Patients' Contexts and Their Effects on Clinicians' Impressions of Conduct Disorder Symptoms

Andres De Los Reyes[a]; Jessecae K. Marsh[b]

[a] Department of Psychology, University of Maryland at College Park, [b] Department of Psychology, Texas Tech University,

PLEASE SCROLL DOWN FOR ARTICLE

Patients' Contexts and Their Effects on Clinicians' Impressions of Conduct Disorder Symptoms

Andres De Los Reyes

Department of Psychology, University of Maryland at College Park

Jessecae K. Marsh

Department of Psychology, Texas Tech University

The purpose of this study was to examine whether contextual information about patients' clinical presentations affected clinicians' judgments of conduct disorder symptoms. Forty-five clinicians read vignettes describing hypothetical patients who displayed one conduct disorder symptom alongside information about the patients' home, school, and peer contexts. Clinicians judged the likelihood of patients meeting conduct disorder criteria. Contextual information highly affected judgments and these effects varied across the 15 conduct disorder symptoms. It is important to note that clinical judgments were not in agreement on the symptoms affected by context.

Diagnostic classifications systems, such as the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev. [*DSM–IV–TR*]), provide for every diagnosis a list of symptoms to guide clinicians through patient evaluations (American Psychiatric Association [APA], 2000). However, when evaluating a patient, information collected during the evaluation varies in its indication of the presence of a symptom. That is, information collected during an assessment of a patient may (a) directly indicate the presence of a diagnosis (e.g., a diagnostic symptom), (b) indirectly suggest the presence of the diagnosis without satisfactorily indicating the presence of a symptom (e.g., risk factor identified in research), or (c) neither directly indicate nor indirectly suggest the presence of the diagnosis or its symptoms. In fact, many patient characteristics, even those identified by empirical research as risk factors for the particular

diagnosis being assessed, often cannot be used as definitive evidence of the presence of a diagnosis in any one patient being assessed (see Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001). This study examines whether *contextual information* about patients' presentations affects clinicians' symptom interpretations.

These issues have particular relevance to the assessment and diagnosis of conduct disorder. Conduct disorder is characterized by children and adolescents (hereafter referred to collectively as "children") displaying aggressive behavior toward others and a general disregard of rules and social norms (APA, 2000). It is important to note that a variety of contextual factors have been identified as posing risk for the condition's development, including dysfunctional parenting, parental psychopathology, rejection by others (e.g., peers, authority figures), and associations with deviant peers (Burke, Loeber, & Birmaher, 2002), with interventions for the condition targeting these factors (Weisz, Hawley, & Jensen Doss, 2004). However, the *DSM* recommends to clinicians to rule out a diagnosis if the symptoms can be *explained* by context (APA, 2000). Indeed, clinicians use context when making vignette-based judgments in which diagnostic information (i.e., symptoms to warrant a diagnosis of conduct disorder) is presented alongside either a contextual or dispositional explanation for

patients' symptoms (e.g., Wakefield, Pottick, & Kirk, 2002). Yet, how do clinicians judge *each* symptom when also provided with contextual information?

An examination of *DSM* guidelines suggests that clinicians may encounter difficulties in consistently incorporating contextual information when making judgments about individual symptoms. For instance, the *DSM* does not provide guidance to clinicians as to which contextual factors are important to consider when making diagnostic decisions. In addition, conduct disorder symptoms vary greatly in their underlying character (e.g., *stays out at night*; *forced someone into sexual activity*), suggesting the possibility that context may not affect clinicians' impressions of conduct disorder symptoms equally. Consistent with this notion, work in cognitive psychology finds that clinicians vary greatly in how they perceive and reason about clinical information and hence make clinical decisions (e.g., Kim & Ahn, 2002). Thus, context effects on symptom judgments may vary by symptom and clinician.

