs underlie their use of contextual information in making a diagnostic judgment. This is an important issue to address because laypeople comprise key stakeholders in determining whether their own behavior or that of significant others (e.g., children, spouses, or coworkers) warrants mental health care (e.g., they are the point of initial contact between a person needing care and a mental health professional; see Marsh et al., 2014).

Because we are providing a first exploration of what factors may predict the use of contextual life information, we begin with this small set of four factors that represent a variety of different ways of thinking about symptoms. In the following experiment, we tested whether the perceived prevalence, diagnosticity, abnormal-

ity, or importance to diagnosis of mental disorder symptoms predicts the extent to which judges are influenced by contextual life information in making clinical ratings of those symptoms. To do this we adopt the methodology of De Los Reyes and Marsh (2011) and Marsh et al. (2014) and test symptoms within conduct disorder. We predict the influence of context will follow a curvilinear relationship as a function of our measures of interest, with symptoms that are seen as the most and least possessing of a given construct (i.e., prevalence, diagnosticity, abnormality, or importance) lending the same information to diagnosis regardless of context. We additionally tested a group of lay participants to ascertain what beliefs predicted their interpretation of contextual factors. Overall, we provide a first exploration of whether the idiosyncratic beliefs people hold about individual mental health symptoms predict the extent to which they believe the diagnostic value of those symptoms for a given mental disorder is static or contextually dependent and whether the nature of those effects varies across experience groups.

## Method

### Participants

We analyzed data from 89 participants. Forty-five participants were mental health clinicians previously collected and examined in De Los Reyes and Marsh (2011; see paper for demographic and clinical experience of these participants). The remaining 44 participants were undergraduate students at a northeastern private university newly collected for the purposes of this paper and not published elsewhere. An additional undergraduate participated but did not provide a full data set and was dropped from analyses. All undergraduates were recruited out of an introductory psychology course and received partial course credit for completing the experiment. Of the undergraduate participants willing to provide demographic information, the participants primarily identified as female (79%), White (78%), and non-Hispanic (95%). The average age of the undergraduate participants was 18.8 years of age.

### Materials

We used the vignettes constructed and extensively pretested for De Los Reyes and Marsh (2011) as our materials. Each vignette presented four characteristics possessed by a hypothetical youth. The first three characteristics described contextual life factors, with one element each related to the child's family, school interactions, and peer relationships. The life factors had been constructed through several rounds of pretesting in previous work to either represent the type of factors expected in a child with conduct disorder (consistent context) or to represent factors not associated with conduct disorder (inconsistent context). Table 1 gives examples of life factors representing consistent and inconsistent contexts. Importantly, through pretesting of these materials, life factors that could reasonably be perceived as a mental disorder symptom were excluded, and thus final vignette materials included only those life factors that no pretest participants identified as symptoms of any mental health concerns. Further information about the creation of the materials can be found in De Los Reyes and Marsh (2011). After the three life factors, a fourth characteristic was listed that was a symptom of conduct disorder (see Table 2).



Table 1  
*Example Context Descriptions*

Type of factor	Consistent context	Inconsistent context
Family	His father recently lost his job after getting into an argument with his boss.	His father recently lost his job after his company eliminated some employment positions due to budget cuts.
Peer relations	One of his best friends moved back to his school after serving a 1-year sentence in juvenile hall.	One of his best friends moved back to his school after spending a year in the hospital undergoing chemotherapy.
School environment	When he is not studying, he spends time on the computer chatting with older strangers that he has met in person without his parents knowing.	When he is not studying, he spends time on the computer chatting with friends he knows from other schools that his parents have met before.

Note. Taken from Marsh, De Los Reyes, and Wallerstein (2014).

The four characteristics were presented as a bulleted list to highlight to participants that each characteristic was a separate fact about the child. Overall, each vignette presented three facts that either would be expected of a child with conduct disorder or not, and then one symptom of conduct disorder.

### Procedure

The procedure was the same as found in De Los Reyes and Marsh (2011) and Marsh et al. (2014). Participants began the experiment by reading a description of the vignette-rating task they would complete. Participants were told they would read information about hypothetical children attending the same middle school that was collected by brief interviews with a guidance counselor. All of the information was described as currently happening or happening within the last six months. To establish that there was some cause for concern for these children, participants were told all of the described youths were selected for assessment because they were failing at least one class (i.e., a sign of possible impairment stemming from mental health concerns). Participants were told that for each child they read about they would make the following rating: "How likely would a youth with the given life factors be found to have conduct disorder if a full clinical evalu-

ation was given," using a scale of a 0 (*not very likely*) to 100 (*very likely*). Finally, participants were told to give their best estimates without worrying about statistical accuracy; that some, all, or none of the children they read about may meet diagnostic criteria for conduct disorder; and that they should feel free to use the entire 0-to-100 scale. All of this introductory information was the same for clinicians and lay participants. Lay participants in addition read information that familiarized them with the diagnostic criteria for conduct disorder: a description of the type of problems that children with conduct disorder experience, a list of the 15 diagnostic symptoms of conduct disorder seen in Table 2, and that three symptoms were needed to make a diagnosis of conduct disorder.

After the introductory information, participants read and made diagnostic likelihood judgments for 30 individual vignettes. The goal of this rating task was to have participants rate a vignette presenting each of the 15 conduct disorder symptoms once in a consistent framing and once in an inconsistent framing, allowing us to compare ratings for the same symptom across the two contexts. To do this, the presentation of vignettes was split into two blocks. In one block, participants rated eight vignettes presenting a consistent context and seven vignettes presenting an inconsistent context, with each vignette presenting a different conduct disorder symptom from Table 2. The other block again presented each of the 15 conduct disorder symptoms in different vignettes, with seven vignettes presenting a consistent context and eight presenting an inconsistent context. Across the two blocks, each diagnostic symptom of conduct disorder was presented once in a consistent context and once in an inconsistent context. Thus, each participant rated all of the symptoms of conduct disorder either in the consistent or inconsistent framing, and then rated the symptoms again in the context they had not been presented in previously. For each participant, the order of vignettes within a block and the order of blocks were randomized.

