Psychological Assessment
2001, Vol. 13, No. 4, 549–565

1040-3590/01/$5.00 DOI: 10.1037//1040-3590.13.4.549

Using Cognitive Science Methods to Assess the Role of Social Information Processing in Sexually Coercive Behavior

Teresa A. Treat, Richard M. McFall, Richard J. Viken, and John K. Kruschke
Indiana University, Bloomington

Seventy-four undergraduate men completed cognitive performance tasks assessing perceptual organization, classification, and category learning, as well as self-report measures relevant to sexual coercion. The stimuli were slides of Caucasian women who varied along affect and physical exposure (i.e., sensuality) dimensions. Data were analyzed using a weighted multidimensional scaling model, signal-detection theory analyses, and a connectionist learning model (RASHNL; J. K. Kruschke & M. K. Johansen, 1999). Individual differences in performance on the classification and category-learning tasks were congruent with individual differences in perceptual organization. Additionally, participants who showed relatively more attention to exposure than to affect were less sensitive to women's negative responses to unwanted sexual advances. Overall, the study demonstrates the feasibility and utility of cognitive science methods for studying information processing in psychopathology.

Clinical psychologists regularly refer to cognitive models when investigating and treating psychological disorders. However, the constructs and methods in these models seldom resemble the constructs and methods found in the models of contemporary cognitive science (see Lang, 1988; MacLeod, 1993; McFall & Townsend, 1998; McFall, Treat, & Viken, 1997, 1998). This disconnect between clinical science and cognitive science is unfortunate. Advances in our understanding of the etiology, treatment, and prevention of psychopathology might be accelerated if clinical scientists attended more closely to theoretical and methodological developments in cognitive science. The present research was based on this premise. We explicitly adapted experimental tasks and analytical techniques from cognitive science in an effort to test clinical hypotheses regarding the role of cognitive processes in men's sexually coercive behavior toward women.

Role of Cognition in Sexual Coercion

Clinical theorists are in agreement that cognitive variables play a critical role in the instigation of sexually coercive behavior between acquaintances, which accounts for the overwhelming majority of sexually aggressive episodes (Abbey, Ross, McDuffie, & McAuslan, 1996; Drieschner & Lange, 1999; Ward, Hudson, Johnston, & Marshall, 1997; Yeater & O'Donohue, 1999). Research in this area initially focused on cognitive constructs drawn from the social and personality traditions, such as attitudes and beliefs. Introspective access to these constructs is assumed, and they typically are assessed using self-report methods. Extensive research in this area has documented clearly that men who report engaging in sexually coercive behavior show more distorted, negative, and rape-supportive attitudes, beliefs, and expectations about women, sex, and sexual violence (see Drieschner & Lange, 1999, and Ward et al., 1997, for reviews).

More recently, clinical researchers have begun to focus on cognitive constructs drawn from cognitive psychology, such as perception, attention, memory, classification, and learning. Several clinical theorists (Hudson & Ward, 2000; Johnston & Ward, 1996; McFall, 1990; Schewe & O'Donohue, 1993; Segal & Stermac, 1990; Ward et al., 1997) have proposed that the likelihood of sexually coercive behavior is a function, in part, of how men organize and process information about women. Extensive research by social and cognitive psychologists has challenged the belief that humans have introspective access to the important cognitive processes underlying their own perception and action (MacLeod, 1993; Nisbett & Wilson, 1977; Wilson, 1994). Thus, research aimed at testing the hypothesized relationship between information processing and sexual coercion should rely on performance-based methods of assessing cognitive processing.

McFall's social information-processing model provides the overarching theoretical framework for the current work (1982, 1990). According to the model, behavior that is judged to be sexually coercive could result from difficulties in one or more processing stages: a *decoding stage*, in which incoming stimulus information is perceived, organized, and interpreted; a *decision-making stage*, in which responses are selected; and an *enactment stage*, in which the selected response is executed. The present work focuses on decoding processes, as decoding difficulties propagate through the information-processing system.

Two lines of research are relevant to the investigation of the role of decoding in sexual violence: One evaluates liberal *decision criteria*, or *biases*, in men's perception of women's affective or sexual-interest cues; the other examines *insensitivity* to women's affective or sexual-interest cues.

Abbey and colleagues initiated the first line of research. They have documented repeatedly that men, both participants and ob-

Teresa A. Treat, Richard M. McFall, Richard J. Viken, and John K. Kruschke, Department of Psychology, Indiana University, Bloomington.

For helpful comments on a previous version of this article, we thank Michael Erickson and Robin Weersing.

Correspondence concerning this article should be addressed to Teresa A. Treat, who is now at Yale University, Department of Psychology, P.O. Box 208205, New Haven, Connecticut 06520. Electronic mail may be sent to teresa.treat@yale.edu.

servers, rate women as displaying more sexual-interest cues during interpersonal interactions or in written scenarios than women (both participants and observers) rate themselves (e.g., Abbey, 1982; Abbey & Harnish, 1995; Abbey, McAuslan, & Ross, 1998; Abbey, Zawacki, & McAuslan, 2000; Shotland & Craig, 1988). These findings are consistent with the implication that men show a liberal decision criterion in their perception of women's affect or sexual-interest cues. Numerous theorists have suggested that men who display this bias should be at increased risk of exhibiting sexually coercive or violent behavior. Consistent with this prediction, the presence of this bias is associated with a self-reported history of sexually coercive behaviors (Bondurant & Donat, 1999; Shea, 1993) and with rape-supportive attitudes (Abbey & Harnish, 1995; Kowalski, 1993).

In the second line of research, several researchers have demonstrated associations between sexual violence and errors in judgments about women's affect in videotaped heterosexual interactions (e.g., Malamuth & Brown, 1994; McDonel, 1995; McDonel & McFall, 1991; Murphy, Coleman, & Haynes, 1986). For example, incarcerated male participants in a study by Lipton, McDonel, and McFall (1987) viewed a series of videotaped heterosexual interactions and classified both the man's and the woman's affect in each interaction as romantic, positive, neutral, negative, or bad mood. Rapists were less accurate than either violent or nonviolent control participants in this classification task. As expected, rapists' decoding deficit was most pronounced when judging women's negative cues: the accuracy of participants' judgments of women's negative cues accounted for 14% of the variance in group membership. In a follow-up study, McDonel and McFall (1991) demonstrated that college men's accuracy in judging women's negative cues correlated negatively with their acceptance of rape-supportive beliefs, their self-reported proclivity to rape if they were assured that they would not be caught and punished, and their judgments about the justifiability of continuing unwanted sexual advances. This line of research is consistent with the possibility that insensitivity to women's affective or sexual-interest cues increases the likelihood of sexual violence.

These two research lines illustrate the promise of investigating the role of decoding in sexual violence. However, future work in this area would benefit from using performance-based tasks, which ask for classification decisions of standardized stimuli. Additionally, increasing the sample size of stimulus classifications would allow researchers to use signal-detection theory approaches to tease apart sensitivity and bias explanations for the results (Green & Swets, 1966; Macmillan & Creelman, 1991; McFall & Treat, 1999).

The present work examines a third decoding-related hypothesis specified by several clinical theorists (Johnston & Ward, 1996; McFall, 1990; Schewe & O'Donohue, 1993; Segal & Stermac, 1990; Ward, Hudson, Johnston, & Marshall, 1997): namely, that the likelihood of sexual violence should be greater among men who attend relatively less to information about women's affect or sexual interest and relatively more to women's physical sexual attributes (e.g., physical exposure, sensuality, provocativeness, and sexual attractiveness). Conversely, the likelihood of sexual violence should be lower among men who attend to information regarding women's sexual interest and affect. In short, men's heterosexual behavior should be predictable, in part, from knowledge about their perceptual organization of information about women. The trick, of course, is to find valid ways to assess perceptual organization.

To our knowledge, few clinical researchers to date have used performance-based methods drawn from cognitive science to assess individual differences in men's relative attention to women's affect and physical sensuality. Prevailing theories in cognitive psychology predict that perceptual processes lie upstream from the operation of other higher order cognitive processes, such as decision making, classification, memory, and learning. Thus, the present work assesses individual differences in men's performance on tasks that assess attention, classification, and learning, and then evaluates their covariation, not only among themselves but also with measures relevant to sexual coercion. Because many readers may be unfamiliar with the theoretical and measurement models used in this research, we turn now to an overview of these models.

Contemporary Cognitive Science

Cognitive scientists have developed a number of performance-based tasks and analytical techniques relevant to the assessment of normative cognitive processes, which presumably are invariant across persons apart from random fluctuations. Additionally, they have developed precise conceptual and quantitative models of the interrelationships among cognitive processes. In 1986, Nosofsky initiated a resolution to prior difficulties in predicting performance across tasks assessing identification, classification, memory, and category learning; he proposed that task-specific transformations of a single underlying perceptual representation *could* predict performance across divergent tasks. In a series of articles, Nosofsky (e.g., 1986, 1992a, 1992b) then demonstrated both the power of multidimensional scaling (MDS) techniques to assess perceptual organization and the use of sophisticated mathematical models to predict performance across tasks, by using task-specific transformations of MDS solutions. Nosofsky's model predicted performance on category-learning tasks only at a fixed moment in the learning process, however; it did not address the process of learning over time as a function of feedback. Thus, Kruschke and colleagues (Kruschke, 1992; Kruschke & Johansen, 1999) have proposed several connectionist models that specify the mechanisms responsible for the process of category learning; we describe in detail one of these models, RASHNL (which stands for Rapid Attention SHifts 'N' Learning), in a later section.

Cognitive scientists have developed solutions for another long-standing problem of interest to clinical scientists as well: how to tease apart and quantify the independent contributions of specific sources of information to performance in identification, classification, and category-learning tasks. One solution, which we describe below, uses methods drawn from signal-detection theory (SDT; Green & Swets, 1966; Macmillan & Creelman, 1991).

Cognitive scientists initially developed and refined these tasks and analytical techniques to explore normative processes of non-social human cognition, typically using simple artificial stimuli. In principle, however, there is no reason why these same tasks and techniques cannot be adapted for studying individual differences in social information processing. In the following subsections, we describe in more detail these three specific methods of quantifying cognitive performance, which should be of interest to clinical scientists: MDS, RASHNL, and SDT.

Multidimensional Scaling

MDS provides a multidimensional spatial representation of a person's perceptual organization of a stimulus set (Davison, 1992; Schiffman, Reynolds, & Young, 1981). Typically, MDS models use similarity ratings of all possible stimulus pairs as input. In the present study, for example, undergraduate men rated the similarity of all possible pairs of 14 photographs of women that varied naturally along two dimensions of particular theoretical interest: physical exposure, which correlates with sensuality and provocativeness, and facial affect. MDS models provide a spatial representation of these 14 stimuli, such that stimuli judged to be very similar are scaled closer together than stimuli judged to be very dissimilar. Cognitive psychologists typically refer to the MDS output as a "psychological space," as it represents participants' perceptual organization of the stimuli.