The purpose of this study was to investigate (a) how contextual factors that do not directly indicate the presence of symptoms affect judgments of diagnostic symptoms overall and (b) whether these contextual factors would influence clinicians' judgments of all disorder symptoms equally. To meet these objectives, we recruited mental health clinician participants to read vignettes describing factors in a child patient's life that were either consistent with intuitive depictions of a child with conduct disorder (e.g., "One of his friends recently started attending another school because he was expelled from his previous school.") or inconsistent with intuitive depictions of a child with conduct disorder (e.g., "His friends' parents tend to like him."). We selected factors for use in the "consistent condition" vignettes based on prior empirical work on key contextual factors that pose risk for the development of conduct disorder (Burke et al., 2002). After reading a vignette, clinicians provided a judgment rating of the likelihood that the child described in the vignette would meet criteria for a conduct disorder diagnosis if a full clinical evaluation of the child were administered.

Overall, we hypothesized that participants' symptom judgments would be affected by context, with symptoms paired with features that are consistent with a diagnosis of conduct disorder resulting in higher likelihood judgments relative to symptoms paired with inconsistent features. However, we further predicted that the extent to which clinical impressions would be influenced by context would vary across diagnostic symptoms. As such, a second study goal was to examine if context influenced clinical impressions of symptoms differentially and, if so, whether clinical impressions were in agreement as to the affected symptoms. It is possible that impressions of certain

"important" symptoms would be reliably unaffected by context. However, previous cognitive research suggests that clinicians' own idiosyncratic theories influence their decision-making (Kim & Ahn, 2002). Therefore, clinicians' impressions might not be in agreement as to the symptoms most affected by context.

METHODS

Participants

We recruited 46 professional mental health clinicians specializing in the treatment of children. Forty-five participants provided complete data and thus were included in all analyses (33 women, age range = 29–61 years, 40 Caucasians and 1 Asian). Of these 45 participants, 40 participants reported certifications or licensures in psychology, 4 in counseling, and 1 in social work (range of certification year = 1976–2010). Thirty-four participants reported their highest degree as Ph.D., 6 as Psy.D., 3 as M.S., 1 as M.S.W., and 1 as M.Ed.[1]

Recruitment and compensation. We contacted clinicians by posting advertisements through mailing lists of professional organizations that cater to clinicians specializing in the care of children. These advertisements included a link that participants could use to access the online experiment. We used the Qualtrics Research Suite software to construct and administer test stimuli (Qualtrics Labs, Inc., 2005). Clinicians completed the experiment online on a computer of their choice and at their own pace. We designed the experiment to be completed in 30 min. As compensation, participants were either entered into a drawing for a $50 gift certificate with a $5 donation in each participant's name made to a child's welfare charity (N = 22) or paid a flat amount of $20 (N = 23).[2]

Main Experimental Procedure

Overview. All procedures were approved by the Internal Review Boards of the universities with which the first author and second author are affiliated. Our main experiment tested whether contextual information affected clinicians' symptom interpretations. Specifically, using a within-subjects design participants made

---

[1] In a few instances, participants did not report demographic data. Specifically, across the 45 participants, 1 participant did not report gender, 6 their age, 4 their race identification, and 1 the year of license certification.

[2] We initially used a lottery format to recruit participants. Because of low recruitment rates, we switched to using the flat rate payment. Findings do not differ if we examine data provided by the lottery participants only or from all of the participants.

TABLE 1
Average Context Effect Scores Across the 15 Diagnostic Symptoms

Diagnostic Symptom	M	SD	Minimum	Maximum
Run away from home overnight	28.31	25.6	−30	88
Bullies others	27.04	24.2	−13	80
Destroyed property without fire	27.04	24.6	−20	80
Cruel to animals	26.24	23.4	−5	80
Stays out at night	25.56	28.2	−30	80
Truant from school	25.40	23.7	−10	85
Broken into a house	25.13	25.1	−20	88
Used a weapon for harm	24.51	24.8	−10	75
Stolen without confronting a person	24.33	26.3	−50	83
Initiates fights	20.67	25.7	−20	85
Stolen while confronting a person	20.44	22.0	−20	80
Lies to avoid and obtain	18.96	21.0	−35	75
Set fires	18.80	21.0	−20	70
Cruel to people	18.76	20.1	−30	70
Forced sexual activity	17.20	21.1	−15	75

the likelihood judgments described previously twice for each of the 15 conduct disorder symptoms seen in Table 1 (i.e., 30 vignettes total). That is, we presented each symptom to participants once alongside three patient characteristics that were not surprising to be seen in a child with conduct disorder ("Consistent

