

Dissimulation in Phallometric Testing of Rapists' Sexual Preferences

Grant T. Harris, Ph.D.,^{1,3} Marnie E. Rice, Ph.D.,¹ Terry C. Chaplin, B.A.,¹ and Vernon L. Quinsey, Ph.D.²

Sexual preferences of 38 rapists were assessed phallometrically with and without a semantic tracking task in a counterbalanced design. Four categories of audio-taped vignettes describing neutral interactions, consenting sex, rape, and nonsexual violence were employed as stimuli. In the semantic tracking task, participants were instructed to press one button when violent events were described in the vignette and another when sexual activities were described. Phallometric assessment with the semantic task better discriminated between rapists and non-sex-offender participants (from an earlier study) than the same assessment without the task. Among four rapists who had previous experience with phallometric testing, there was a very strong correlation between deviance scores and tracking accuracy. Results suggest that the semantic task may improve discriminant validity, particularly among sex offenders who have had previous experience with phallometric assessment.

KEY WORDS: phallometric assessment; faking; rape; sex offenders.

INTRODUCTION

Phallometric assessments are commonly employed to determine the sexual preferences of sex offenders. Variables affecting the discriminant validity of phallometric assessments have received extensive study. Harris *et al.* (1992) found that discriminant validity was enhanced by the use of z (rather than raw) scores to compensate for individual differences in responsivity, deviance indices (computed

¹Mental Health Centre, 500 Church Street, Penetanguishene, Ontario, L9M 1G3, Canada.

²Queen's University, Kingston, Ontario, Canada.

³To whom correspondence should be addressed.

by calculating the mean difference between responses to deviant and nondeviant stimuli) rather than responses to deviant stimuli alone, and stimuli depicting brutal sexual coercion.

Depending upon the use of procedures found to enhance it (Harris *et al.*, 1992), studies have obtained varying amounts of discrimination between rapists and other men (Lalumière and Quinsey, 1993, 1994). The ability of sex offenders to fake their preferences could undermine discriminant validity. Since Freund's initial investigations (Freund, 1961), there have been several demonstrations of instructional control of penile tumescence (Adams *et al.*, 1992; Henson and Rubin, 1971; Lalumière and Earls, 1992; Laws and Rubin, 1969; Quinsey and Bergersen, 1976; Rosen *et al.*, 1975; Rubin and Henson, 1975). Researchers have attempted to discern signs that participants are attempting to fake (Freund *et al.*, 1988; Smith and Over, 1987), and to eliminate it by employing very brief stimuli, different modalities of stimulus presentation, and "priming" in which a stimulus that is preferred (or said to be preferred) is presented immediately before a deviant stimulus. None of these have been completely successful in eliminating faking by motivated subjects (Abel *et al.*, 1975; Julien and Over, 1988; Malcolm *et al.*, 1985; Quinsey and Carrigan, 1978; Wormith *et al.*, 1988; Wydra *et al.*, 1983).

We developed a semantic tracking task to reduce faking in phallometric assessments. Participants were instructed to press one button when sexual activity was described and another when violence was described (Quinsey and Chaplin, 1988). This was to force subjects to attend to relevant features of the stimuli without interfering with erectile responses. Using stimuli that had been shown to discriminate rapists from nonrapists (Quinsey *et al.*, 1981), 16 non-sex offenders were tested four times. In the first and fourth sessions, participants received normal instructions (relax, listen carefully to the stories, and pretend that you are the person talking). In the second session, subjects received faking instructions (pretend you're sexually interested in rape and nonsexual violence but not in consenting sex). In the third session, the semantic tracking task (press one button for sex and the other for violence) was added to faking instructions. For a random half of the subjects, the tracking task alone was presented in the second session, faking instructions alone in the third, and both tasks in the fourth session. Participants were told various strategies to try to fake interest in nonarousing stimuli and lack of interest in arousing stimuli (e.g., think about something else), but were free to use any strategy. Subjects who successfully faked when instructed to do so were paid more. Fifteen of the 16 subjects significantly faked their responses in the absence of the tracking task, but only 2 succeeded when also performing the semantic task. Both reported they memorized the stimuli in earlier sessions so they would not have to attend closely when faking.

Although the semantic task reduced faking in non-sex offenders, its usefulness with rapists remained unclear. This issue is somewhat difficult to address because, when the procedures recommended in Harris *et al.* (1992) are used, phallometric assessments already achieve excellent discrimination. To improve classification

accuracy, therefore, the semantic task would have to make the responses of rapists even more deviant. In the present study, we tested rapists with and without the semantic tracking task using a counterbalanced, within-subjects design.