The weighted MDS (WMDS) algorithm captures individual differences in the importance of the stimulus dimensions to participants' similarity ratings by estimating individual-specific weights for the dimensions of the group multidimensional space. The WMDS model assumes that all participants perceive the relative positioning of stimuli along each dimension in the same way but allows individuals to stretch and shrink the dimensions of the group configuration differentially. Figure 1 depicts two idealized perceptual organizations corresponding to extremely exposure-oriented (EO) and affect-oriented (AO) participants. An extremely EO participant bases his similarity ratings almost entirely on exposure and virtually ignores affect. Thus, exposed and unexposed women are well separated in this participant's unique perceptual space, but happy and unhappy women are scaled very close to one another. This EO participant would receive a very high attention weight for exposure and a very low attention weight for affect. In contrast, an extremely AO participant stretches the group configuration along the affect dimension and shrinks the group configuration along the exposure dimension, resulting in a large attention weight for affect and a small attention weight for exposure.

Nosofsky capitalized conceptually on the WMDS model to account for performance across classification, identification, memory, and learning tasks within a single framework; he assumed that the underlying stimulus representation remained constant across tasks, but he allowed task-specific stretching and shrinking of the perceptual dimensions (Nosofsky, 1986, 1992a, 1992b). This idea makes intuitive sense. For example, increasing attention to affect might enhance discrimination between old and new stimuli in a recognition memory task, if previously viewed stimuli tended to show negative affect and novel stimuli tended to show positive affect. Similarly, increasing attention to exposure might facilitate performance on an exposure category-learning task in which exposed and unexposed women are in different categories.

Our use of MDS techniques in the present study differs from typical applications in social or clinical psychology, where investigators use MDS atheoretically, as an exploratory tool, in the hope of discovering or revealing unknown dimensions underlying social perception (see Jones, 1983, for a review). In contrast, we have adopted the theory-driven, hypothesis-testing approach of cognitive scientists, who typically use structured, well-characterized stimuli for which the underlying dimensions influencing participants' perceptions have been scaled normatively. Our application

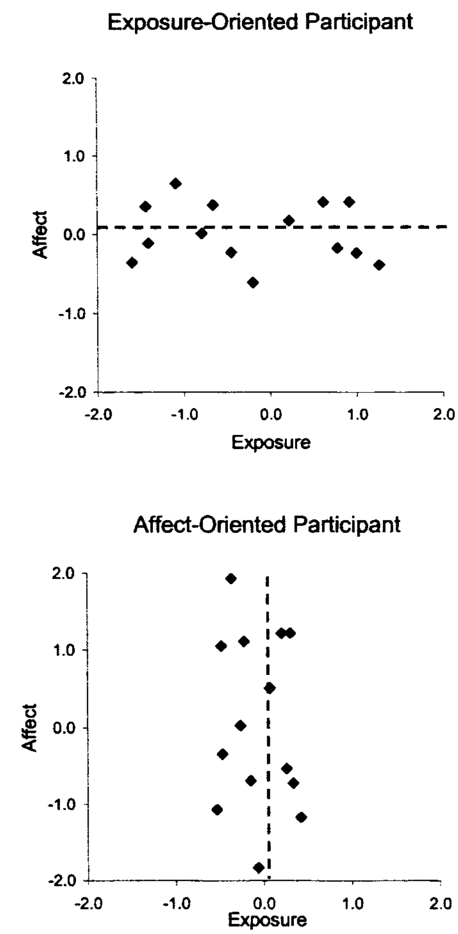

Figure 1. Extremely affect-oriented and exposure-oriented participants' unique perceptual organizations (hypothetical).

differs from typical applications in cognitive science, however, by focusing on individual differences and by using more socially relevant stimuli (Treat et al., in press).

RASHNL: A Connectionist Model of Category Learning

Whereas Nosofsky's model addressed performance at a fixed moment in the category-learning task only, Kruschke's generalizations of this model captured changing performance during the process of learning, as well (Kruschke, 1992; Kruschke & Johansen, 1999). Kruschke and Johansen's RASHNL model instantiates several mechanisms by which participants might learn a category structure, and allows researchers to examine the role of these mechanisms in participants' category learning.

Fitting a formal process model to the learning data in the present study should enhance our interpretation of these data in terms of cognitive science constructs. We also use analysis of variance (ANOVA) techniques to evaluate our predictions about the learning data. An adequate process model should reproduce the major quantitative and qualitative effects observed in the ANOVAs. Supplementing the ANOVAs with formal process modeling should deepen our understanding of the ANOVA results by allowing us to interpret them within the theoretical framework of cognitive science. To facilitate understanding of the complexities of RASHNL, we emphasize the ways in which the model might be able to account for the expected ANOVA findings in terms of cognitive science concepts.

In the present study, participants completed two category-learning tasks in which they classified numerous photographs of women into one of two unspecified categories and received trial-by-trial feedback. Feedback was based on the woman's affect for one task and on the woman's exposure for the other task. We anticipated that EO participants would learn the exposure task more quickly than AO participants and that AO participants would learn the affect task more quickly than EO participants. In other words, we expected a significant Group (EO vs. AO) × Task (affect vs. exposure) interaction. The presence of this interaction would tell us little about *how* participants learned the affect and exposure category structures. However, fitting RASHNL to participants' category-learning data would allow us to draw preliminary inferences about the relevance of RASHNL's learning mechanisms to EO and AO participants' performance, as well as to evaluate RASHNL's applicability to an evaluation of individual differences in learning processes with complex, socially relevant stimuli.

Kruschke and Johansen's (1999) RASHNL model instantiates three relevant mechanisms that presumably influence participants' learning. The first mechanism is the perceived relative salience of the stimulus dimensions, the second is the rate at which dimension-specific attention shifts, and the third is the rate at which associations form between stimuli and correct category labels. The first mechanism, the perceived relative salience of stimulus dimensions, should facilitate performance on a learning task that is congruent with the group's perceptual organization and degrade performance on a learning task that is incongruent with the group's perceptual organization. For example, AO participants, who show relatively greater attention to affect than to exposure, should learn the affect task more quickly than EO participants, as the affect category structure is more congruent with AO than with EO participants' perceptual organizations. Thus, the fit of RASHNL to participants' data should improve substantially when the relative salience parameter is estimated separately for AO and EO participants, rather than estimated across AO and EO participants (i.e., when the parameter estimate is "group specific"). Inclusion of group-specific estimates of this parameter should allow the model to predict the expected Group × Task interaction.

The second mechanism, the rate at which dimension-specific attention shifts, suggests that learning may result, in part, from increasing attention to stimulus dimensions that maximally separate stimuli in different categories, and from decreasing attention to stimulus dimensions that minimally separate stimuli in different categories. Consider, for example, how the EO participant depicted in the top panel of Figure 1 might learn the affect category structure. This participant's perceptual organization provides minimal separation of women on the basis of their affect; numerous women fall near the relevant category boundary, depicted by a dashed line in the figure. The EO participant might solve the affect task problem, however, by decreasing attention to (i.e., shrinking) the exposure stimulus dimension and increasing attention to (i.e., stretching) the affect stimulus dimension—in essence, by changing his perceptual organization such that it is more similar to that of the AO participant. Similarly, the AO participant depicted in the bottom panel of Figure 1 could learn the exposure category structure by increasing attention to the exposure dimension and decreasing attention to the affect dimension. This mechanism essentially reflects whether participants modify their perceptual organizations when a different distribution of attention to affect or exposure may be more appropriate.

The third mechanism, the rate at which associations form between stimuli and correct category labels, suggests that learning occurs, in part, as participants link different regions of the psychological space with the correct category labels. For example, AO participants learn the affect category structure in part by associating women who show positive affect with one category label and women who show negative affect with the other category label.

Modeling will address the relevance of these three mechanisms to EO and AO participants' learning data. Additionally, we evaluate the necessity of group-specific parameter estimates by comparing the quantitative and qualitative fits of models with and without each group-specific parameter. Inclusion of a group-specific relative salience parameter should facilitate prediction of the expected Group × Task interaction. In contrast, a model including a group-specific estimate of only the attention-strength shift rate or the association learning rate parameter should predict a Group effect but should not predict the expected Group × Task interaction. All three models will include a group-specific parameter and will have the same number of parameters, but we predict that the qualitative and quantitative fit of the group-specific relative salience model will be greater.

Signal-Detection Theory

Overall accuracy (i.e., percentage correct) in a classification task depends on two separable factors that are familiar concepts from SDT (Green & Swets, 1966; Macmillan & Creelman, 1991; McFall & Treat, 1999): *sensitivity*, how well the participant discriminates between stimuli in different categories; and *bias*, where the participant draws the boundary between the two categories. Although cognitive scientists commonly separate the contribution of sensitivity and bias to performance on a variety of cognitive tasks, clinical scientists rarely have made this distinction in clinical research, even though deficits in information processing could reflect either influence. For example, errors in the affect category-learning task described above might be due either to an insensitivity to differential cues for positive and negative affect (Lipton et al., 1987) or to a bias toward viewing most woman as showing positive affect (Abbey, 1982). Using SDT to analyze performance allows us to tease apart the relative contributions of sensitivity and bias. In the present study, we expected individual differences in participants' perceptual organizations to predict individual differences in participants' relative *sensitivity* to affect versus exposure cues, in both the classification and learning tasks. However, we also examined the extent to which *bias* contributed to classification errors on the learning tasks.

Present Study

This study illustrates the adaptation of tasks and analytical techniques drawn from cognitive science to the assessment of problems of interest to clinical scientists, specifically, the role of cognitive processing in sexually coercive behavior (Johnston &

Ward, 1996; McFall, 1990; Schewe & O'Donohue, 1993; Segal & Stermac, 1990). Before specifying the research questions to be evaluated in the present study, we first describe our decisions regarding a recurring sampling dilemma faced by all researchers in this content area.

Sampling Issues

Obtaining a representative sample of men who exhibit sexually coercive behavior has proven to be an impossibility for researchers thus far. Commonly used sampling strategies are to study men who have been incarcerated for sexual assault or who admit engaging in sexually coercive behavior. Both approaches have problems. The first samples only men who are reported, arrested, and convicted of sexual assault—a small minority of the number of men engaging in sexually coercive behavior (Koss, Gidycz, & Wisniewski, 1987). The second may be affected by social desirability and by failure to perceive one's behavior as coercive. In the present research, it would be paradoxical to expect men who are relatively inattentive to women's affect (i.e., men classified as EO) to attend to women's negative reactions to their coercive behavior. In this particular theoretical context, therefore, we did not expect to find an association between performance on the cognitive tasks and a self-reported history of sexually coercive behavior. Given the potential interest in this question from alternative theoretical perspectives, however, we also conduct exploratory analyses to evaluate whether EO participants report engaging in more sexually coercive behavior on a version of the Coercive Sexuality Scale (CSS; Rapaport & Burkhart, 1984).

Instead, we examined the link between decoding processes and sexual behavior by assessing the association between cognitive task performance and performance on a heterosocial perception measure that is linked theoretically to the likelihood of exhibiting sexually coercive behavior. On the Heterosocial Perception Survey (HPS; McDonel & McFall, 1991), participants rate the justifiability of a man continuing to make sexual advances toward a woman whose response to his advances is increasingly negative. We expect that men who attend relatively little to women's affective cues (i.e., men classified as EO) will be less sensitive to the negativity of the woman's response. In other words, EO participants' justifiability ratings should decrease much *less* than AO participants' ratings, as the woman's response becomes increasingly negative.