After the vignette phase, participants made a series of ratings about the conduct disorder symptoms alone, without any contextual information being present. Participants judged the likelihood a youth displaying each conduct disorder symptom alone would receive a diagnosis of conduct disorder if a full evaluation were given. These are the exact same judgments as for the previous vignettes and thus provide baseline diagnostic likelihood ratings for each symptom outside of the influence of any contextual factors. Participants also made judgments for each symptom of its prevalence (i.e., "what percentage of youths diagnosed with conduct disorder possess that feature"), diagnosticity (i.e., "what percentage of youths who display that feature do you think would

Table 2  
*Fifteen Diagnostic Symptoms of Conduct Disorder*

Diagnostic symptom
Often bullies threatens or intimidates others
Often initiates physical fights
Has used a weapon that can cause serious physical harm to others
Has been physically cruel to people
Has been physically cruel to animals
Has stolen while confronting a victim
Has forced someone into sexual activity
Has deliberately engaged in fire setting with the intention of causing serious damage
Has deliberately destroyed others property (other than fire setting)
Has broken into someone else's house building or car
Often lies to obtain goods or favors or to avoid obligations
Has stolen items of nontrivial value without confronting a victim
Often stays out at night against his parents' wishes and has been doing so since elementary school
Has run away from home overnight multiple times
Is often truant from school and has been doing so since elementary school

Note. Taken from Marsh, De Los Reyes, and Wallerstein (2014).

meet the criteria for having a diagnosis of conduct disorder”), abnormality (i.e., “rate how *abnormal* you believe it is for the average, normal youth to possess each feature”), and importance to diagnosis (i.e., “rate how *important* you believe the feature to be in diagnosing a youth with conduct disorder. By “important” we mean how critical is it for the feature to be present in a youth’s behavior in order for a mental health clinician to feel justified in diagnosing the youth with conduct disorder”). All of these ratings were made on 0-to-100 scales with higher ratings indicating higher likelihood, percentages, abnormality, and importance. The completion order of the five symptom judgments was randomized for each participant.

Finally, participants ended the experiment by providing demographic information (e.g., age, gender, ethnicity/race). Clinicians additionally provided professional information (e.g., clinical orientation and years of clinical experience). All elements of the experiment were self-paced with the experiment designed to be completed in approximately 30 min at a comfortable pace. The experiment was presented through a computerized program built using the Qualtrics Research Suite. Clinicians completed the experiment online on a computer of their own choosing. Lay participants completed the experiment on a computer in the first author’s laboratory.

### Data-Analytic Plan

The primary goal of the present analyses was to identify how the influence of contextual information on diagnosis likelihood depends on judgments of the prevalence, diagnosticity, abnormality, and importance of a given symptom. The outcome variable in the analyses was the difference between participants’ diagnostic likelihood ratings when the symptom is accompanied by consistent contextual information and when the symptom is accompanied by inconsistent contextual information (referred to as the context effect score). This represents the same difference score used as the main variable of interest in De Los Reyes and Marsh (2011) and Marsh et al. (2014).<sup>1</sup> As suggested earlier, the analyses considered both linear and quadratic associations between judgments of the variables of prevalence, diagnosticity, abnormality, and importance.

Because each participant made ratings for 15 symptoms, we used linear multilevel (or mixed) model analyses (Raudenbush & Bryk, 2002) to test these questions. Multilevel models can test the associations between both quantitative and categorical predictors and quantitative outcomes, while accounting for the inherent dependence of the repeated measurements. One way they do so is by allowing for person-level “random effects” or individual differences in the process of interest. This approach can also simultaneously model stimulus-level effects, which are important for obtaining results that are both unbiased and generalizable (Judd, Westfall, & Kenny, 2012).

In the primary analysis, prevalence, diagnosticity, abnormality, and importance were entered as both linear and quadratic predictors of context effect scores.<sup>2</sup> To reduce the bias associated with using difference scores as outcome variables, we included the average of the consistent and inconsistent ratings of the symptom as a covariate (see Seidman & Burke, 2015, for more information on this approach). This average rating and each of the four evaluations were centered at their grand mean before being entered as

predictors. We included random intercepts at both the person and symptom levels; none of the random slopes was significant, so they were not included in the final models. We assumed a compound symmetry residual covariance structure, which assumes equal residual covariances for each pair of symptoms.

Because clinician and lay theories of mental disorders may vary qualitatively as discussed previously, we chose to analyze the two samples separately. Although this approach may overemphasize group differences (e.g., just because an effect is significant for one group and not significant for the other does not imply that the difference between the two groups is significant), we believe it makes the most sense to analyze these samples separately to see what beliefs influence context effects in each population. In the Appendix, we present analyses done on all participants simultaneously that explicitly tested for differences between the two groups.

## Results

### Between-Participant Agreement

We began by examining agreement between participants on the context effect scores and ratings of our belief measures for the different symptoms. We used Kendall’s coefficient of concordance (i.e., Kendall’s *W*) as an indicator of agreement. De Los Reyes and Marsh (2011) had previously reported very low levels of agreement on context effect scores for this sample of clinicians. That is, participants were highly variable in how the presence of consistent versus inconsistent contextual life factors influenced their diagnostic likelihood judgments across symptoms. In these analyses, we expand this finding by presenting agreement analyses for our predictors of interest and for all study variables in the laypeople sample. Table 3 contains these coefficients for the context effect score, baseline diagnostic likelihood (in the absence of contextual information), importance, abnormality, prevalence, and diagnosticity. As seen previously in clinicians, laypeople showed very low agreement on context effect scores across symptoms. The other study variables showed somewhat more concordance across participants, with clinicians generally showing higher agreement than laypeople. However, even the highest of these scores only represents a modest degree of agreement—if interpreted as a correlation, a value of 0.45 would be considered a “medium” effect.