METHOD

Participants

Thirty-eight male rapists (\bar{x} age = 31.9, SD = 7.14 years) participated as part of a clinical evaluation of sexual preferences. All had been charged with, convicted of, or admitted to criminal offenses (rape, sexual assault, murder, kidnapping, forcible confinement) in which at least one woman (16 years or older) was the victim of physically forceful sexual assault. For 32, the most serious offense was rape or sexual assault; for 5, homicide (or attempted) was the most serious; and, for 1, sexual harassment was the most serious charge. Thirty-six rapists received a primary psychiatric diagnosis of personality disorder, 1 was psychotic, and 1 was mentally retarded.

Stimuli and Procedure

The audiotaped stimuli for this study have been described elsewhere (Quinsey *et al.*, 1981). There were 18, each describing an interaction between a man and a woman in four categories: 3 described a neutral interaction (e.g., a trip to a travel agent), 5 described consenting sexual activity, 5 described the violent and brutal rape of a stranger, 5 described the brutal nonsexual beating of a woman by a man. The stories, written in the second person, were recorded in two different random orders read in a male voice. After 20 rapists had been tested, the stories were rerecorded in a different random order in a female voice (but told in the second person) used for the remaining 18 participants. In all cases, testing was performed by the same male research assistant.

Participants sat alone in a comfortable chair in a sound-attenuated room. A mercury-in-silastic strain gauge was placed around the middle of the shaft of the penis and a board holding two push buttons was placed across the arms of the chair. While the participant listened to the stimuli, his penile tumescence was monitored using a Parks Model 270 plethysmograph (frequent calibrations ensured that plethysmograph output was linearly related to penile circumference). Penile expansion was recorded for the interval between 2 and 120 sec after stimulus onset. The 2-sec period directly before stimulus onset was the baseline for that trial.⁴ The

⁴Because the erectile response depends in turn on a vascular response (blood supply to the penis), no stimulus contingent penile response could occur within 2 sec (Bancroft and Mathews, 1967), and this is a true baseline.

experimenter ensured that the participant returned to baseline before beginning the next trial. Three warm-up trials (two neutral stories and one consenting sexual interaction) were presented before the experimental stimuli; penile responses were not recorded.

A randomly selected half of the participants were instructed to push buttons for the first half of the trials, and the remaining participants used the buttons during the second half of the session. Instructions were: "Please listen carefully to the following stories and imagine you are involved in each situation. Whenever you hear an action word (verb) which describes sexual touching (e.g., stroke, fondle, etc.), please press the red (right) button. Whenever you hear an action word (verb) which describes violence (e.g., kick, punch), please press the black (left) button. Push the button as soon as you hear a sexual or violent action." Participants practiced the button-pushing before the formal testing session began.

The rerecording of the stimuli and switch to a new random order coincided with the move of our laboratory to a new location. Except as noted above, this move resulted in no changes to our stimuli or procedures except: the control of experimental sessions and recording of responses switched from electromechanical programming relays and a polygraph to an IBM-XT computer. Button presses yielded slightly different scores: temporal duration versus raw frequencies. Stimuli were presented by means of a speaker in a soundproof chamber versus over headphones in a sound-attenuated room.

Treatment of the Data

The testing yielded data of two types: penile responses and data from the button-pressing task. In addition to the raw maximum penile expansion in each trial, phallometric responses from each participant and each session were converted to z scores. Again, following our earlier findings (Harris *et al.*, 1992), all participants were included with no minimum response criterion. Finally, we computed an index of sexual deviance: the greater of the mean z score response to rape or nonsexual violence minus mean z score response to consenting sexual stories.

The button-pressing data were of two types: for the first 20 participants, the duration each button was pressed; and for the other 18, the number of times each button was pressed. To have an index of button-pressing accuracy, simple rules were derived: in each neutral trial, 1 point was given for pressing neither button; on consenting trials, 1 point was given for pressing the "sex" button only; on nonsexual violence trials, 1 point was given for pressing the "violence" button only; and in rape trials, 1 point was given for pressing the "sex" button at least once, 1 point for pressing the "violence" button first, and 1 point for pressing the "violence" button more (because these trials all described more violence than sex). Thus, each participant had an accuracy score ranging from 0 to 14.

RESULTS

The mean raw responses and z -score conversions for all rapists are shown in Fig. 1. Several things are immediately apparent. First, the participants exhibited deviant preferences with or without the semantic tracking task. Nonrapists tested with the same stimuli showed a very strong preference for the consenting stimuli (Quinsey *et al.*, 1981; Quinsey and Chaplin, 1988), irrespective of the narrator. Second, button-pressing led to profiles where, on average, rapists exhibited an absolute preference for rape stimuli. Third, raw scores (which refer to the electrical output of the apparatus calibrated to be linearly related to penile circumference) suggest that the new laboratory might have produced higher levels of arousal overall (perhaps due to the use of a female voice), but that deviance was unaffected. Fourth, the differences overall seem small.