We examined our research questions in an unselected sample of undergraduate men. In the area of sexual coercion, college men serve not as convenience samples but as samples of intrinsic theoretical interest, for two reasons: the difficulties in obtaining a representative sample of men exhibiting sexually coercive behavior, as discussed previously, and the particularly high prevalence rates of and research interest in sexual coercion among college men (Berkowitz, 1992; Craig, 1990; Koss et al., 1987; Malamuth, Sockloskie, Koss, & Tanaka, 1991). Prevalence estimates vary depending on whether assessment instruments focus primarily on coerced intercourse (e.g., variants of Koss & Oros's [1982] Sexual Experiences Survey) or on a broader spectrum of coercive behaviors (e.g., variants of Rapaport & Burkhart's [1984] Coercive Sexuality Scale). In studies adopting the latter assessment approach, the majority of college men report engaging in some form of sexual coercion (i.e., 61%, 64%, and 55%, in work by Rapaport & Burkhart, 1984, Rapaport & Posey, 1991, and Hannan & Burkhart, 1993, respectively).

Research Questions

The present study addresses five interrelated questions: First, does the WMDS model provide a useful description of individual differences in men's attention to women's affect and physical exposure? Second, do individual differences in perceptual organization predict individual differences in implicit classification, such that EO participants' classifications are more sensitive to exposure than AO participants' classifications, and AO participants' classifications are more sensitive to affect than EO participants' classifications? Third, do individual differences in perceptual organization predict individual differences in category learning, such that EO participants learn the exposure category structure more quickly than AO participants, and AO participants learn the affect category structure more quickly than EO participants? Both the second and third questions examine the congruence of participants' classification behavior with their relative attention to affect and exposure, as quantified by the WMDS model. The implicit classification task assesses categorization in the absence of systematic feedback, whereas the category-learning task examines categorization in the presence of systematic feedback. Thus, the former question evaluates whether participants spontaneously impose a category structure that is consistent with their perceptual organization, whereas the latter investigates whether participants learn a category structure more quickly when it is more congruent with their perceptual organization.

Fourth, does fitting RASHNL to the category-learning data provide a helpful description of the mechanisms underlying participants' performance on the category-learning tasks? And fifth, are individual differences in perceptual organization associated with responses on a social perception measure relevant to sexually coercive behavior, such that EO participants show decreased sensitivity to women's negative responses to unwanted sexual advances?

Method

Design

Seventy-four undergraduate men at a midwestern university received partial credit toward their course requirements in introductory psychology by participating in the study. Information on the age and ethnicity of the sample was not obtained, as unselected samples drawn from this undergraduate participant pool are very uniform (i.e., 90% report their age as between 18 and 22 and their ethnicity as Caucasian). Participants completed five tasks: a similarity-ratings task, an implicit classification task, a category-learning task with feedback based on affect classification (high or low), a category-learning task with feedback based on exposure classification (high or low), and two paper-and-pencil self-report measures relevant to sexual coercion.

Development of Stimuli for Cognitive Tasks

The stimuli used in the similarity-ratings, implicit classification, and category-learning tasks were color slides of Caucasian women, taken from newsstand magazines and mass-marketing catalogs; we included only Caucasian women in the stimulus set for reasons of homogeneity described above. The slides were selected in a series of pilot studies conducted by a female experimenter. Initially, two groups of undergraduate men ($ns = 23$ and 21) provided normative judgments of 106 potential stimuli on 10-point scales for 14 attributes (e.g., attractiveness, mood, activity level). Three additional groups of undergraduate males ($ns = 18$, 20, and 24) rated the similarity of all possible pairs of a subset of 26 of these stimuli. In a WMDS analysis of participants' similarity ratings, the two most salient dimensions corresponded to affect (positive to negative) and degree of physical exposure (modest to revealing clothing).

On the basis of these pilot data, we selected another set of 26 stimuli from the original photo pool of 106 stimuli to fill the two-dimensional, affect by exposure, stimulus space as uniformly as possible. We selected several representative stimuli for each quadrant of the stimulus space whenever possible, but stimuli were distributed sparsely in the high-exposure, negative affect quadrant. Two new groups of undergraduate men ($ns = 14$ and 7) made normative judgments on 10-point scales for 14 dimensions, including exposure, affect, attractiveness, and sensuality. Ratings of exposure and affect were uncorrelated, $r(19) = .20$, *ns*. The range of average ratings was similar for exposure (8.00) and affect (7.80), but the variance was greater for ratings of exposure (7.05) than affect (4.66).

Normative ratings of exposure correlated moderately to strongly with ratings of more subjective attributes: sensuality, $r(24) = .64$, $p < .01$; "naughtiness" (vs. "niceness"), $r(24) = .81$, $p < .01$; vulnerability, $r(24) = .722$, $p < .01$; attractiveness, $r(24) = .615$, $p < .01$; and the woman's sexual arousal, $r(24) = .84$, $p < .01$. Others have reported similar associations (Abbey, Cozzarelli, McLaughlin, & Harnish, 1987). The relatively objective characteristic of exposure was selected as the dimension label. Normative ratings of affect correlated moderately to strongly with ratings of likability, $r(24) = .55$, $p < .01$, and approachability, $r(24) = .712$, $p < .01$.

All 26 slides were used in the classification and learning tasks.[1] The normative data were used to classify 12 slides as examples of high exposure and 14 slides as examples of low exposure. Similarly, 12 slides were classified as high positive affect and 14 slides as low positive affect. Only 14 representative slides were used in the similarity-ratings task, as a prohibitively large number of similarity ratings would have been necessary if all 26 slides had been used.

Self-Report Measures

Heterosocial Perception Survey. On the HPS (McDonel & McFall, 1991), participants read three scenarios in which a man is with a woman and is interested in having sex with her for the first time. The three scenarios vary primarily in how long and how well the man has known the woman. At the end of each scenario, the participant reads five descriptions of a sexual advance initiated by the man, as well as the woman's response; these descriptions are identical across the three scenarios. Both the intimacy of the sexual advance and the negativity of the woman's response increase across the five descriptions. The participant's task is to rate the justifiability of continued sexual advances, given the woman's reaction, on a 101-point scale (0 = *absolutely no justification in continuing*; 100 = *completely justified in continuing*). The participant makes a justifiability rating for each of the five sexual advance levels for each of the three scenarios, for a total of 15 justifiability ratings.

We scored the HPS by adding 1 to each participant's justifiability rating for the five levels of sexual advance described after each of the three scenarios, so that the resulting 15 ratings ranged from 1 to 101. A total justification score was calculated, as recommended by McDonel and McFall (1991), by summing participants' responses across all scenarios and all sexual advance levels. In a sample of 50 male college students, the total justification score has proven to be a valid predictor of rape-supportive attitudes and rape proclivity measures, as well as decoding accuracy of female negative cues, but, as expected, not of female positive cues or male cues (McDonel & McFall, 1991).

A novel index, the relative justification score, also was calculated from HPS ratings. A mean justifiability rating was computed for each of the five levels of sexual advance by averaging the justifiability ratings at each level across the three scenarios. Next, four ratios of the average justifiability ratings at Level 1 to the average justifiability ratings at Levels 2 through 5 were computed. The mean of these four ratios was the relative justification score. A value of 1.0 indicated that a participant rated the justifiability of continued sexual advances to be the same across all five sexual-advance levels, whereas larger values indicated that a participant rated the justifiability of continued sexual advances as lower for the four final sexual-advance levels than for the first level. Larger scores indicated greater sensitivity to a woman's increasingly negative responses. The relative justification score index was developed specifically for this study. Emergence of the predicted EO–AO group differences on this measure in the present study would support the predictive validity of this index.

Coercive Sexuality Scale. On the most recent version of the CSS provided to us by Burkhart at the outset of the study, participants reported how frequently they had engaged in each of seven sexually coercive behaviors (Hannan & Burkhart, 1993; Rapaport & Burkhart, 1984; Rapaport & Posey, 1991). The least coercive item was "I have placed my hand on a woman's breast, thigh, or crotch against her wishes," and the most coercive item was "I have had sexual intercourse with a woman against her wishes." Frequency was indicated on a 5-point Likert scale as *never, once, twice, several times*, or *often*.

We scored the CSS by summing the number of distinct coercive acts admitted out of seven possible acts to obtain a total coercion score; scores ranged from 0 to 7. Burkhart (personal communication, June 2001) reported that the test–retest reliability of this index over a 2-week period with a sample of college men exceeded .90. Additionally, the total coercion score derived from several different versions of the CSS shows moderate relationships with rape-supportive attitudes, rape-proclivity measures, and laboratory analogues of sexual aggression (e.g., Hall & Hirschman, 1994; Rapaport & Burkhart, 1984; Rapaport & Posey, 1991).

Procedure

Participants were run in four groups ($ns = 21$, 19, 23, and 21) in a small classroom. They first read and signed the consent form provided by the two experimenters (one male, one female). Then participants were told that the study was an investigation of men's perceptions of women, and that they would be making judgments about a variety of women pictured in slides during the first part of the experiment. Participants then completed the following tasks.

Similarity ratings. Participants rated the similarity of all possible pairs of the 14 stimuli (91 pairs) on a 10-point scale (0 = *very different* to 9 = *very similar*). Order of presentation of pairs and location of individual stimuli (right side or left side of pair) was randomized. Paired slides were projected on a 1.5 × 1.5-m screen for approximately 8 s; participants marked their ratings on an answer form. This task took approximately 15 min (see Footnote 1).

Implicit classification. Participants were told that their task would be to view individual slides of women and judge whether each woman did or did not have an unnamed characteristic. Participants were told that they would have to decide what this characteristic was on the basis of the feedback provided. Participants viewed each of the 26 slides in each of four trial blocks; order of slides varied across blocks. After a 3-s presentation of each slide, participants indicated on a response sheet whether the woman had (*yes*) or did not have (*no*) the characteristic.

At the end of each block of 26 slide presentations and classification judgments, an oral list of the 26 "correct" classifications was presented. Feedback was provided at the end of the block of trials and was random and unrelated to any obvious characteristics of the slides. This was done to

[1] Photocopies of the stimuli and full instructions are available from Teresa A. Treat.

make it difficult for participants to associate specific slides with feedback, so that we could evaluate individual differences in participants' spontaneous classifications of the stimuli in the absence of systematic feedback. Because classifications on the first block were made prior to any feedback, participants were informed that they would have to guess on that block what the characteristic was and to respond accordingly. Participants completed the implicit classification task in approximately 10 min (see Footnote 1).

Category learning. In the affect and exposure category-learning tasks, participants again viewed individual slides of women and judged whether each woman did or did not have the unidentified characteristic of interest. Participants received accurate feedback on the correct classification of each slide before viewing the next. At the start of these two learning tasks, participants were informed that the characteristic determining classifications in the new task might or might not be the same as in the previous task. The affect- and exposure-learning tasks were counterbalanced; half of the participants completed the affect task first, half completed the exposure task first. The two category-learning tasks took approximately 15 min to complete (see Footnote 1).