While participants did not agree on how to treat individual symptoms, they may have agreed on how the different variables

<sup>1</sup> There are a number of ways to approach the analysis of data such as these, including a regressed change approach (in which ratings for consistent contexts are regressed on the linear and quadratic versions of prevalence, diagnosticity, abnormality, and importance variable, adjusting for ratings in inconsistent contexts) and a multivariate approach (in which context type is entered as a within-subject predictor and allowed to interact with the linear and quadratic versions of prevalence, diagnosticity, abnormality, and importance variable). We present the difference score version both for consistency with previous research (e.g., De Los Reyes & Marsh, 2011) and because it most directly assesses the relationship of interest—namely, how beliefs about symptoms predict the extent to which the individual takes contextual information into account when evaluating the symptoms.

<sup>2</sup> De Los Reyes and Marsh (2011) did test for simple bivariate correlations between the measured factors and context effect scores. No such correlations were significant. This finding could have been obtained because of the failure of that statistical approach to account for random effects or because of a curvilinear relationship in the data.

Table 3  
Kendall's Coefficient of Concordance for Each of the Study Variables Across Symptoms

	Clinicians	Laypeople
Context score	0.03	0.03
Baseline rating	0.34*	0.21*
Prevalence	0.40*	0.16*
Diagnosticity	0.07*	0.11*
Abnormality	0.45*	0.30*
Importance	0.23*	0.35*

Note. Context score agreement for clinicians was originally calculated and reported in De Los Reyes and Marsh (2011). The value reported here is based on our new analysis of that previous data set but represents the same value as originally reported.

\*  $p < .05$ .

related to each other. To test this, we tested between-participants agreement of the relationships between variables. With 15 ratings per participant, we computed the correlations between each of the study variables for each participant. The distribution of each correlation can be represented using a boxplot, as we have in Figure 1. These boxplots give a sense of the range of correlations seen between each pair of study variables. The boxplots indicate that, while some of the study correlations had a general tendency toward a positive or negative association, others had average correlations close to zero. Symptoms rated high in prevalence

tended to be rated low in abnormality (as well as importance, among laypeople); those high in diagnosticity were generally seen as high in abnormality and importance; and symptoms high in abnormality were also rated high in importance. However, many of these correlations spanned much of the possible range of values (e.g., the range of correlations between importance and diagnosticity was  $-.69$  to  $.94$  for clinicians and  $-.65$  to  $.97$  for laypeople), and none represented an unequivocally positive or negative association. Thus, this analysis reinforces the conclusion of the concordance analysis above that both clinicians and laypeople have idiosyncratic ways of representing these basic features of the symptoms.

### Influence of Context

Moving on to the influence of context, we first tested whether the basic context effect of De Los Reyes and Marsh (2011) and Marsh et al. (2014) replicated in our newly collected laypeople sample. We used a mixed model to analyze these data. Symptoms paired with inconsistent contextual information ( $M = 31.4$ ) were rated as significantly less likely to result in diagnosis than symptoms paired with consistent contextual information ( $M = 59.7$ ),  $t(86) = 11.24$ ,  $p < .001$ , replicating the basic context effect found in the clinician sample as reported in the previous paper. Compared to the baseline ratings ( $M = 55.5$ ), the inconsistent context resulted in significantly lower diagnostic likelihood ratings,  $t(86) = 9.60$ ,  $p < .001$ . As with Marsh and colleagues (2014),

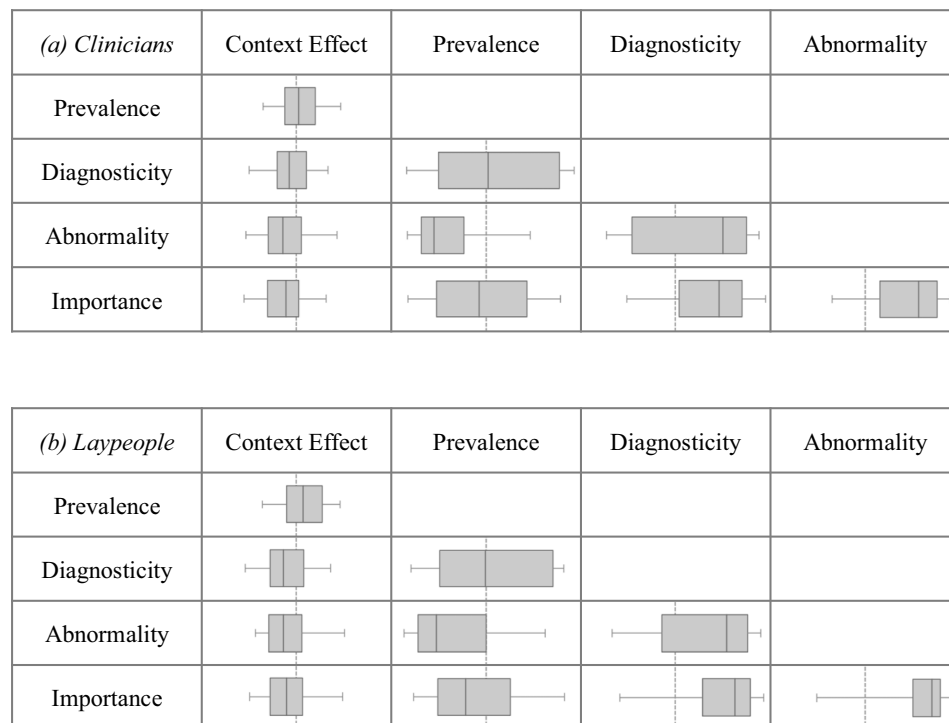


Figure 1. Boxplots of the within-participant correlations between context effect scores (congruent–incongruent diagnostic likelihood ratings) and participants' ratings of symptom prevalence, diagnosticity, abnormality, and importance for (a) clinicians and (b) laypeople. Each cell represents the range  $r = -1$  to  $r = +1$ , with the dashed vertical line representing  $r = 0$ . For clarity of presentation, whiskers extend to the 5th and 95th percentiles of each distribution with outliers not depicted.

symptoms paired with consistent contextual information were rated slightly higher than those without contextual information, however the difference did not reach significance in this sample,  $t(86) = 1.65, p = .103$ .<sup>3</sup> The pattern of results remained when adjusting for ratings of prevalence, diagnosticity, abnormality, and importance. Thus, the data from the current study replicate the basic results of the previous two studies in this line of research.