Analyses of variance on the raw and z -score data yielded no significant effects for the interaction of category by condition (pressing vs. not pressing). However, a planned comparison between the two (pressing vs. not pressing) consenting data points for the z scores yielded a statistically significant difference, $t(37) = 1.69$, $p = 0.05$, one-tailed, while the two corresponding data points for the rape stimuli were identical. The comparison of the z -score deviance differentials yielded a nonsignificant trend in the expected direction: 0.04 (SD = 1.26) without the semantic task vs. 0.22 (SD = 1.16) with the semantic task, $t(37) = 1.14$.

The mean overall tracking accuracy score was 7.4 (SD = 3.1) out of a maximum of 14.⁵ There was a marginally significant relationship between tracking accuracy and deviance differential, $r(38) = 0.23$, $p < 0.08$, one-tailed. We hypothesized that the value of the button-pressing task would be related to prior experience with phallometric testing and examined the 4 participants who had been tested before. There was a strong relationship between tracking accuracy and deviance in the expected direction, $r = 0.98$, $p < 0.05$, one-tailed. Examination of the initial test scores of these 4 participants revealed that the profiles obtained using the tracking task were more similar to their initial testing profiles (obtained before the present experiment), than were the profiles obtained without the tracking task. This was especially true for the two subjects who were more accurate in their tracking.

Finally, in drawing conclusions about the tracking task, it was important to show that tracking would not also make non-sex offenders yield more deviant phallometric responses. Thus, Fig. 2 shows the comparison between the z -score data from the rapists and 8 nonoffenders tested using the semantic tracking task but otherwise normal instructions from Quinsey and Chaplin (1988). The tracking

⁵This mean indicates that, although the accuracy scoring scheme might seem generous, it certainly did not produce a response ceiling—only 3 participants achieved perfect scores. It is also of interest that the participant with a diagnosis of psychosis scored above the mean (9), while the one with a diagnosis of mental retardation scored only slightly below the mean (5), and both were within a standard deviation of the overall mean.