Sexual coercion measures. Participants then completed the HPS and a 7-item version of the CSS (Burkhart, personal communication, June 2001; Rapaport & Burkhart, 1984; Rapaport & Posey, 1991). To encourage candid responding, we stressed the anonymity of the data collection process to participants before they completed the questionnaires. Participants completed the two questionnaires in approximately 10 min.

Debriefing. After completion of the questionnaires, participants were debriefed and thanked for their participation. The entire experiment lasted approximately 55 min.

Results

Question 1: Does the WMDS Model Provide a Useful Description of Individual Differences in Men's Attention to Women's Affect and Physical Exposure?

Data reduction. A nonmetric WMDS model was fit to participants' similarity ratings using ALSCAL (Young & Lewyckyj, 1996). The euclidean distance metric was assumed in all MDS and modeling analyses.[2] A two-dimensional configuration was imposed, and the stimulus coordinates were constrained to be the average normative ratings for exposure and affect. Thus, only attention weights for exposure and affect were estimated.[3] S-stress and stress (badness-of-fit measures ranging from 0 to 1) were .51 and .35, respectively. R^2, the proportion of variability in participants' transformed similarity ratings that could be explained by the MDS model, was .35. The fit of the model to participants' similarity ratings was adequate, given the constraints placed on the analysis and the predictive utility of the attention weights in later analyses.[4] Attention weights for exposure ranged from .22 to .88 (M = .50), and attention weights for affect ranged from .06 to .44 (M = .25).[5]

Classifying participants as EO or AO. Further analyses of these data were based on each participant's flattened subject weight (FSW), which transforms each participant's pair of attention weights for exposure and affect into a single index of relative attention (MacCallum, 1977; Schiffman et al., 1981; Young & Lewyckyj, 1996).[6] FSW was positive if the participant's relative attention to exposure was greater than the sample's average relative attention to exposure and was negative if it was less than the sample's average relative attention to exposure. Participants whose FSW fell in the top third of the sample (those who showed the highest relative attention to exposure) were classified as exposure oriented (EO; n = 24); those participants whose FSW fell in the bottom third of the sample were classified as affect oriented (AO; n = 24).

Summary. The WMDS model provided a parsimonious description of participants' similarity-ratings data. As substantial individual differences in participants' relative attention to expo-

[2] Selecting the appropriate distance metric is a critical first step when fitting MDS models, as the euclidean and city-block metrics frequently provide different estimates of the psychological distance between two stimuli, particularly when the two stimulus locations are very far apart in psychological space. Typically, the euclidean metric underlies judgments of integral stimuli, which are composed of dimensions that are perceived in a more holistic fashion, whereas the city-block metric underlies judgments of stimuli with readily separable dimensions (Shepard, 1964, 1987). The euclidean distance metric is used in all analyses reported in this article, as it consistently resulted in better fits than the city-block metric for our stimulus sets.

[3] An unconstrained two-dimensional WMDS solution also was estimated, in which both attention weights and stimulus coordinates were free parameters. Stimulus locations in the resulting configuration related strongly to the normative ratings for affect and exposure. In two multiple regression analyses predicting either the affect or the exposure norms from the coordinates of the two-dimensional constrained solution, the resulting multiple correlations were .55 for affect and .95 for exposure. The affect and exposure regression lines lay close to the diagonals of the two-dimensional stimulus space, however, and the resulting attention weights represented attention to affect and exposure only indirectly. Thus, adopting a confirmatory MDS approach allowed us to estimate attention to the two attributes of primary theoretical interest more precisely and independently.

[4] Evaluation of the adequacy of the MDS model fit necessitates generation and analysis of simulated random data for comparison. Following the procedure of MacCallum (1981), we simulated 10 sets of 74 matrices of random data; each matrix value was drawn from a uniform distribution in the interval (0, 1). The constrained model was fit to each of the 10 data sets, and s-stress, stress, and R^2 were recorded for each solution. The average s-stress (.61) and stress (.42) values for the constrained model were markedly worse than the values observed in the present study (.51 and .35, respectively). Similarly, the average R^2 for the constrained analyses of the random data (.05) was substantially lower than the observed R^2 in the present study (.35). This pattern of results suggests that both measurement error and attention to affect and exposure were marked in the present data set.

[5] Attention weights should be construed in a relative, rather than in an absolute, fashion, as their magnitude depends on several factors, including the physical salience of the dimensions, the range of variability within each dimension, and participants' differential attention to the dimensions. Presumably, for example, we could reverse the observed ordering of the average attention weights by greatly increasing variability in affect and decreasing variability in exposure in the stimulus set.

[6] This transformation is necessary, as each participant's pair of attention weights specifies the endpoint of a participant-specific *vector* in a two-dimensional (exposure by affect) "weight space," rather than a *point* in this weight space. Vector length in the weight space indicates the fit of the model to a participant's data, and vector direction corresponds to the participant's relative attention to exposure and affect. The angle of each participant's vector in the weight space is interpretable and can be compared meaningfully across participants, in contrast to the endpoint of the vector (i.e., the magnitude of the attention weights). FSW quantifies the angular, or directional, information present in the pair of attention weights in a single value that can be compared meaningfully across participants.

sure and affect were present, consideration of their utility for prediction of classification and learning seemed warranted.

Question 2: Do Individual Differences in Perceptual Organization Predict Individual Differences in Implicit Classification?

Data reduction. An SDT index of sensitivity, d', was computed for affect and exposure for each block in both learning tasks. For each block of 26 trials, four frequencies were tabulated: hits (*yes* responses when the woman was in the category of interest), misses (*no* responses when the woman was in the category of interest), false alarms (*yes* responses when the woman was not in the category of interest), and correct rejections (*no* responses when the woman was not in the category of interest). These four frequencies were tabulated twice: once by scoring each participant's answers as if exposure were the underlying characteristic and a second time as if affect were the underlying characteristic.

These scores then were used to compute a d' for exposure and a d' for affect for each participant. Because the direction of the category structure was arbitrary (e.g., *yes* could indicate positive affect or negative affect), high sensitivity to an attribute could be indicated either by a high sensitivity score (larger than the 0.0 expected for random responding) or by a low sensitivity score (below 0.0). Thus, the absolute value of d' indexed sensitivity in the implicit classification analyses. On this task, a high sensitivity score for either affect or exposure indicated that the participant was behaving as though affect or exposure were the underlying characteristic. High sensitivity for an attribute early in the task suggests that the participant's spontaneous category structure was based on that attribute. A high sensitivity score in later blocks might represent the development of an illusory correlation.

SDT bias indices were not analyzed for this task, as their values would be identical for the affect and exposure scorings of participants' responses. Bias scores reflect the base rate of *yes* responses on the implicit classification task, and an individual's base rate of *yes* responses would be the same for both the affect and the exposure scorings of participants' responses.

Data analysis. Sensitivity data from the implicit classification task were analyzed in a three-factor ANOVA; group (EO or AO) was a between-participants factor, and block and attribute were within-participant factors. There was a significant attribute effect, $F(1, 41) = 29.928$, $p < .001$, with participants showing greater sensitivity to exposure than to affect. As expected, there also was a significant Group $\times$ Attribute interaction, $F(1, 41) = 6.044$, $p < .05$, which is illustrated in Figure 2. In simple effects analyses, EO participants showed greater relative sensitivity to exposure, $t(43) = 2.402$, $p < .05$, and AO participants showed greater relative sensitivity to affect, $t(43) = -2.318$, $p < .05$. Effect sizes for these group differences were moderate to large in magnitude (.72 and .69 for exposure and affect, respectively, using Cohen's [1988] d and Hedges & Olkin's [1985] pooled *SD* estimates). There were no other main effects or interactions.[7] The absence of a Block effect suggests that EO and AO participants retained their initial set throughout the classification task. Inspection of the block-by-block sensitivities suggests that most EO and AO participants "tried" various alternative classification structures after finding that their initial structure was "incorrect" at the end of the first block (e.g., numerous AO participants tried exposure on the second block), but they tended to return to their initial set after their alternative strategies were refuted by the random feedback.

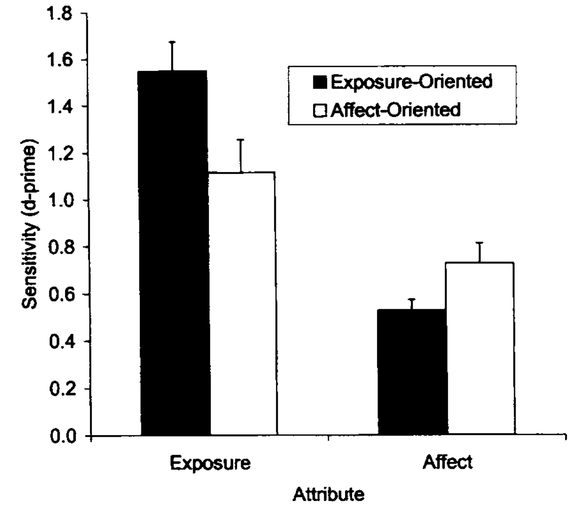

Figure 2. Exposure-oriented and affect-oriented participants' sensitivities to exposure and affect on implicit classification tasks.

Summary. Individual differences in relative attention to exposure and affect were associated with individual differences in sensitivity to affect and exposure on an implicit classification task. Participants showed relatively greater sensitivity when the attribute was congruent, rather than incongruent, with their perceptual organization.

Question 3: Do Individual Differences in Perceptual Organization Predict Individual Differences in Category Learning?

Data reduction. SDT indices of sensitivity and bias (d' and c) were computed for each block of the affect and exposure category-learning tasks. Because the direction of the category structure no longer was arbitrary, the raw values of d', rather than the absolute values of d', were used in further analyses.

Data analysis. Sensitivity data from the category-learning tasks were analyzed in a four-factor ANOVA; group (EO or AO) and ordering (affect first or exposure first) were between-participants factors, and block and task were within-participant factors. Seventy-eight percent of the participants obtained perfect scores on the fourth block of the exposure task. This ceiling effect created distributional and heterogeneity of variance problems for the planned ANOVA, so the fourth block was dropped from the overall analyses and is considered separately below.

[7] In addition to computing the sensitivity of each participant's responses to affect and exposure, we also calculated percentage congruent with affect and exposure scorings of participants' responses, as percentage congruent makes less stringent assumptions than the SDT process model. A parallel Group $\times$ Block $\times$ Attribute ANOVA on percentage congruent indices yielded similar, but slightly weaker, results. This suggests that bias influences on performance were present but minimal and that descriptive use of the SDT model was not inappropriate.