### Primary Analyses

We now turn to the primary analyses, which examine linear and curvilinear relationships between ratings of symptom prevalence, diagnosticity, abnormality, and importance and the effects of consistent versus inconsistent contextual information on diagnostic likelihood according to the analytic plan outlined above. Table 4 summarizes the linear and quadratic regression estimates for our variables of interest. Starting with the statistical indicators (prevalence and diagnosticity), for symptom prevalence neither the linear nor quadratic terms reached significance when examining clinicians ( $ts < 0.35, ps > .730$ ) – prevalence was unrelated to the magnitude of context effects. Lay participants did show a marginal linear effect of prevalence,  $t(593) = 1.93, p = .054$ , such that higher prevalence ratings were associated with somewhat stronger context effects. The quadratic term remained nonsignificant in the lay sample,  $t(593) = 1.05, p = .232$ . See Figure 2 for a graphical representation of these patterns.

Symptom diagnosticity, in contrast, showed the hypothesized curvilinear relationship with the influence of context (see Figure 2). For the clinicians, the linear component was nonsignificant,  $t(607) = 0.34, p = .739$ , and the quadratic term was marginal,  $t(607) = 1.66, p = .096$ . We reran the analysis removing the nonsignificant quadratic terms for other variables, and when we did so this quadratic term became significant,  $t(603) = 2.10, p = .036$ . For the laypeople, there was a marginal linear component,  $t(593) = 1.79, p = .074$ , and a significant quadratic component,  $t(593) = 2.09, p = .037$ , indicating that symptoms very high or very low in diagnosticity were less susceptible to context effects than those symptoms moderate in diagnosticity.

Turning to the more intuitive indicators (abnormality and importance variables), clinicians showed a significant linear associ-

ation between abnormality and the influence of context,  $t(607) = 3.58, p < .001$ , with no notable curvature,  $t(607) = 0.10, p = .924$ . As seen in Figure 2, for clinicians, the influence of context diminished as symptom abnormality increased. In contrast, laypeople did not show a significant linear association,  $t(593) = 0.19, p = .850$ , but they did exhibit a significant quadratic,  $t(593) = 2.10, p = .036$ . Interestingly, this direction of this quadratic term suggests that, for laypeople, context effects were stronger for symptoms that are very high or very low in abnormality (see Figure 2).

Finally, for the importance variable, clinicians did show significant linear,  $t(607) = 2.75, p = .006$ , and quadratic,  $t(607) = 2.08, p = .038$ , components. As seen in Figure 2, the direction of these effects supports the expected pattern: Symptoms rated as either very high or very low in importance were less susceptible to the presence of consistent or inconsistent contextual information than those symptoms that were moderate in importance. For the laypeople, both terms failed to reach significance ( $ts < .24, ps > .817$ ), indicating that the importance variable did not influence the magnitude of contextual effects.

### General Discussion

The goal of this study was to examine factors that predict how individual symptoms of disorder are reinterpreted given the context they present in. We focused our attention on factors that should be relevant for diagnostic decisions—namely, judges' perceptions of the statistical properties of disorder symptoms (i.e., prevalence and diagnosticity), as well as two more qualitative impressions of disorder symptoms (i.e., importance and abnormality). In clinicians, we found that the statistical measure of diagnosticity and the qualitative measure of importance demonstrated a curvilinear relationship with context effects. In other words, a symptom seen as moderately important to diagnosis or moderately diagnostic of a specific disorder may be seen as a strong sign of disorder in a consistent context but as a weak sign of disorder in an inconsistent context. A symptom that was high or low on these measures of importance or diagnosticity would be seen as an equal sign of disorder regardless of the contextual life factors with which it presented. The other two measures did not show curvilinear relationships, with a linear relationship existing for abnormality ratings and no significant relationship for prevalence ratings.

Lay participants also showed a curvilinear relationship between diagnosticity ratings and context effects, with symptoms rated highest and lowest on diagnosticity being least influenced by context. However, lay participants' results did not resemble clinicians' judgments in other ways. There was no relationship between the importance variable and context for lay participants. The lay sample showed an inverted curvilinear pattern with respect to symptom abnormality (i.e., symptoms of moderate abnormality were less susceptible to context effects than those of low or high abnormality) and a positive linear effect of prevalence, such that more prevalent symptoms were more susceptible to context ef-

Table 4  
*Linear and Quadratic Regression Estimates Predicting Context Effect Scores*

	Clinicians	Laypeople
Intercept	29.58*	28.68*
Average rating	0.33*	0.09
Prevalence (linear)	0.00	0.10
Prevalence (quadratic)	-0.002	-0.002
Diagnosticity (linear)	-0.01	-0.09
Diagnosticity (quadratic)	-0.002 <sup>a</sup>	-0.004*
Abnormality (linear)	-0.16*	0.10
Abnormality (quadratic)	0.000	-0.003*
Importance (linear)	-0.12*	0.01
Importance (quadratic)	-0.003*	0.000

<sup>a</sup> As noted in the text, this effect becomes significant when the nonsignificant quadratic terms of prevalence and abnormality are removed from the analysis.

\*  $p < .05$ .

<sup>3</sup> Given that previous data was not analyzed using multilevel modeling, we reanalyzed the clinician data of De Los Reyes and Marsh (2011) to determine if we replicated the previous results using this statistical technique. All findings replicated as in the original paper with this technique. See De Los Reyes and Marsh for further data presentation. Furthermore, the pattern of results stayed the same when adjusting for prevalence, diagnosticity, abnormality, and importance in the multilevel model.