TABLE 2
Example Sets of Consistent and Inconsistent Context Versions of a Vignette

Consistent Context	Inconsistent Context
• His parents have had a difficult time finding him a regular babysitter because people often refuse to babysit him more than once.	• His parents have had a difficult time finding him a regular babysitter because his parents need a sitter at inconvenient times during the day.
• His friends' parents tend not to like him.	• His friends' parents tend to like him.
• He doesn't like some of his classmates because they wouldn't let him cheat off their tests.	• He doesn't like some of his classmates because they try to cheat off his tests.
• He deliberately set a fire with the intention of causing serious damage.	• He often lies to obtain things he wants and to avoid obligations.

Note: Underlining denotes the stem that is identical between the two versions. The underlining was not used in the actual experiment. Each row represents a different type of characteristic that was presented to the participants. These characteristics described the following areas of the patient's life and were presented in each vignette in the order depicted: family relations, peer relations, school environment, and finally a diagnostic symptom of conduct disorder. Reading down either column provides an example of a single vignette that would have been provided to participants on a single screen of the experiment.

Context" [CC] condition). We also presented the same symptom alongside features surprising to be seen in a child with conduct disorder ("Inconsistent Context" [IC] condition). Vignette presentation was blocked such that all of the symptoms were rated once (half in the CC condition and half in the IC condition) before being rated again in the other context. Table 2 shows examples of CC and IC vignettes.

Online study procedure. After completing an online consent form, participants began the experiment by reading a set of instructions that orientated them to the upcoming vignette-rating task of the experiment. Participants were told that they would be asked to make ratings about children who were showing different behaviors. The children were described as attending the same middle school and had been randomly selected for assessment from a pool of students who were failing at least one class. The participants were told that guidance counselors had completed brief initial interviews and the participants were going to read the initial information the counselors had collected in separate vignettes. Furthermore, the participants were told that everything they were going to read was either currently true of the child or occurred in the last 6 months.[3]

Participants were instructed that they would be asked to make a likelihood rating for each vignette worded as, "How likely would a youth with the given life factors be found to have Conduct Disorder if a full clinical evaluation was given," using a scale of a 0 (*not very likely*) to 100 (*very likely*). Participants were told not to worry about being statistically accurate but rather to provide ratings that best matched their clinical impressions of the children. Further, participants were told that some, all, or none of the children they were to read about may meet diagnostic criteria for conduct disorder and were encouraged to use the entire 0-to-100 scale. Participants rated 30 vignettes (15 CC vignettes and 15 IC vignettes). We blocked vignette order so that participants rated a separate vignette depicting each of the 15 diagnostic symptoms once (half in the CC condition and half in the IC condition) before rating the symptom again in the other context. The order of vignettes was randomized within the two blocks, and the presentation order of the two blocks was randomized across participants.

After rating the vignettes, participants completed another round of ratings using the same scale where they

---

[3]This introductory passage highlighted that the children about whom participants read evidenced impairment. This fact is important to establish so that the children described in the vignettes could be *potentially* evidencing the type of functional impairments that need to be present to warrant a diagnosis of conduct disorder (see APA, 2000).

rated each diagnostic symptom with no contextual information provided (i.e., likelihood that a youth with Symptom X would receive a diagnosis of conduct disorder if a full evaluation was given). These No Context (NC) ratings were completed after the vignettes so as to not overemphasize the contextual manipulation of the experiment.[4] The participants ended the experiment by providing the demographic and professional information described previously.