The pattern of mean differences in sensitivity for the first three blocks is illustrated in Figure 3. There was a significant Block effect, $F(2, 84) = 182.892$, $p < .001$, indicating that participants were learning to classify the stimuli correctly. A significant Task effect, $F(1, 42) = 18.465$, $p < .001$, reflected lower sensitivity to affect during the affect task than to exposure during the exposure task. Finally, there was a significant Group × Task interaction, $F(1, 42) = 4.195$, $p < .05$. Simple effects analyses showed a significant group difference for exposure, $t(44) = 1.895$, $p < .05$, and a nonsignificant tendency toward a group difference for affect, $t(44) = 1.313$, $p < .10$. Effect sizes for these two group comparisons were moderate for exposure and small for affect (.56 and .39, respectively, using Cohen's [1988] *d* and Hedges & Olkin's [1985] pooled *SD* estimates).

Although the fourth-block scores could not be included in the ANOVA, the pattern of results in the fourth block was consistent with our predictions. A significantly higher percentage of AO participants (59.1%) than EO participants (29.2%) obtained perfect scores on the fourth block of the affect-learning task, $z = -2.05$, $p < .05$. In the fourth block on the exposure-learning task, there was a nonsignificant tendency for a greater percentage of EO participants (87.5%) than AO participants (68.2%) to obtain perfect scores, $z = 1.59$, $p < .10$. The magnitude of the effect sizes for these group differences was moderate (.48 and .61 for exposure and affect, respectively, using Cohen's [1988] *h* for computation of the effect size for the difference in two proportions).

A parallel ANOVA of bias scores revealed a significant Block effect, $F(2, 84) = 10.208$, $p < .001$; the bias toward responding *yes* decreased from the first block ($M = -.14$) to the third block ($M = -02$). This analysis also yielded significant interactive effects for Block × Group × Ordering, $F(2, 84) = 3.399$, $p < .05$; Block × Task, $F(2, 84) = 4.460$, $p < .05$; and Block × Task × Group × Ordering, $F(2, 84) = 4.335$, $p < .05$. These complex interactions did not appear to reflect any interpretable pattern of mean differences, however, and mean bias scores under all conditions actually were very close to zero, reflecting low levels of bias in either direction.[8]

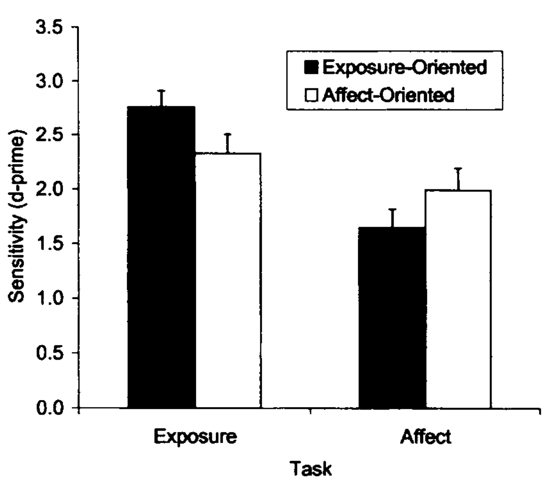

Figure 3. Exposure-oriented and affect-oriented participants' sensitivities to exposure and affect averaged across the first three blocks of exposure and affect category-learning tasks.

Summary. Individual differences in relative attention to exposure and affect predicted individual differences in performance on the exposure and affect category-learning tasks. Both AO and EO participants showed better performance across blocks when learning the exposure category structure, in keeping with the greater emphasis on exposure in both groups' perceptual organizations. More important, both groups also showed relatively better performance when learning the category structure that was more congruent with their perceptual organization.

Question 4: Does Fitting RASHNL to the Category-Learning Data Provide a Helpful Description of the Mechanisms Underlying Participants' Performance on the Category-Learning Tasks?

Overview of RASHNL. As shown in Figure 4, RASHNL is a three-layer connectionist model of category learning. It learns to map stimulus input to specific output categories by means of an exemplar layer, which represents the presented stimuli, or exemplars. In the present instantiation, two input-layer nodes encode stimulus values along the psychological dimensions of exposure and affect, 26 exemplar-layer nodes represent the 26 stimuli used in the category-learning tasks, and 2 category nodes correspond to participants' binary decisions. When a stimulus is presented to the network, the activation of the input nodes, weighted by dimension-specific attention strengths, propagates to the exemplar layer. Then exemplar activation, weighted by exemplar- and category-specific association weights, propagates to each of the 2 category nodes. Next, the two sets of weights connecting the three layers (i.e., attention strengths and association weights) are modified to minimize error (i.e., the discrepancy between the dichotomous feedback and the actual category node activations). The Appendix details activation propogation and the modification of the attention strengths and association weights.

Three parameters are of particular interest, as they capture the three learning mechanisms described in the introduction. First, the relative salience of exposure and affect is indicated by σ_{exp}, the salience of exposure. As the saliences of exposure and affect are constrained to sum to 1.0, this parameter captures the relative salience of these attributes. An estimate of relative salience differing from .5 is consistent with a Task effect, as the separation of the stimuli in the two categories will be greater along the more salient dimension. Second, the rate at which dimension-specific attention shifts is represented by $\lambda\gamma$, the attention-strength shift rate. Values greater than zero suggest that initial differences in the relative salience of affect and exposure are overcome during the course of learning, such that the learning curves for the two tasks converge. Thus, nonzero values of this parameter are consistent with a Group × Task × Block interaction. Third, the rate at which associations form between specific stimuli and correct category

[8] In addition to SDT sensitivity and bias indices, the percentages congruent with affect and exposure were computed for each block of the affect and exposure tasks. Parallel ANOVAs of percentage congruent showed similar results to those observed for sensitivity. Thus, as in the previous analyses of the implicit classification task data, the conclusions based on percentage congruent were similar to, but slightly weaker than, the conclusions based on sensitivity, reflecting the minimal influence of bias on participants' performances.

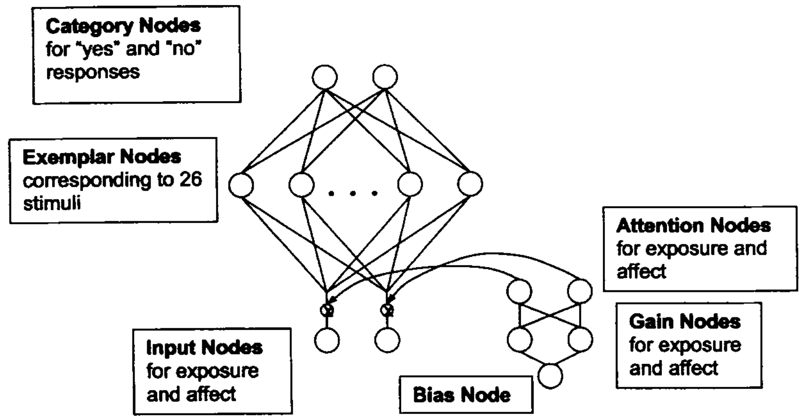

Figure 4. RASHNL model. Adapted from "A Model of Probabilistic Category Learning," by J. K. Kruschke and M. K. Johansen, 1999, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* p. 1095. Copyright 1999 by the American Psychological Association.

labels is indicated by λ_w, the association-weight learning rate. A parameter value of zero indicates that participants are not learning the category structure. Nonzero values indicate that participants are learning to associate regions of the affect-by-exposure psychological space with the correct category labels and are consistent with a Block effect.

Two other free parameters are estimated as well, although they are of less consequence to the present discussion. The probability mapping constant, ϕ, describes the decisiveness of participants' classifications; larger values indicate that small differences in the activation of the category nodes translate into large differences in category choice probabilities. The exemplar specificity parameter, c, captures the distinctiveness of the stimuli, with larger values indicating greater distinctiveness.

Modeling. The RASHNL model was fit to the 16 proportion-correct values (i.e., percentage correct divided by 100) obtained by the EO and AO groups on each of four blocks on the affect and exposure category-learning tasks; the four observed learning curves are illustrated in Figure 5. Note the Task × Group interaction, as well as the Task and Blocks effects, which emerged in the ANOVAs presented earlier. We evaluated the adequacy of several competing models both quantitatively, in terms of their fit statistics, and qualitatively, as indicated by their ability to reproduce the three observed ANOVA effects.

Following Kruschke and Johansen (1999), we fit RASHNL to the data using a stepwise search algorithm, which explores the parameter space iteratively and systematically, to find parameter values that minimized root-mean-squared deviation (RMSD). There were 16 degrees of freedom in the data: 2 (group) × 2 (task) × 4 (block). The order in which participants completed the tasks was not incorporated into the model, as interactions involving ordering were not significant in the ANOVAs. In contrast, we included block (i.e., we modeled performance for all four blocks, rather than averaging performance across four blocks), given the significant block effects in the ANOVAs.

We evaluated the quantitative and qualitative fit of four models. Model 1 included the five free parameters necessary to fit RASHNL and contained no group-specific parameters. Models 2, 3, and 4 estimated one of these parameters separately for EO and AO participants: exposure salience (Model 2), association-weight learning rate (Model 3), or attention-strength shift rate (Model 4). Note that Model 1 is nested within the remaining three models (i.e., Model 1 is a restricted version of Models 2 through 4). Figure 6 illustrates the predictions of Models 1 through 4. Table 1 presents fit indices and parameter estimates for each of these four models.

The likelihood-ratio statistic, G^2, is a measure of the discrepancy between the observed probabilities and the probabilities predicted by each model. The likelihood-ratio test can be used to compare the fit of two hierarchically nested models. The difference between G^2 values for two hierarchically related models is distributed as a χ^2 distribution, with degrees of freedom equal to the difference in the number of free parameters estimated by the two models (see Wickens, 1989, for a detailed overview of the use of G^2 for model comparison purposes). If the observed difference in G^2 values—the "likelihood-ratio test statistic"—exceeds the appropriate χ^2 critical value, then the restricted model is rejected, and the additional parameters in the more general model are presumed necessary.

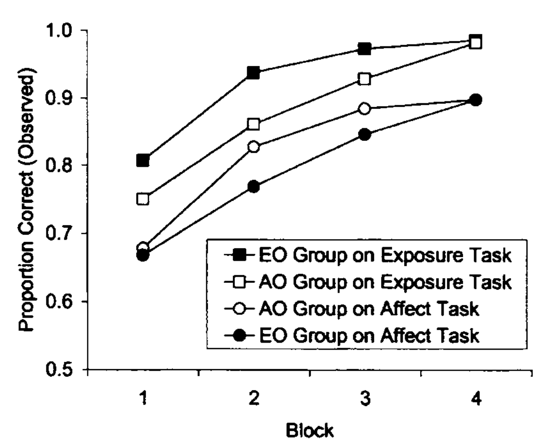

Figure 5. Observed percentage correct values for exposure-oriented (EO) and affect-oriented (AO) participants on exposure and affect category-learning tasks.