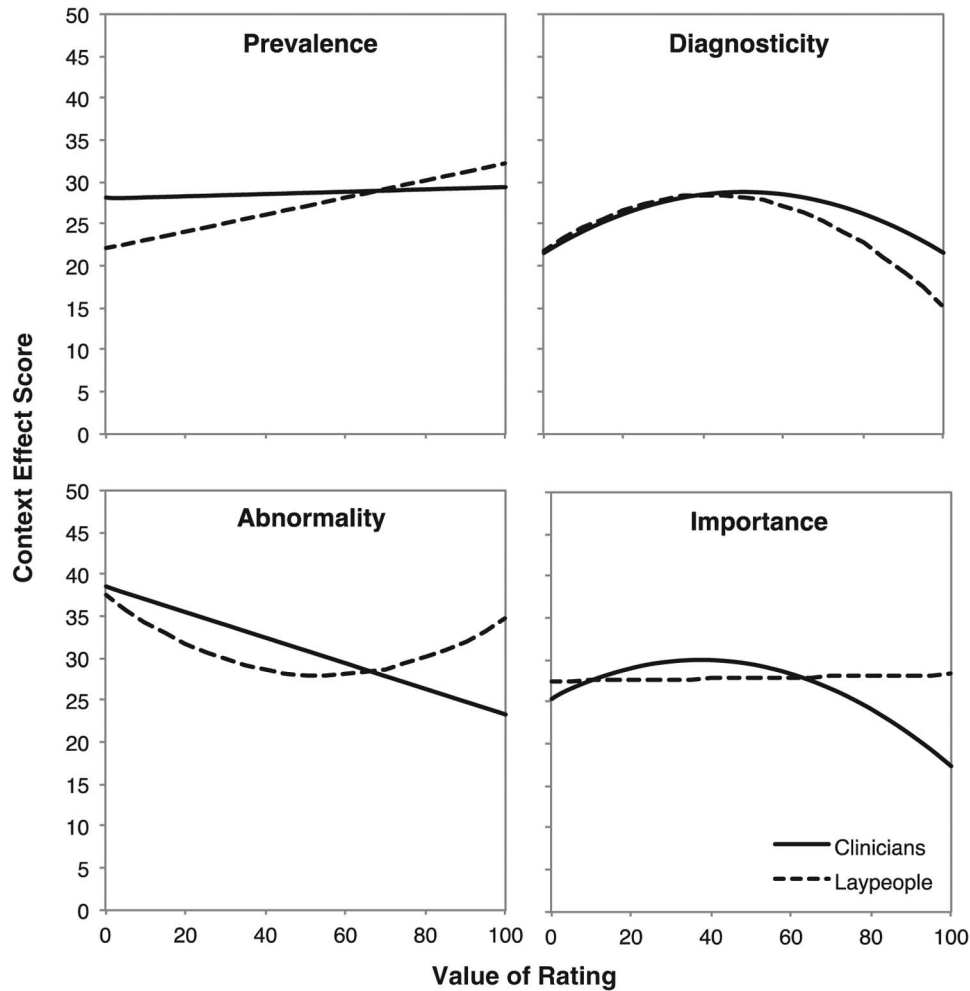


Figure 2. Influence of congruent versus incongruent contextual information on diagnostic likelihood judgments for clinicians and laypeople as a function of rated prevalence, diagnosticity, abnormality, and importance of the symptom for diagnosis.

fects. In short, the data of the lay participants indicate that what predicts laypeople's use of context for a given symptom differs in many respects from predictors of clinicians' use of context.

Our findings suggest that beliefs about disorder symptoms are predictive of how context will be used to interpret those symptoms. However, our agreement analyses suggest there is not consensus on these beliefs among clinicians or laypeople. There was moderate (but not necessarily good) agreement with respect to the variables of symptom importance, abnormality, and prevalence, but low agreement (even among clinicians) on symptom diagnosticity. Going beyond symptom-level agreement, we also found that the correlations among these factors (importance, abnormality, diagnosticity, prevalence, and context effects) varied widely across participants as well (see Figure 1). In general, the correlations spanned most of the possible range across participants suggesting, for instance, that symptoms that are seen as highly important by some participants are also seen as highly diagnostic, whereas for other participants highly important symptoms are seen as low in diagnosticity. The lack of agreement on these predictive variables

highlights how difficult it is to predict a priori how an assessor may see a specific symptom as indicative of disorder in a given context without knowing additional details of the assessor's own beliefs about the disorder symptoms in question.

More broadly, our results suggest a new obstacle to be overcome in improving reliability in clinical decision making. Previous research has documented that clinicians show low agreement in clinical decisions (Garb, 2005; Jensen-Doss, Osterberg, Hickey, & Crossley, 2013; Landis & Koch, 1977), and the level of this agreement depends on such factors as training background (Davidsen & Fosgerau, 2014) and the underlying reason behind the assessment (Large & Nielssen, 2008). We have suggested that this lack of agreement in diagnosis is also fueled by a lack of agreement on how context should be used to evaluate specific symptoms (De Los Reyes & Marsh, 2011; Marsh et al., 2014). The current findings suggest additional factors contributing to variance in clinical decision making. Specifically, lack of agreement in how context should be used to form clinical judgments about mental health symptoms may also stem from a

lack of consensus in other beliefs about the symptoms themselves, such as their diagnosticity, importance to diagnosis, and abnormality. Our work suggests that if we want to understand when a rater will interpret a patient as experiencing a mental disorder, we will need to understand not only the contextual life factors surrounding that patient, but also details of the rater's beliefs about the symptoms of the disorder the patient may display. This adds yet another layer of idiosyncrasy to the identification of mental disorders and understanding why clinicians do not agree on diagnoses.