Materials.   In our main experiment we presented participants with vignettes describing hypothetical children that each included one symptom of conduct disorder alongside three characteristics of the child's life. Based on prior work of robust risk factors of conduct disorder (see Burke et al., 2002), these three characteristics were composed of one feature about the child's family life, one feature about the child's peer relations, and one feature describing the child's school environment. We constructed each characteristic so that it either would (CC) or would not (IC) represent one of three different empirically based conduct disorder risk factors: (a) dysfunctional parenting or parental psychopathology, (b) rejection by others (peers, adult authority figures), and (c) associations with deviant peers. The contextual characteristics either depicted an IC or a CC context. Table 2 shows an example of contextual features used in the IC and CC conditions. As can be seen in this table, we created an IC and a CC version of each characteristic in such a way that conveyed the same basic structure (e.g., nearly equal word count and grammatical structure). This ensured that we did not convey extra information in one condition versus another.

To develop materials for the experiment, we conducted two different pretests. First, we created the individual characteristic features that would be presented in the two conditions. In total, we pretested 163 matched pairs, or 326 individual contextual factor items. We tested these 163 matched pairs on 29 graduate students receiving training in mental health research and practice who had yet to begin clinical internship training outside their degree-granting program, using recruitment procedures similar to those used for the main experiment (e.g., online postings to professional organizations, e-mails sent directly to faculty and graduate students at mental health programs). Participants made judgments on these pairs of stimuli using a 0-to-100 scale similar to that

used in our main experiment. In addition, participants were asked to judge (a) if the given characteristic was actually an instantiation of a conduct disorder symptom and (b) if the given characteristic was a symptom of another mental disorder by listing the name of the corresponding disorder. We entered each participant into a drawing to win a $30 online gift certificate.

From this pretest, we excluded any characteristics that at least one participant identified as a conduct disorder symptom or another mental disorder's symptom. We analyzed the remaining items to find pairs that showed significant differences via independent *t* tests in likelihood ratings scores between consistent and inconsistent presentations ($ps < .05$). We selected 62 item pairs that evidenced these differences.

In a second round of pretesting, we tested the extent to which participants would rate the CC and IC features differently when presented as they would be in the main experiment (i.e., sets of three features). We created 60 sets of features that would look like the final experiment, that is, have one feature from each of the three risk factor categories. We created a CC and IC version of each, and we recruited a sample of 35 participants with more clinical experience than participants recruited for the first pretest (i.e., clinical interns who had already progressed past the initial years of graduate or professional training in clinical psychology, psychiatry, and social work). As compensation, each participant was entered into a drawing to win a $30 online gift certificate.

Similar to Pretest 1, for Pretest 2 we selected sets of three characteristics for use in the main experiment that showed a significant difference in ratings between their CC and IC format. We used a stringent *p* value criterion of less than .01 to identify 30 stimuli (15 pair sets) for the main tests.

RESULTS

Overall Context Effects

When clinicians' 15 likelihood ratings were averaged within condition we observed higher ratings in the CC condition ($M = 54.1$, $SD = 19.3$) than in the IC condition ($M = 30.8$, $SD = 14.2$; Figure 1), $F(1, 44) = 120.1$, $p < .001$, $\eta^2 = .73$.[5] Given these differences, we were interested in how the CC and IC ratings compared to the noncontextualized judgments (NC). There was a significant main effect among the CC, IC, and NC ($M = 53.9$, $SD = 19.9$) conditions, $F(2, 88) = 53.1$, $p < .001$, $\eta^2 = .54$. Planned paired *t* tests comparing each of the context conditions to the NC ratings revealed a

---

[4]Clinicians provided NC ratings as part of a series of other posttest ratings for the diagnostic symptoms, including prevalence, diagnosticity, abnormality, and importance to diagnosis. None of these other ratings were significantly related to the results and are not discussed further. Information on these ratings can be acquired from the authors.