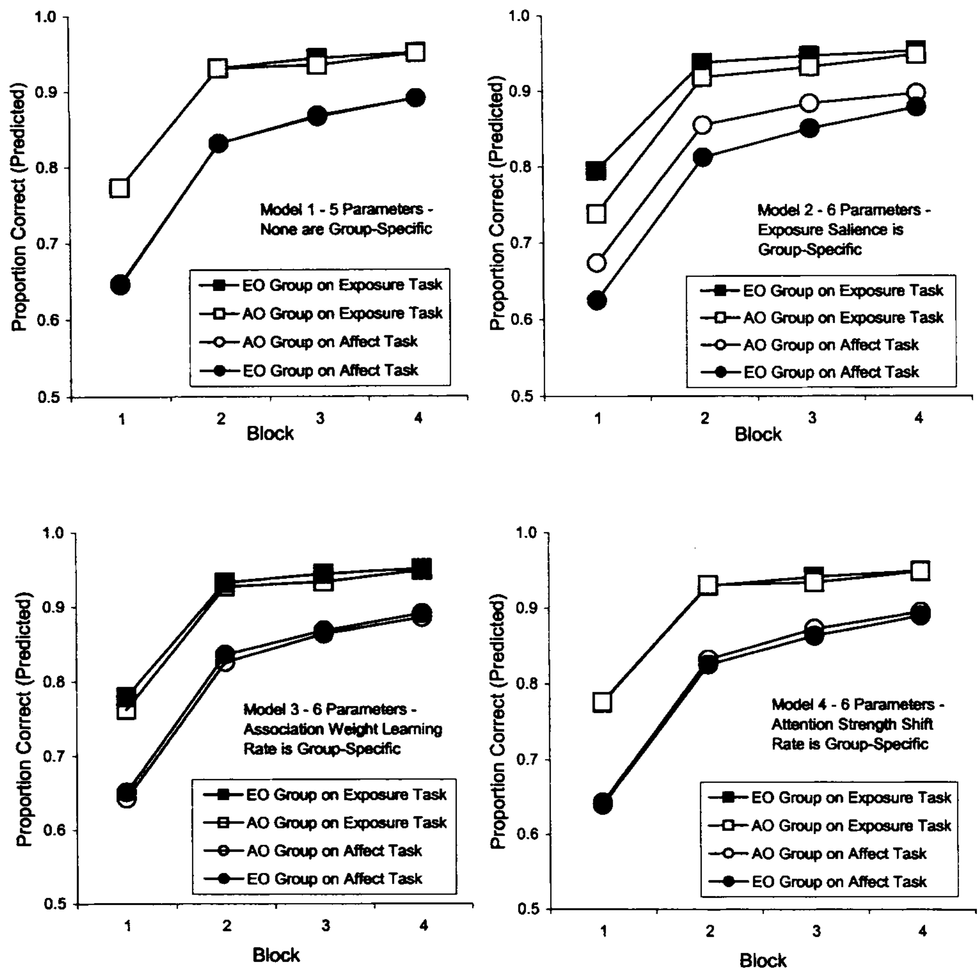

Figure 6. Predicted percentage correct values for exposure-oriented (EO) and affect-oriented (AO) participants on exposure and affect category-learning tasks for the four models listed in Table 1.

As Model 1 is nested within each of the remaining three models, it provides both quantitative and qualitative benchmarks by which to evaluate the incremental validity of the other models. The best fitting RMSD for Model 1 was .032, and the model accounted for 89.8% of the variability in the 16 data points. As this model contained no group-specific parameters, it could not predict the Group × Task interaction but rather predicted only the Task and Block effects.

Relative to these Model 1 benchmarks, Model 2, which estimated the relative salience parameter separately for EO and AO participants, clearly provided the best quantitative fit to the data. This model showed an RMDS of .027 and an r^2 of .927. Additionally, likelihood-ratio tests indicated that Model 2 fit significantly better than Model 1 $\chi^2(1, N = 16) = 26.91, p < .001$. Not surprisingly, Model 2 also provided the best qualitative account of the data. It predicted not only the observed Task and Block effects, but also the Task × Group interaction.

In contrast, neither Model 3 nor Model 4 showed a clear quantitative or qualitative advantage over Model 1, in spite of the addition of a group-specific parameter in each case. The fit indices of Model 3 and 4 barely improved on those of Model 1, and neither model could accommodate the Group × Task interaction in the observed data. This is not surprising, as neither model's structure allowed prediction of the observed Group × Task interaction, except by capitalizing on sample variability. Likelihood-ratio tests also demonstrated that the fits of Models 3 and 4 did not exceed that of Model 1, $\chi^2(1, N = 16) = 2.76$ and -0.97, respectively.[9] Thus, the model including the group-specific relative salience parameter (Model 2) was privileged in its ability to predict the Group × Task interaction; both the quantitative and qualitative fits of this model were superior to those of the competing six-parameter models (Models 3 and 4).

The parameter estimates of the best fitting model, Model 2, were readily interpretable. As expected, EO participants showed a larger exposure salience estimate (.614) than AO participants (.445), allowing the model to predict the Task and Group × Task effects.

[9] The negative value of the likelihood-ratio test statistic that is obtained when comparing Models 1 and 4 is a by-product of optimizing RMSD, rather than log likelihood, and is not cause for concern, given its small magnitude.

Table 1
RASHNL Modeling Results

Model (no. parameters)	Group-specific parameter	RMSD	σ_{exp}	λ_w	λ_γ	c	ϕ	r^2	G^2
1 (5)	None	.032	.554	.100	.000	1.976	5.079	.898	117.97
2 (6)	Exposure salience	.027	.614[a] .445[b]	.092	.000	1.847	5.241	.927	91.06
3 (6)	Association-weight learning rate	.031	.544	.103[a] .089[b]	.000	1.911	5.145	.899	115.21
4 (6)	Attention-strength shift rate	.031	.593	.116	.000[a] .042[b]	0.169	4.780	.899	118.94

Note. RMSD = root mean squared deviation; σ_{exp} = salience of exposure; λ_w = association-weight learning rate; λ_γ = attention-strength shift rate; c = specificity of exemplar nodes; ϕ = probability mapping constant.
[a] Exposure-oriented group estimate. [b] Affect-oriented group estimate.

The association-weight learning rate estimate was .092, suggesting that participants were learning to associate regions of the psychological space with the correct category label; the nonzero parameter value allowed the model to predict the observed Block effect. It is interesting to note that the attention-strength shift rate estimate was .000, suggesting that participants were not shifting attention to different attributes over the course of learning. The attention-strength shift rate value of zero is consistent theoretically with the finding that the euclidean distance metric accounts better for participants' responses than the city-block distance metric. Typically, the euclidean metric provides a better fit when stimulus dimensions are perceived in a more holistic fashion, in which case change in relative attention to the dimensions is more difficult (Nosofsky & Palmeri, 1996; Shepard, 1964). This zero-value estimate also is consistent with the absence of a Group × Task × Block interaction in the ANOVA results reported earlier.

Summary. RASHNL fit the category-learning data well, with the addition of a group-specific parameter representing differences between EO and AO participants' relative attention to exposure and affect, and suggested that participants in neither group altered their relative attention to affect and exposure stimulus information during the learning tasks.

Question 5: Are Individual Differences in Perceptual Organization Associated With Responses on a Social Perception Measure Relevant to Sexually Coercive Behavior?

Data reduction. The total justification score (from the HPS), the relative justification score (from the HPS), and the total coercion score (from the CSS) were computed for each participant, as described in the Method section. To correct for positive skew, we applied a square root transformation to all three distributions. All analyses were conducted on the transformed data, but summary information presented in the text and figures is based on the untransformed data.

Data analysis: Heterosocial Perception Survey. The HPS total justification score did not differ significantly for EO (M = 207.13) and AO (M = 254.17) participants, and the effect size was small to moderate (.39; Cohen's [1988] d and Hedges & Olkin's [1985] pooled SD estimates). However, the AO group showed significantly higher HPS relative justification scores (M = 27.29) than the EO group (M = 17.08), $t(46) = -2.417$, $p < .05$; the effect size for this group difference was moderate to large (.70). For AO participants, the perceived justifiability of unwanted sexual advances was more dependent on the degree of negative reaction from the hypothetical woman than it was for EO participants. As the AO and EO classifications reflected individual differences in relative attention to exposure and affect, the observed group differences on the HPS relative justification score could be related to (a) group differences in relative attention to exposure and affect, (b) absolute levels of attention to exposure, or (c) absolute levels of attention to affect. To explore these possibilities, we correlated the HPS relative justification scores for AO and EO participants with their FSWs and with their attention weights for exposure and affect. The HPS relative justification score showed moderate correlations with attention toward affect, $r(46) = .418$, $p < .001$, and with FSW, $r(46) = -.402$, $p < .01$, and a weaker correlation with attention toward exposure, $r(46) = -.286$, $p < .05$. Thus, those participants showing greater attention to affect on the similarity-ratings task demonstrated greater sensitivity to the negativity of a woman's affect on the HPS.

Data analysis: Coercive Sexuality Scale. AO and EO groups did not differ significantly on the CSS total coercion score, which was the total number of sexually coercive acts a participant reported. AO men endorsed an average of 1.58 acts, whereas EO men reported an average of 1.96 acts; the associated effect size was small (.32; Cohen's [1988] d and Hedges & Olkin's [1985] pooled SD estimates). Of the AO and EO participants, 70.8% reported engaging in at least one sexually coercive behavior, which is comparable to the percentage of college males who reported engaging in at least one sexually coercive behavior on a similar measure in studies by Rapaport and Burkhart (61.0%; 1984), by Rapaport and Posey (64.0%; 1991), and by Hannan and Burkhart (55.0%; 1993).

Convergence of HPS and CSS. The CSS total coercion score of AO and EO participants correlated moderately with the HPS total justification score, $r(46) = .417$, $p < .01$, but was unassociated with the HPS relative justification score, $r(46) = -.099$, *ns*. The HPS total and relative justification scores showed moderate convergence, $r(46) = .355$, $p < .05$.

Summary. No significant group differences emerged on the CSS total coercion score or the HPS total justification score. AO and EO participants differed significantly on the HPS relative justification score, however: EO participants were less sensitive to the negativity of a woman's response to an unwanted sexual advance. The CSS total coercion score correlated moderately with the HPS total justification score but not with the HPS relative justification score.

Discussion

Clinical scientists' theorizing in the area of sexual violence suggests that episodes of sexually coercive behavior between acquaintances are associated with deficits in decoding processes (Johnston & Ward, 1996; McFall, 1990; Schewe & O'Donohue, 1993; Segal & Stermac, 1990; Ward et al., 1997). Prior research suggests possible roles for insensitivity to and a liberal bias in perception of women's sexual interest in the instigation of sexual aggression. Social information-processing theorists also implicate attentional deficits, such as decreased attention to a woman's affect and increased attention to her physical exposure. Investigation of this hypothesis has been limited, in part, by a dearth of appropriate measurement models in clinical psychology. Fortunately, cognitive scientists' theoretical models of perceptual organization are highly relevant and are associated with well-developed measurement models. Additionally, cognitive scientists' models not only address perceptual organization but also delineate other related higher order cognitive processes (e.g., memory, learning, and classification) and their interrelationships. Thus, the present study examined the feasibility and utility of adapting theoretical and measurement models drawn from cognitive science to study the role of cognitive processing in a specific clinical problem, sexually coercive behavior. More specifically, we used the WMDS model to describe individual differences in men's perceptual processing of stimuli depicting women and explored the linkage between participants' perceptual processes and self-reported sexually coercive behavior and perceptions. Cognitive scientists' theoretical models suggest that perceptual organization constrains the operation of higher order cognitive processes. Thus, we also used SDT and RASHNL to evaluate how individual differences in construal affected later classification and learning.