### Research and Clinical Implications

Our results illuminate the importance of people's beliefs to the clinical decision-making process. An improved understanding of these beliefs and their links to clinical decision making may inform methods for training mental health professionals and thus improving reliability in clinical decision making. Namely, if professionals could be trained to more accurately assess disorder symptoms' diagnosticity, importance to diagnosis, and abnormality, their beliefs on these variables could become more standardized. For example, more explicit training in graduate programs on how often a symptomatic behavior is seen in nondiagnosed populations may help standardize beliefs about the diagnosticity of disorder symptoms. Such training may improve clinicians' abilities to incorporate contextual life information in similar ways to each other. With such factors being used in more similar ways, it may be easier to then implement training programs to better advise trainees how life factors should actually influence the evaluation of individual disorder symptoms. However, given the current controversy in how context is treated in the *DSM* (e.g., Wakefield & First, 2012), it is not clear what form these training regimens would take.

Given that we were providing a first exploration of what beliefs predicted the use of context, we restricted our exploration to four possible factors to provide a tractable first pass on this question. However, there may be a myriad of other factors that influence the evaluation of diagnostic symptoms and predict the influence of context outside of our measured variables. For example, other more intuitive judgments, such as how causally central a symptom seems to be (e.g., Kim & Ahn, 2002) or how hard it is to imagine removing a given symptom and a person still having the disorder (e.g., Sloman, Love, & Ahn, 1998) may predict the extent to which context influences interpretation of that symptom. We task future research to explore what other beliefs reliably influence the interpretation of contextual information in clinician and lay populations.

Finally, the differences between what factors were predictive of the use of contextual information in clinicians and laypeople illustrate the role clinical experience has in shaping theories about mental illness. Importance to diagnosis was a significant predictor for clinicians and not laypeople. The lack of a relationship is not surprising for laypeople because they do not engage in formal diagnostic practice that could have encouraged weighting some symptoms as more important than others. That clinicians do show a relationship between importance ratings and context effect scores suggests that the idea of tagging select symptoms as key identifiers of a given disorder is something that develops with clinical experience. Likewise, the different patterns of relationship with context for prevalence and abnormality ratings likewise suggest that clin-

ical experience shapes how people think about what is abnormal and what is prevalent in a disorder category. Future research should explore how these idiosyncratic beliefs about the nature of prevalence, abnormality, and importance develop as experience and training is gained in formal clinical practice.

### Limitations

The limitations of our study reveal directions for future research. First, our sample of clinicians was predominantly White (as reported in De Los Reyes & Marsh, 2011) and future research is needed to see if our effects generalize to clinicians from multiple ethnic and racial backgrounds. Second, we examined how contextual information affects clinicians' impressions using diagnostic likelihood ratings. Yet, it remains unclear how these impressions of diagnostic likelihood impact actual clinical decisions such as referring a child for further evaluation or forming a treatment plan. Third, the laypeople we sampled in this study were a group of undergraduates who were not actively engaged in making diagnostic decisions for others, such as might be case with parents or teachers, meaning they may incorporate contextual information less than laypeople charged with helping children gain access to care. Given that both clinicians and laypeople evidenced the same type of context effects, there is reason to believe parents and teachers may show similar trends as those of undergraduates in the present study (see Marsh et al., 2014 for further discussion). As such, we recommend future research examining how much more laypeople charged with making decisions for someone with a mental health concern incorporate and are influenced by contextual information.

### Conclusion

Ultimately, our results suggest that the use of contextual information when interpreting mental health symptoms is both prevalent and complicated. Depending on the personal beliefs and theories of a given professional or layperson about disorder symptoms, different conclusions may be drawn about whether the exact same symptom is indicative of a given mental disorder. In essence, for a given evaluator, there appears to be an idiosyncratic set of symptoms that will be most susceptible to the disorder-congruence or -incongruence of contextual details. Our findings suggest that this set will include for clinicians those symptoms that are viewed as moderate in importance or diagnosticity or low in abnormality by that particular evaluator. Research that furthers investigates what beliefs underlie the use of context is necessary to help us better understand how people identify and recognize symptoms of mental illness when they first present.

### References

- Ahn, W., Levin, S., & Marsh, J. K. (2005). Determinants of feature centrality in clinicians' concepts of mental disorders. In B. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 67-72). Mahwah, NJ: Erlbaum, Inc.
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 1: Sensitivity and specificity. *British Medical Journal*, *308*, 1552. <http://dx.doi.org/10.1136/bmj.308.6943.1552>