[5]An eta-squared value greater than or equal to .25 is considered a large effect (Tabachnick & Fidell, 2007).
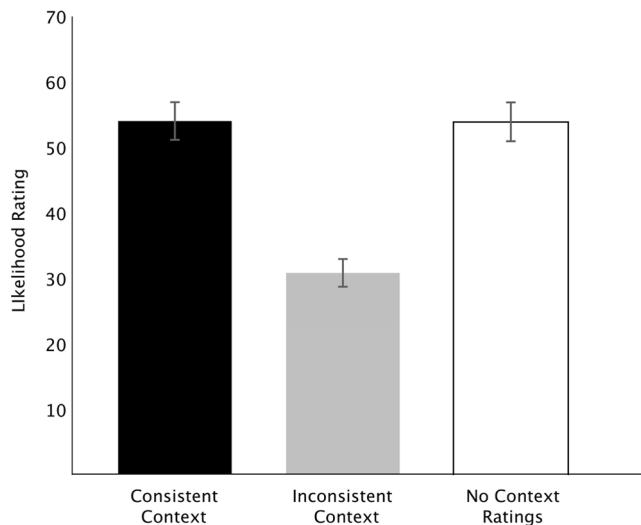
FIGURE 1 Average likelihood ratings for No Context ratings compared to Consistent Context and Inconsistent Context conditions.

significant difference between the IC and NC ratings, $t(44) = 8.4$, $p < .001$, $d = 1.2$, but no difference between CC and NC ratings, $t(44) = .04$, $p = .96$, $d = 0$ (Figure 1).

Context Effects by Individual Symptoms

We constructed difference scores to examine how context affected individual symptoms by subtracting each clinician's IC rating for that symptom from their CC rating for the same symptom (i.e., context effect scores; see Table 1). We were interested in whether judgment ratings on any one symptom were more affected by contextual information than those ratings on any other symptom, and we did not observe a significant effect in this regard, $F(14, 616) = 1.5$, $p = .09$, $\eta^2 = .03$.[6]

Context Effects by Individual Clinicians

Table 1 also shows that each symptom received a wide range of context effect scores, suggesting that clinical impressions varied as to observed context effects. A Kendall's W test to assess overall agreement between clinicians' context effect scores confirmed this, Kendall's $W = .03$, $df = 14$, $p = .17$.

As a second test of variability by clinician, we ranked for each clinician the context effect scores across the 15 symptoms. Diagnostic symptoms ranked 1 had the largest, positive context effect score and 15 the lowest. We tallied how many times a symptom ranked in the

[6]This analysis does not differ if we collapse the 15 symptoms' context effect scores into the four categories provided in the *DSM–IV* (aggression, lying or theft, rule breaking, and property destruction) and conduct the analysis of variance over those scores, $F(3, 132) = 1.1$, $p = .33$, $\eta^2 = .02$

TABLE 3
Number of Times Each Diagnostic Symptom Received a Context Effect Score Within Listed Rank Groupings

	Rank Groupings				
Diagnostic Symptom	1–3	4–6	7–9	10–12	13–15
Run away from home overnight	18	8	5	8	6
Bullies others	12	9	13	8	3
Destroyed property without fire	14	11	7	7	6
Cruel to animals	16	8	11	5	5
Stays out at night	15	8	5	11	6
Truant from school	11	14	8	7	5
Broken into a house	12	11	6	9	7
Used a weapon for harm	11	10	12	6	6
Stolen without confronting a person	11	7	11	10	6
Initiates fights	9	10	12	7	7
Stolen while confronting a person	12	8	7	10	8
Lies to avoid and obtain	9	11	7	9	9
Set fires	11	6	10	7	11
Cruel to people	5	14	10	9	7
Forced sexual activity	7	10	11	8	9

Note: Ties were all given the higher rank. As such, fewer symptoms were possibly ranked in the lower ranks.

"top three," "bottom three," and ranks in between (Table 3). No single symptom appeared more often as the most, Rank 1 to 3: $\chi^2(14, N = 173) = 13.6$, $p = .4$, or least influenced by context, Rank 13 to 15: $\chi^2(14, N = 101) = 7.8$, $p = .8$. Rank groupings in between top and bottom also did not vary, Ranks 4 to 6: $\chi^2(14, N = 145) = 7.8$, $p = .9$; Ranks 7 to 9: $\chi^2(14, N = 135) = 11.3$, $p = .6$; Ranks 10 to 12: $\chi^2(14, N = 121) = 4.5$, $p = .9$.