Individual differences in initial perceptual processing and organization appeared to exert an important influence on performance on later classification and learning tasks in the same stimulus domain, even with complex social stimuli. Participants' sensitivities to exposure and affect on the classification task were congruent with their underlying perceptual organization: EO participants were more sensitive to exposure than AO participants; AO participants were more sensitive to affect than EO participants. On the category-learning tasks, AO participants performed better than EO participants on the affect-learning task; EO participants performed better than AO participants on the exposure-learning task. In other words, participants whose perceptual organization was more congruent with the underlying category structure showed better performance than those whose perceptual organization was less congruent. One implication of this finding is that men who focus relatively more on women's sexual characteristics, rather than on their affective cues, may find it more difficult to learn competent responding when they must rely on affective cues for feedback. Interestingly, both AO and EO participants showed greater sensitivity to exposure on the classification task and learned the exposure category structure more quickly. These results are consistent with the greater average attention weight for exposure in the WMDS analyses. Overall, the results indicate that MDS representations of perceptual organization provide useful and valid descriptions of individual differences in perception of social stimuli and that perceptual organization has an important effect on classification and learning.

The RASHNL modeling allowed us to draw suggestive inferences about the mechanisms underlying participants' learning, facilitated interpretation of the ANOVA results in terms of cognitive science concepts, and suggested numerous research questions for future investigation. The best fitting model suggested that the perceived relative salience of exposure and affect differed for EO and AO participants. This difference allowed each group to solve the congruent category structure more quickly than the other group. This model also estimated the rate at which participants shifted attention from one dimension to another to be zero. This, in turn, suggested that EO and AO participants were not altering their attention to the exposure and affect dimensions to maximize the distance between stimuli in different categories and to minimize the distance between stimuli in the same category.

Overall, these parameter estimates suggest that EO and AO participants' perceptual organizations, as indexed by their relative attention to exposure and affect, were quite stable and unperturbed by the feedback received during the task. Thus, altering participants' perceptions may necessitate more than providing corrective feedback when participants learn an unspecified category structure. Under what conditions might participants' perceptual organizations show varying degrees of "stability"? For example, might participants' relative attention to exposure and affect be more malleable if an explicit description of the category structure were provided, or if new women were presented throughout the task? Eventually, individual differences in the stability of participants' perceptual organizations also will be of interest, as variability in the responsiveness of participants' construal processes to changing circumstances may be a useful predictor of participants' adaptation to and behavior in these new circumstances. For example, might EO participants who show less flexible perceptual organizations be at greater risk of exhibiting behavior judged to be coercive than EO participants who shift their attention to women's affective cues more readily?

Individual differences in perceptual organization were unrelated to a self-reported history of sexually coercive behavior on the CSS in the present study. This was not a surprising outcome in the current theoretical context, in which it would be paradoxical to expect self-reported coercion to be related to perceptual differences. That is, participants with deficits in their perceptions of heterosexual cues should be less likely to recognize and report accurately when their own behavior is coercive. Future evaluation of the relationship between individual differences in perceptual structure and sexual coercion would benefit from a comparison of the cognitive processing patterns of externally identified, rather than self-identified, coercive men and relevant controls.

Interestingly, AO and EO participants differed significantly on a *relative* justifiability measure based on the HPS. Compared with the EO participants, AO participants decreased their ratings of justifiability more as the female target's reaction became more negative. In this hypothetical scenario, at least, EO men were less sensitive to differences in the magnitude of a woman's negative behavioral cues. To evaluate this finding further, we split both the AO and EO groups into even numbers of more and less extremely AO- and EO-oriented participants ($n = 12$ for each subgroup). Figure 7 presents the mean HPS relative justification score for these four groups of participants. A one-way ANOVA showed the expected Group effect, $F(3, 44) = 3.601$, $p < .05$. A post hoc Dunnett test showed that the extreme EO group differed significantly from each of the other groups (all $p < .05$). Thus, as would be expected, being extremely EO appears to place a participant at particular risk of decreased sensitivity to women's sexual interest in another context.

Cognitive scientists typically use formal process models, such as WMDS, RASHNL, and SDT, under more ideal circumstances in which model assumptions are less likely to be violated. Thus, our results and conclusions should be viewed cautiously. It is possible, for example, that we might draw different conclusions if we were to use different measurement models to evaluate perceptual organization, classification, and category learning. The strength and qualitative consistency of the findings with our theoretical expectations mitigate against this possibility. Nonetheless, replication and extension of our inferences are critical to further evaluation of their validity.

Future research should incorporate current knowledge about the context-specificity of sexually coercive behavior, as the likelihood of sexual violence varies strikingly with numerous situational factors (Abbey et al., 1996; Marx, Van Wie, & Gross, 1996; Muehlenhard & Linton, 1987). For example, alcohol consumption is estimated to occur prior to or during the overwhelming majority of coercive sexual encounters. Thus, we would anticipate that theoretically relevant manipulations, such as alcohol consumption or inducement of sexual arousal, should (a) alter many participants' cognitive processing patterns by exerting main and interactive effects on attention, memory, learning, and other higher order cognitive processes and (b) increase the magnitude of the associations between cognitive processing and indices relevant to sexual violence, such as the HPS relative justification score. Eventually, this line of basic research on the linkage between cognitive processing and sexual violence should lead to the development of performance-based assessment instruments that would allow clinicians to characterize a patient's cognitive processing deficits, identify patient-specific treatment targets, and evaluate patient-specific treatment progress.

Modification and enhancement of the measurement and methodological strategies used in the present study also should be fundamental to future research efforts. It will be important to assess the generalizability of the present results when using a stimulus set containing photographs of undergraduate women, rather than models depicted in newsstand magazines. Future efforts also should examine the potential incremental utility of alternative performance-based strategies for assessing perceptual organization, classification, and learning, such as speeded assessments of perceptual organization, explicit classification tasks, and category-learning tasks involving more complex category structures and probabilistic feedback. Finally, further development and investigation of the HPS could tease apart two simultaneous influences on participants' justifiability ratings: the negativity of the woman's responses and the intimacy of the man's sexual advance.

The present study demonstrates the feasibility and utility of adapting the theoretical and measurement models developed by cognitive scientists to study the role of cognitive processing in clinical problems. Even though these models were developed to examine normative processing of nonsocial stimuli, they clearly generalize to less ideal situations and can be used to investigate individual differences in processing socially relevant stimuli. The results demonstrated that the close association among perceptual organization, classification, and category learning observed in basic research (Ashby, 1992; Nosofsky, 1992a, 1992b) also is reflected in an association between individual differences in perceptual organization and individual differences in classification and category learning with socially relevant stimuli. Thus, the theoretical and measurement models of normative cognitive processes appear to provide promising strategies for examining further the role of cognition in psychopathology. In a reciprocal fashion, continued examination of the role of cognitive processing in psychopathology provides an opportunity for cognitive scientists to develop unified accounts of cognitive processing that capture not only normative perception of simple, artificial stimuli but also systematic individual differences in complex social perception.

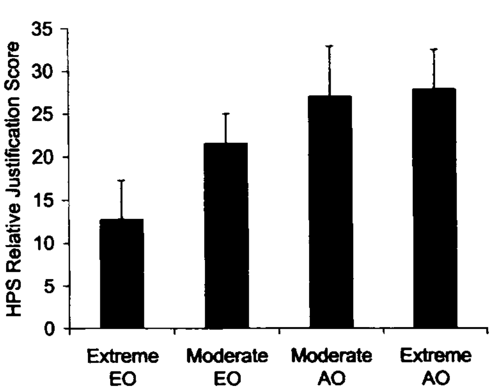

Figure 7. Average Heterosocial Perception Survey (HPS) relative justification scores for extreme exposure-oriented (EO), moderate EO, moderate affect-oriented (AO), and extreme AO participants.

References

Abbey, A. (1982). Sex differences in attributions for friendly behavior: Do males misperceive females' friendliness? *Journal of Personality and Social Psychology, 42,* 830–838.

Abbey, A., Cozzarelli, C., McLaughlin, K., & Harnish, R. J. (1987). The effects of clothing and dyad sex composition on perceptions of sexual intent: Do women and men evaluate these cues differently? *Journal of Applied Social Psychology, 17,* 108–126.

Abbey, A., & Harnish, R. J. (1995). Perception of sexual intent: The role of gender, alcohol consumption, and rape supportive attitudes. *Sex Roles, 32,* 297–313.

Abbey, A., McAuslan, P., & Ross, L. T. (1998). Sexual assault perpetration by college men: The role of alcohol, misperception of sexual intent, and sexual beliefs and experiences. *Journal of Social and Clinical Psychology, 17,* 167–195.

Abbey, A., Ross, L. T., McDuffie, D., & McAuslan, P. (1996). Alcohol and dating risk factors for sexual assault among college women. *Psychology of Women Quarterly, 20,* 147–169.

Abbey, A., Zawacki, T., & McAuslan, P. (2000). Alcohol's effects on sexual perception. *Journal of Studies on Alcohol, 61,* 688–697.

Ashby, F. G. (1992). *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Berkowitz, A. (1992). College men as perpetrators of acquaintance rape and sexual assault: A review of recent research. *Journal of American College Health, 40,* 175–181.

Bondurant, B., & Donat, P. L. N. (1999). Perceptions of women's sexual interest and acquaintance rape: The role of sexual overperception and affective attitudes. *Psychology of Women Quarterly, 23,* 691–705.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Craig, M. E. (1990). Coercive sexuality in dating relationships: A situational model. *Clinical Psychology Review, 10,* 395–423.

Davison, M. L. (1992). *Multidimensional scaling.* Malabar, FL: Krieger.

Drieschner, K., & Lange, A. (1999). A review of cognitive factors in the etiology of rape: Theories, empirical studies, and implications. *Clinical Psychology Review, 19,* 57–77.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hall, G. C. N., & Hirschman, R. (1994). The relationship between men's sexual aggression inside and outside the laboratory. *Journal of Consulting and Clinical Psychology, 62,* 375–380.

Hannan, K. E., & Burkhart, B. (1993). The topography of violence in college men: Frequency and comorbidity of sexual and physical aggression. *Journal of College Student Psychotherapy, 8,* 219–237.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

Hudson, S. M., & Ward, T. (2000). Interpersonal competency in sex offenders. *Behavior Modification, 24,* 494–527.

Johnston, L., & Ward, T. (1996). Social cognition and sexual offending: A theoretical framework. *Sexual Abuse: Journal of Research and Treatment, 8,* 55–80.

Jones, L. E. (1983). Multidimensional models of social perception, cognition, and behavior. *Applied Psychological Measurement, 7,* 451–472.

Koss, M. P., Gidycz, C. A., & Wisniewski, N. (1987). The scope of rape: Incidence and prevalence of sexual aggression and victimization in a national sample of students in higher education. *Journal of Consulting and Clinical Psychology, 55,* 162–170.

Koss, M. P., & Oros, C. J. (1982). Sexual Experiences Survey: A research instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology, 50,* 455–457.