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Berens, C., Witteman, C. L. M., & van de Ven, M. O. M. (2011). Is understanding why necessary for treatment choices? *European Journal of Psychological Assessment, 27*, 81–87. <http://dx.doi.org/10.1027/1015-5759/a000047>
- Carta, M. G., Angermeyer, M. C., Matschinger, H., Holzinger, A., Floris, F., & Moro, M. F. (2013). Perception of depressive symptoms by the Sardinian public: Results of a population study. *BMC Psychiatry, 13*, 57. <http://dx.doi.org/10.1186/1471-244X-13-57>
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152. [http://dx.doi.org/10.1207/s15516709cog0502\\_2](http://dx.doi.org/10.1207/s15516709cog0502_2)
- Cortner, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin, 111*, 291–303. <http://dx.doi.org/10.1037/0033-2909.111.2.291>
- Davidson, A. S., & Fosgerau, C. F. (2014). What is depression? Psychiatrists' and GPs' experiences of diagnosis and the diagnostic process. *International Journal of Qualitative Studies on Health and Well-Being, 9*, 24866. <http://dx.doi.org/10.3402/qhw.v9.24866>
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1674. <http://dx.doi.org/10.1126/science.2648573>
- de Kwaadsteniet, L., Hagmayer, Y., Krol, N. P., & Witteman, C. L. (2010). Causal client models in selecting effective interventions: A cognitive mapping study. *Psychological Assessment, 22*, 581–592. <http://dx.doi.org/10.1037/a0019696>
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*. Advance online publication. <http://dx.doi.org/10.1037/a0038498>
- De Los Reyes, A., Bunnell, B. E., & Beidel, D. C. (2013). Informant discrepancies in adult social anxiety disorder assessments: Links with contextual variations in observed behavior. *Journal of Abnormal Psychology, 122*, 376–386. <http://dx.doi.org/10.1037/a0031150>
- De Los Reyes, A., Ehrlich, K. B., Swan, A. J., Luo, T., Van Wie, M., & Pabón, S. C. (2013). An experimental test of whether informants can report about child and family behavior based on settings of behavioral expression. *Journal of Child and Family Studies, 22*, 177–191. <http://dx.doi.org/10.1007/s10826-012-9567-3>
- De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology, 37*, 637–652. <http://dx.doi.org/10.1007/s10802-009-9307-3>
- De Los Reyes, A., & Marsh, J. K. (2011). Patients' contexts and their effects on clinicians' impressions of conduct disorder symptoms. *Journal of Clinical Child and Adolescent Psychology, 40*, 479–485. <http://dx.doi.org/10.1080/15374416.2011.563471>
- De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9*, 123–149. <http://dx.doi.org/10.1146/annurev-clinpsy-050212-185617>
- Flores, A., Cobos, P. L., López, F. J., & Godoy, A. (2014). Detecting fast, online reasoning processes in clinical decision making. *Psychological Assessment, 26*, 660–665. <http://dx.doi.org/10.1037/a0035151>
- Flores, A., Cobos, P. L., López, F. J., Godoy, A., & González-Martín, E. (2014). The influence of causal connections between symptoms on the diagnosis of mental disorders: Evidence from online and offline measures. *Journal of Experimental Psychology: Applied, 20*, 175–190. <http://dx.doi.org/10.1037/xap0000025>
- Freedman, R., Lewis, D. A., Michels, R., Pine, D. S., Schultz, S. K., Tamminga, C. A., . . . Yager, J. (2013). The initial field trials of *DSM-5*: New blooms and old thorns. *The American Journal of Psychiatry, 170*, 1–5. <http://dx.doi.org/10.1176/appi.ajp.2012.12091189>
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89. <http://dx.doi.org/10.1146/annurev.clinpsy.1.102803.143810>
- Hartley, A. G., Zakriski, A. L., & Wright, J. C. (2011). Probing the depths of informant discrepancies: Contextual influences on divergence and convergence. *Journal of Clinical Child and Adolescent Psychology, 40*, 54–66. <http://dx.doi.org/10.1080/15374416.2011.533404>
- Hsieh, D. K., & Kirk, S. A. (2003). The effect of social context on psychiatrists' judgments of adolescent antisocial behavior. *Journal of Child Psychology and Psychiatry, 44*, 877–887. <http://dx.doi.org/10.1111/1469-7610.00172>
- Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3*, 29–51. <http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091419>
- Jensen-Doss, A., Osterberg, L. D., Hickey, J. S., & Crossley, T. (2013). Agreement between chart diagnoses and standardized instrument ratings of youth psychopathology. *Administration and Policy in Mental Health and Mental Health Services Research, 40*, 428–437. <http://dx.doi.org/10.1007/s10488-012-0436-6>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*, 54–69. <http://dx.doi.org/10.1037/a0028347>
- Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131*, 451–476. <http://dx.doi.org/10.1037/0096-3445.131.4.451>
- Kim, N. S., & LoSavio, S. T. (2009). Causal explanations affect judgments of the need for psychological treatment. *Judgment and Decision Making, 4*, 82–91. <http://journal.sjdm.org/81028/jdm81028.pdf>
- Kim, N. S., Paulus, D. J., Gonzalez, J. S., & Khalife, D. (2012). Proportionate responses to life events influence clinicians' judgments of psychological abnormality. *Psychological Assessment, 24*, 581–591. <http://dx.doi.org/10.1037/a0026416>
- Kim, N. S., Paulus, D. J., Nguyen, T. P., & Gonzalez, J. S. (2012). Do clinical psychologists extend the bereavement exclusion for major depression to other stressful life events? *Medical Decision Making, 32*, 820–830. <http://dx.doi.org/10.1177/0272989X12443417>
- Kirk, S. A., & Hsieh, D. K. (2004). Diagnostic consistency in assessing conduct disorder: An experiment on the effect of social context. *American Journal of Orthopsychiatry, 74*, 43–55. <http://dx.doi.org/10.1037/0002-9432.74.1.43>
- Kirk, S. A., & Hsieh, D. K. (2009). Do perceptions of dysfunction and normality mediate clinicians' judgments of adolescent antisocial behavior? *Social Service Review, 83*, 245–266. <http://dx.doi.org/10.1086/599981>
- Kirk, S. A., Wakefield, J. C., Hsieh, D. K., & Pottick, K. J. (1999). Social context and social workers' judgment of mental disorder. *Social Service Review, 73*, 82–104. <http://dx.doi.org/10.1086/515798>
- Krol, N. P. C. M., De Bruyn, E. E. J., & Van Den Bercken, J. H. L. (1995). Intuitive and empirical prototypes in childhood psychopathology. *Psychological Assessment, 7*, 533–537. <http://dx.doi.org/10.1037/1040-3590.7.4.533>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. <http://dx.doi.org/10.2307/2529310>
- Large, M. M., & Nielssen, O. (2008). Factors associated with agreement between experts in evidence about psychiatric injury. *Journal of the American Academy of Psychiatry and the Law, 36*, 515–521.
- Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives*