Finally, we examined the context effect rank orderings for individual clinicians (for a similar approach, see Kim & Ahn, 2002). Specifically, we averaged the

TABLE 4
Average Context Effect Score by Rank Across the 15 Diagnostic Symptoms

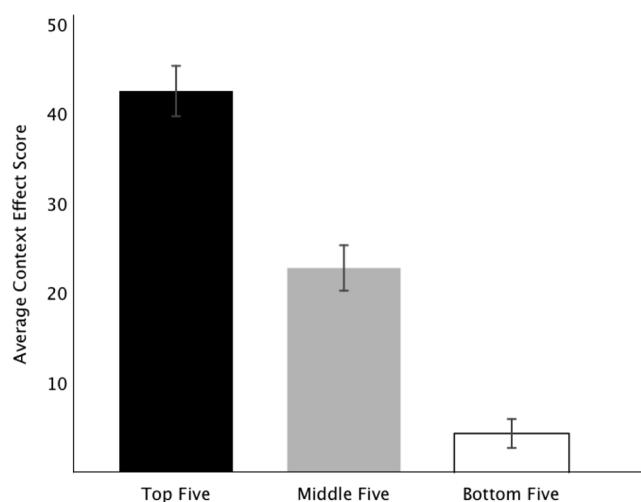| Rank	M Context Effect Scores	SD	Minimum	Maximum
1	54.82	20.9	15	88
2	46.71	19.6	10	88
3	41.47	19.8	0	85
4	36.82	19.0	0	80
5	32.84	19.4	0	75
6	29.87	19.1	0	75
7	26.56	18.1	0	70
8	22.87	17.4	0	70
9	18.60	16.5	−1	65
10	16.13	16.1	−5	60
11	13.29	14.0	−10	50
12	9.84	12.2	−10	45
13	6.07	11.4	−10	43
14	2.02	11.8	−30	40
15	−9.51	13.7	−50	30

FIGURE 2   Average context effect scores grouped by rank order. *Note:* Context effect scores were calculated by subtracting each participant's Inconsistent Context (IC) rating from their own Consistent Context (CC) rating. Higher context effect scores represent greater CC ratings relative to IC ratings. The bar labeled Top Five represents the average of Rank 1 through 5, Middle Five the average of Rank 6 through 10, and Bottom Five the average of Rank 11 through 15.

context effect score for each rank across the participants (i.e., we averaged the context effect score across all participants for each of the 15 ranks; see Table 4). Across individual clinicians' ranks, context effect scores varied significantly, $F(14, 616) = 157.9$, $p < .001$, $\eta^2 = .78$. It is important to note that we also observed a significant linear relation across the 15 ranks, $F(1, 44) = 271.9$, $p < .001$, $\eta^2 = .86$. In Figure 2 we present a simpler presentation of these effects, in which we collapsed across the ranks to create top, middle, and bottom ranks. Analyses of these collapsed ranks also revealed a significant main effect, $F(2, 88) = 212.2$, $p < .001$, $\eta^2 = .82$. Paired t tests revealed significant differences between each of the three rank groups: top five ($M = 42.5$, $SD = 18.8$) versus middle five ($M = 22.8$, $SD = 17.0$), $t(44) = 13.5$, $p < .001$, $d = 2$; top five versus bottom five ($M = 4.3$, $SD = 10.8$), $t(44) = 16.3$, $p < .001$, $d = 2.4$; middle five versus bottom five, $t(44) = 11.2$, $p < .001$, $d = 1.6$. The differences between ranks indicate that clinicians do not judge all symptoms equally. Rather, context exerts its effects on clinicians' impressions for some symptoms and not others.