Kowalski, R. M. (1993). Inferring sexual interests from behavioral cues: Effects of gender and sexually relevant attitudes. *Sex Roles, 29,* 13–36.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1083–1119.

Lang, P. J. (1988). Fear, anxiety, and panic: Context, cognition, and visceral arousal. In S. Rachman & J. D. Maser (Eds.), *Panic: Psychological perspectives* (pp. 219–236). Hillsdale, NJ: Erlbaum.

Lipton, D. N., McDonel, E. C., & McFall, R. M. (1987). Heterosocial perception in rapists. *Journal of Consulting and Clinical Psychology, 55,* 17–21.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

MacCallum, R. C. (1977). Effects of conditionality on INDSCAL and ALSCAL weights. *Psychometrika, 42,* 297–305.

MacCallum, R. (1981). Evaluating goodness of fit in nonmetric multidimensional scaling by ALSCAL. *Applied Psychological Measurement, 5,* 377–382.

MacLeod, C. (1993). Cognition in clinical psychology: Measures, methods or models? *Behaviour Change, 10,* 169–195.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

Malamuth, N. M., & Brown, L. M. (1994). Sexually aggressive men's perceptions of women's communications: Testing three explanations. *Journal of Personality and Social Psychology, 67,* 699–712.

Malamuth, N. M., Sockloskie, R. J., Koss, M. P., & Tanaka, J. S. (1991). Characteristics of aggressors against women: Testing a model using a national sample of college students. *Journal of Consulting and Clinical Psychology, 59,* 670–681.

Marx, B. P., Van Wie, V., & Gross, A. M. (1996). Date rape risk factors: A review and methodological critique of the literature. *Aggression and Violent Behavior, 1,* 27–45.

McDonel, E. C. (1995). An information-processing theory of the measurement of social competence. In W. O'Donohue & L. Krasner (Eds.), *Theories of behavior therapy: Exploring behavior change* (pp. 415–443). Washington, DC: American Psychological Association.

McDonel, E. C., & McFall, R. M. (1991). Construct validity of two heterosocial perception skill measures for assessing rape proclivity. *Violence and Victims, 6,* 17–30.

McFall, R. M. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment, 4,* 1–33.

McFall, R. M. (1990). The enhancement of social skills: An information-processing analysis. In W. L. Marshall & D. R. Laws (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 311–330). New York: Plenum Press.

McFall, R. M., & Townsend, J. T. (1998). Foundations of psychological assessment: Implications for cognitive assessment in clinical science. *Psychological Assessment, 10,* 316–330.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50,* 215–241.

McFall, R. M., Treat, T. A., & Viken, R. J. (1997). Contributions of cognitive theory to new behavioral treatments. *Psychological Science, 8,* 174–176.

McFall, R. M., Treat, T. A., & Viken, R. J. (1998). Contemporary cognitive approaches to studying clinical problems. In D. K. Routh & R. J. DeRubeis (Eds.), *The science of clinical psychology: Accomplishments and future directions* (pp. 163–197). Washington, DC: American Psychological Association.

Muehlenhard, C. L., & Linton, M. A. (1987). Date rape and sexual aggression in dating situations: Incidence and risk factors. *Journal of Counseling Psychology, 34,* 186–196.

Murphy, W. D., Coleman, E. M., & Haynes, M. R. (1986). Factors related to coercive sexual behavior in a nonclinical sample of males. *Violence and Victims, 1,* 255–278.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M. (1992a). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363–393). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. (1992b). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43,* 25–53.

Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin and Review, 3,* 222–226.

Rapaport, K., & Burkhart, B. R. (1984). Personality and attitudinal characteristics of sexually coercive college males. *Journal of Abnormal Psychology, 93,* 216–221.

Rapaport, K. R., & Posey, C. D. (1991). Sexually coercive college males. In A. Parrot & L. Bechhofer (Eds.), *Acquaintance rape: The hidden crime* (pp. 217–228). New York: Wiley.

Schewe, P., & O'Donohue, W. (1993). Rape prevention: Methodological problems and new directions. *Clinical Psychology Review, 13,* 667–682.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications.* Orlando, FL: Academic Press.

Segal, Z. V., & Stermac, L. E. (1990). The role of cognitions in sexual assault. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories and treatment of the offender* (pp. 161–174). New York: Plenum.

Shea, M. E. C. (1993). The effects of selective evaluation on the perception of female cues in sexually coercive and noncoercive males. *Archives of Sexual Behavior, 22,* 415–433.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1,* 54–87.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Shotland, R. L., & Craig, J. M. (1988). Can men and women differentiate between friendly and sexually interested behavior? *Social Psychology Quarterly, 51,* 66–73.

Treat, T. A., McFall, R. M., Viken, R. J., Nosofsky, R. M., MacKay, D. B., & Kruschke, J. K. (in press). Assessing clinically relevant perceptual organization with multidimensional scaling techniques. *Psychological Assessment.*

Ward, T., Hudson, S. M., Johnston, L., & Marshall, W. L. (1997). Cognitive distortions in sex offenders: An integrative review. *Clinical Psychology Review, 17,* 479–507.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale, NJ: Erlbaum.

Wilson, T. D. (1994). The proper protocol: Validity and completeness of verbal reports. *Psychological Science, 5,* 249–252.

Yeater, E. A., & O'Donohue, W. (1999). Sexual assault prevention programs: Current issues, future directions, and the potential efficacy of interventions with women. *Clinical Psychology Review, 19,* 739–771.

Young, F. W., & Lewyckyj, R. (1996). *ALSCAL user's guide* (5th ed.). Chapel Hill: L. L. Thurstone Psychometric Laboratory, University of North Carolina.

Appendix

Activation Propagation and Modification of Attention Strengths and Association Weights Within the RASHNL Model

Activation Propagation

The activation of each of the two input nodes i corresponds to $a_i^{\text{in}} = \psi_i$, where ψ_i is the average normative rating for the stimulus on dimension i (either exposure or affect) and the superscript "in" indicates that this is an input node.

Activation of each of the 26 exemplar nodes corresponds to the psychological similarity of the input stimulus to the exemplar node. More similar exemplar nodes are activated more strongly, as the input stimulus is more likely to fall in the exemplar node's "receptive field." Similarity is an exponentially decreasing function of the weighted euclidean distance between the input stimulus and the exemplar node in psychological space. Activation of each exemplar node is defined as

$$a_j^{\text{ex}} = exp\left\{ -c\left[\sum_{i}^{\text{in}} \alpha_i\,(\sigma_i|\psi_{ji} - a_i^{\text{in}}|)^2\right]^{\frac{1}{2}}\right\},$$

where ψ_{ji} is the average normative rating for stimulus j on dimension i and the superscript "ex" indicates that this is an exemplar node. The constant c, which is referred to as the *specificity* of the exemplar nodes, determines the width of each exemplar's receptive field. Large values of c correspond to a very small (i.e., highly specific) receptive field, such that exemplars are activated only by highly similar input stimuli. The *salience* of the ith dimension, σ_i, is assumed to be stable across the experiment and refers to how "stretched out" each psychological dimension is. The salience of exposure and affect sum to 1.0, so σ_i essentially captures the relative salience of exposure and affect. The *attention strength* on the ith dimension, α_i, also captures how stretched out the exposure or affect dimension is. In contrast to dimensional saliences, however, dimensional attention strengths can change over the course of the experiment, as a dimension becomes more or less relevant to the current learning task.

Each attention strength, α_i, is a function of an underlying dimensional *gain*, γ_i, as given by

$$\alpha_i = \exp\left(\gamma_i \Big/ \sum_{j}^{\text{in}} \exp(\gamma_j)\right).$$

This formalization forces the attentional strengths to be nonnegative and capacity constrained, such that the two attention strengths sum to unity. The two gains are initialized at zero (i.e., the two attention strengths are initialized at .50) but can become any real value.

Activation of each of the two category nodes corresponds to the sum of weighted exemplar-node activations:

$$a_k^{\text{cat}} = \sum_{j}^{\text{ex}} w_{kj}^{\text{cat}} a_j^{\text{ex}},$$

where the superscript "cat" indicates that this is a category node and w_{kj}^{cat} is the association weight from exemplar node j to category node k. Initial values for the association weights are zero, reflecting the lack of association between the stimuli and specific outcomes before the learning task begins.

Response probabilities are obtained from category node activations through Luce's (1959) choice rule, such that

$$\Pr(K) = \exp\left(\phi a_K^{\text{cat}} \Big/ \sum^{\text{cat}} \exp(\phi a_K^{\text{cat}})\right),$$

where K corresponds to a particular response, and ϕ is the probability mapping constant. Large values of ϕ map a small difference in the activations of the two-category node into a large difference in the response probabilities for the two categories. Small values of ϕ render less extreme response probabilities, even when large differences in category node activations are present.

Attention Shifting and Association Learning

Standard backpropagation is used to change attention strengths and association weights by gradient descent on sum-squared error. On each trial, the error generated by the model is given by

$$E = \frac{1}{2}\sum_{k}^{\text{cat}} (t_k - a_k^{\text{cat}})^2,$$

where t_k indicates the *teacher value*, or feedback, given to each category node. The value of t_k is 1 if the stimulus is a member of category k and 0 otherwise.

On each trial, learning progresses in two serial stages: Attention shifts rapidly to minimize error and association weights change to reduce remaining error. Recall that dimensional attention strengths, α_i, are a function of dimensional attention gains, γ_i. Thus, the attention strength for dimension A is modified indirectly, by shifting the dimensional gain for A proportionally to the negative of the error gradient with respect to the dimensional gains:

$$\Delta\gamma_A = -\lambda_\gamma \sum_{i}^{\text{in}} \sum_{j}^{\text{ex}} \sum_{k}^{\text{cat}} (t_k - a_k^{\text{cat}})\, w_{kj}^{\text{cat}} a_j^{\text{ex}} c\sigma_i |\psi_{ji} - a_i^{\text{in}}|\, (\kappa_{iA}\alpha_A - \alpha_i\alpha_A),$$

where $-\lambda_\gamma$ is the *attention-strength shift rate*, and $\kappa_{iA} = 1$ if $i = A$ and 0 otherwise. In contrast to ALCOVE, trial-by-trial attention shifts in RASHNL can be quite large. Large attention shifts are accomplished by iterating this equation 10 times on each trial and recalculating error after each iteration.

Association weights are modified after attention is shifted, again using gradient descent on error:

$$\Delta w_{kj}^{\text{cat}} = -\lambda_w (t_k - a_k^{\text{cat}}) a_j^{\text{ex}},$$

where λ_w is the *association-weight learning rate*. The associative weights from the bias node to the gain nodes are also adjusted by means of gradient descent on error, where error is defined as the difference between the shifted value and the initial, preshift value. That is, the shifted value acts as the teacher, or target, for the gain nodes.

Interested readers who would like to learn more about RASHNL are referred to Kruschke and Johansen's (1999) article, which provides a more detailed description and evaluation of RASHNL.

Received November 30, 2000
Revision received June 18, 2001
Accepted June 25, 2001 ■