- on *Psychological Science*, 9, 355–387. <http://dx.doi.org/10.1177/1745691614535216>
- Marsh, J. K., & Ahn, W. (2012). Memory for patient information as a function of experience in mental health. *Applied Cognitive Psychology*, 26, 462–474. <http://dx.doi.org/10.1002/acp.2832>
- Marsh, J. K., De Los Reyes, A., & Wallerstein, A. (2014). The influence of contextual information on lay judgments of childhood mental health concerns. *Psychological Assessment*, 26, 1268–1280. <http://dx.doi.org/10.1037/pas0000012>
- Mulvey, E. P., & Reppucci, N. D. (1988). The context of clinical judgment: The effect of resource availability on judgments of amenability to treatment in juvenile offenders. *American Journal of Community Psychology*, 16, 525–545. <http://dx.doi.org/10.1007/BF00922769>
- Murphy, G. L., & Wright, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 144–155. <http://dx.doi.org/10.1037/0278-7393.10.1.144>
- Pottick, K. J., Kirk, S. A., Hsieh, D. K., & Tian, X. (2007). Judging mental disorder in youths: Effects of client, clinician, and contextual differences. *Journal of Consulting and Clinical Psychology*, 75, 1–8. <http://dx.doi.org/10.1037/0022-006X.75.1.1>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive Science*, 36, 919–932. <http://dx.doi.org/10.1111/j.1551-6709.2012.01253.x>
- Seidman, G., & Burke, C. T. (2015). Partner enhancement, verification, and derogation and emotional responses to daily conflict. *Journal of Social and Personal Relationships*, 32, 304–329. <http://dx.doi.org/10.1177/0265407514533227>
- Shafto, P., & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 641–649. <http://dx.doi.org/10.1037/0278-7393.29.4.641>
- Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, 22, 189–228. [http://dx.doi.org/10.1207/s15516709cog2202\\_2](http://dx.doi.org/10.1207/s15516709cog2202_2)
- Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23, 457–482. [http://dx.doi.org/10.1016/0010-0285\(91\)90016-H](http://dx.doi.org/10.1016/0010-0285(91)90016-H)
- Wakefield, J. C. (2013). The *DSM-5* debate over the bereavement exclusion: Psychiatric diagnosis and the future of empirically supported treatment. *Clinical Psychology Review*, 33, 825–845. <http://dx.doi.org/10.1016/j.cpr.2013.03.007>
- Wakefield, J. C., & First, M. B. (2012). Placing symptoms in context: The role of contextual criteria in reducing false positives in *Diagnostic and Statistical Manual of Mental Disorders* diagnoses. *Comprehensive Psychiatry*, 53, 130–139. <http://dx.doi.org/10.1016/j.comppsy.2011.03.001>
- Wakefield, J. C., Kirk, S. A., Pottick, K. J., & Hsieh, D. (1999). Disorder attribution and clinical judgment in the assessment of adolescent antisocial behavior. *Social Work Research*, 23, 227–238. <http://dx.doi.org/10.1093/swr/23.4.227>
- Wakefield, J. C., Kirk, S. A., Pottick, K. J., Hsieh, D. K., & Tian, X. (2006). The lay concept of conduct disorder: Do nonprofessionals use syndromal symptoms or internal dysfunction to distinguish disorder from delinquency? *The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie*, 51, 210–217.
- Wakefield, J. C., Pottick, K. J., & Kirk, S. A. (2002). Should the *DSM-IV* diagnostic criteria for conduct disorder consider social context? *The American Journal of Psychiatry*, 159, 380–386. <http://dx.doi.org/10.1176/appi.ajp.159.3.380>
- Wittman, C. L. M., Harries, C., Bekker, H. L., & Van Aarle, E. J. M. (2007). Evaluating psychodiagnostic decisions. *Journal of Evaluation in Clinical Practice*, 13, 10–15. <http://dx.doi.org/10.1111/j.1365-2753.2006.00689.x>
- Yopchick, J. E., & Kim, N. S. (2009). The influence of causal information on judgments of treatment efficacy. *Memory & Cognition*, 37, 29–41. <http://dx.doi.org/10.3758/MC.37.1.29>
- Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child and Adolescent Psychology*, 42, 139–159. <http://dx.doi.org/10.1080/15374416.2012.736358>

(Appendix follows)



## Appendix

### Additional Analyses Comparing Clinicians and Laypeople

These analyses assess the overall pattern of results (i.e., averaging across the two samples) for the primary analyses and explicitly test the differences between clinicians and lay participants in each parameter. We created an effect-coded variable to indicate group membership (coded  $-0.5$  for laypeople and  $0.5$  for clinicians) and interacted it with each of the terms in the model. The terms that do not include this variable indicate the average pattern (ignoring group membership), and the terms that do include this variable test for differences between clinicians and laypeople. For symptom prevalence, after accounting for the other symptom ratings, neither the linear nor quadratic terms reached significance in the combined analysis,  $ts < 1.46$ ,  $ps > .148$ . Likewise, neither of these effects differed significantly between the two samples,  $ts < 1.41$ ,  $ps > .161$ . For symptom diagnosticity, the linear term was nonsignificant in the overall analysis,  $t(1,215) = 1.56$ ,  $p = .119$ , and the quadratic term was significant,  $t(1,215) = 2.55$ ,  $p = .011$ . The analysis suggested a marginal difference between clinicians and laypeople in the magnitude of the linear component,  $t(1,215) = 1.67$ ,  $p = .094$ , but no differences between the samples in the quadratic component,  $t(1,215) = 0.57$ ,  $p = .569$ . Looking next at symptom abnormality, in the overall analysis both the linear,  $t(1,215) = 1.83$ ,  $p = .067$ , and quadratic,  $t(1,215) = 1.74$ ,

$p = .082$ , terms were marginal after adjusting for the other ratings. Because the quadratic was only marginal, we reran the analysis without the quadratic term, and there was a significant linear relationship,  $t(1,217) = 2.83$ ,  $p = .005$ , between abnormality and the influence of context. It was negative in sign, suggesting a decrease in the influence of context as abnormality increased. In addition, there was a marginal difference between clinicians and laypeople in the magnitude of the linear term,  $t(1,215) = 1.93$ ,  $p = .054$ , but not in the quadratic term,  $t(1,215) = 1.39$ ,  $p = .166$ . Finally, for symptom importance, averaging across clinicians and laypeople, neither the linear,  $t(1,215) = 1.91$ ,  $p = .168$ , nor the quadratic,  $t(1,215) = 0.85$ ,  $p = .358$ , components reached significance, suggesting that the magnitude of the observed context effects did not vary as a function of perceived symptom importance, adjusting for the other symptom evaluations. There were no significant differences between the two groups in these effects ( $ts < 2.20$ ,  $ps > .138$ ).

Received November 26, 2014

Revision received April 23, 2015

Accepted April 27, 2015 ■