DISCUSSION

This study yielded two main findings. First, clinicians rated hypothetical child patients displaying a conduct disorder symptom as more likely to receive a diagnosis when the contexts in which they express their symptoms were consistent with conduct disorder, relative to incon-

sistent. Second, these effects varied across symptoms, with low agreement as to which judgments of specific symptoms were most influenced by context effects. Thus, although the *DSM* recommends ruling out diagnoses in the presence of explanatory contextual information, contextual information influences clinicians' judgments of some symptoms. In fact, it appears that context can influence clinicians to believe a symptom is more likely to be present than when no context is present.

Surprisingly, clinicians making ratings in the IC condition reduced their ratings compared to the noncontextualized NC ratings. One might surmise that these children should be even *more* worrisome to clinicians because they express symptoms without a contextual explanation (Wakefield et al., 2002). It is possible that a clinician seeing that a child comes from a seemingly "normal" background and also expresses a conduct disorder symptom may interpret this symptom as a lesser form of disruptive behavior than an actual symptom. Alternatively, the clinician may ignore the diagnostic symptom as an unreliable index of the disorder. These issues merit further study.

Is Using Context Useful or Harmful?

An interesting issue is whether context effects are inherently beneficial or detrimental to clinical decision making. We see potential advantages and disadvantages. First, a disadvantage is that a child who expresses symptoms in a context inconsistent with this expression could be denied much needed treatment or experience delays in receiving treatment. This could have implications for identifying differences in diagnostic rates among different socioeconomic class members who may differentially experience contexts consistent with a conduct disorder diagnosis. At the same time, contextual information might serve as a useful guide to interpreting behaviors potentially indicating a conduct disorder diagnosis. Indeed, we designed the contextual factors displayed in the CC condition based on empirical research documenting the risks they pose for the development of conduct disorder (Burke et al., 2002). Thus, these factors may have been effective heuristics for forming clinical impressions for children rated in this condition. In any event, clearer answers to these issues await further study.

Limitations

The limitations of our study reveal directions for future research. First, our sample was predominantly Caucasian, 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 diagnostic decisions. Given our effects and their variability across clinicians, it might be that context effects greatly impact the diagnostic decisions of some clinicians and not others. Furthermore, it is an open question whether contextual information influences the treatments that clinicians prescribe as appropriate to a given patient. Relatedly, we did not examine factors that may predict whether contextual information will affect clinicians' judgments. Clinicians develop personal theories of the interrelations among a disorder's symptoms and, in particular, beliefs as to the symptoms that "have to be present" for one to diagnose a patient (Kim & Ahn, 2002). Similarly, these theories might be intertwined with theories about contextual information, and future research needs to focus on understanding these theories.

Implications for Research, Policy, and Practice

Our findings have important implications for clinical science and practice. In particular, our findings identify a factor that might strengthen diagnostic classification of conduct disorder. That is, for judgments of some symptoms clinicians use contextual factors indicated by prior work as posing risk for conduct disorder. It would be important, then, for future research to test how diagnostic systems could promote consistent use of contextual information. Alternatively, it is possible that a symptom criterion should be eliminated if clinicians cannot reliably take into account related contextual information when making clinical decisions. Thus, future work should identify ways to capitalize on clinical applications of contextual information to conduct disorder symptom judgments.

REFERENCES

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

Burke, J. D., Loeber, R., & Birmaher, B. (2002). Oppositional defiant disorder and conduct disorder: A review of the past 10 years, part II. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*, 1275–1293.

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.

Kraemer, H. C., Stice, E., Kazdin, A. E., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry*, *158*, 848–856.

Qualtrics Labs, Inc. (2005). *Qualtrics Labs, Inc.* (Version 7546) [Computer software]. Provo, UT: Author.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

Wakefield, J. C., Pottick, K. J., & Kirk, S. A. (2002). Should the *DSM-IV* diagnostic criteria for conduct disorder consider social context? *American Journal of Psychiatry*, *159*, 380–386.

Weisz, J. R., Hawley, K. M., & Jensen Doss, A. (2004). Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. *Child and Adolescent Psychiatric Clinics of North America*, *13*, 729–815.