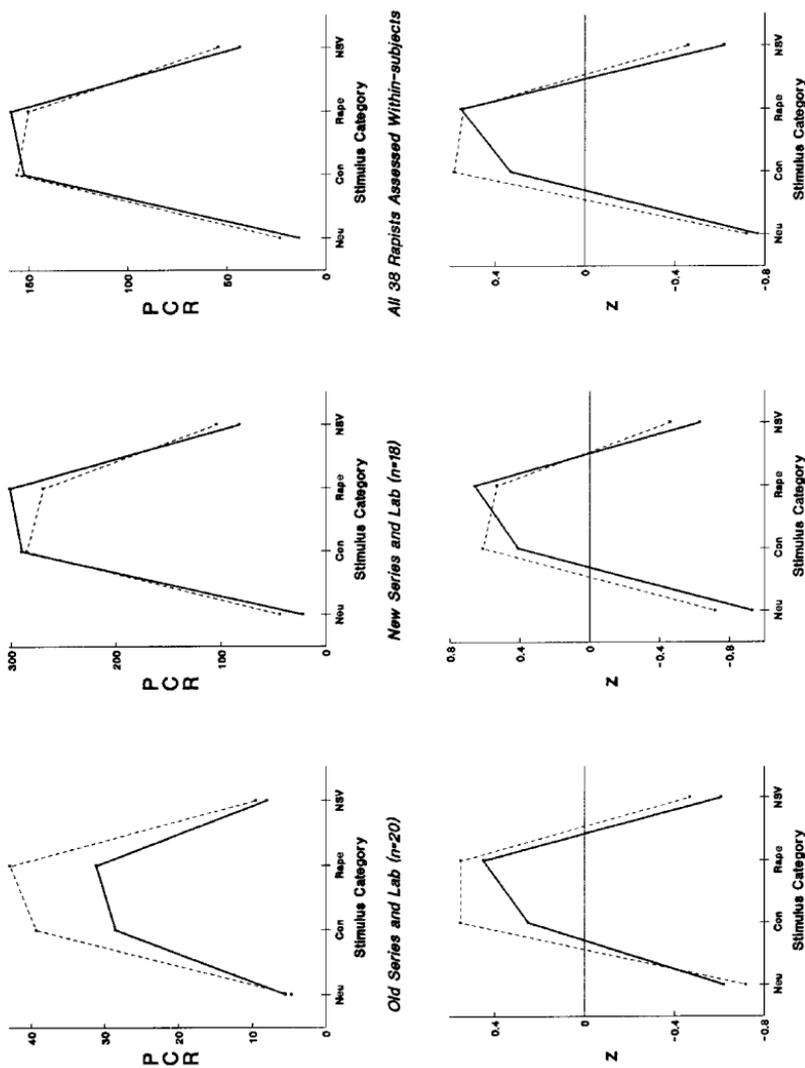


Fig. 1. Raw (top row) and z-score (bottom row) phallometric data for subjects assessed with (solid line), and without (dashed line) the semantic tracking task. Neu = neutral stimuli, Con = Consenting sex, NSV = Nonsexual violence.

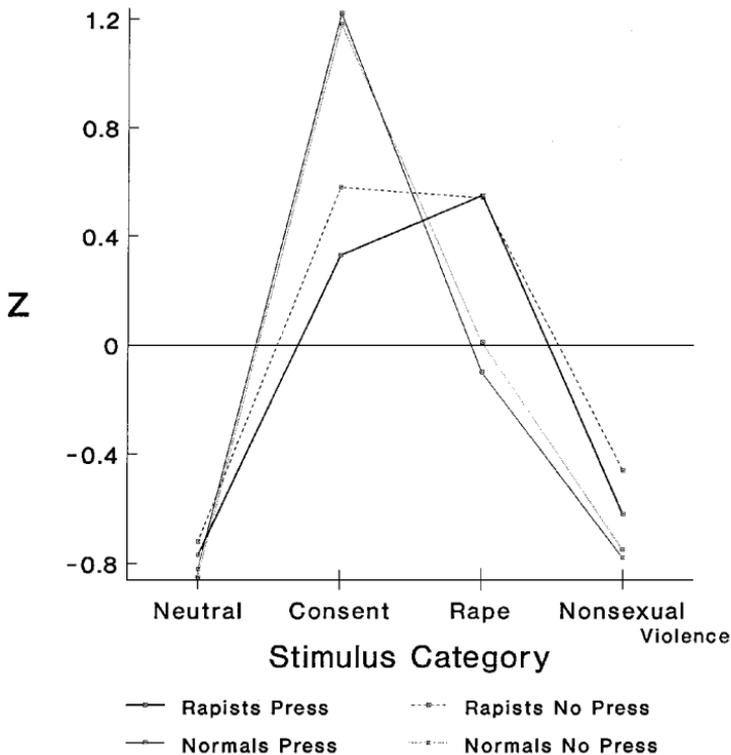


Fig. 2. Phallometric z-score data for the rapists in the present study and 8 nonoffenders from Quinsey and Chaplin (1988) tested with and without the semantic task and under otherwise normal (not faking) instructions.

task made only the rapists' responses more deviant. As a concise evaluation of the button-pressing procedure's improvement in discriminant validity, we used the relative operating characteristic (ROC; Rice and Harris, 1995). Briefly, an ROC is the empirically derived plot of the sensitivity–specificity trade-off for an experiment, and the area under that function is an index of discrimination, where an area of 1.0 indicates perfect accuracy. The ROC area for the 38 rapists versus the 8 nonoffender participants⁶ using the z-score deviance differential without buttons was 0.77 (SE = 0.068) compared to 0.87 (SE = 0.056) with buttons. ROC areas are mathematically and conceptually equivalent to the common language effect size (McGraw and Wong, 1992) and, even without buttons, these effect

⁶The participants in the Quinsey and Chaplin (1988) study were “preselected” to have nondeviant sexual preferences (resulting in the exclusion of 3 men). These ROCs, therefore, might slightly overestimate the difference between rapists and nonoffenders (but are consistent with data from unselected samples; Lalumière and Quinsey, 1993, 1994; Rice *et al.*, 1994). There is no reason, however, to expect that the effect of the semantic task was affected by this small amount of selection.

sizes are large by conventional standards (Rice and Harris, 1995). Nevertheless, an inferential test for within-subjects ROCs (Hanley and McNeil, 1983) showed that the semantic task improved the discriminant validity of phallometric assessment, $z = 1.62$, $p < 0.06$, one-tailed. Because most rapists yielded deviant results whether or not they pressed the buttons, the present results yielded data points only for that portion of the ROC where both hits and false alarms were high (i.e., high sensitivity and low specificity). Under these circumstances, it was not possible to be certain about the slope of the standardized ROC. Unitary slope depends on equal variance in the two populations and it appeared that rapists had larger variance (in their deviance differentials) than nonrapists, a finding consistent with the results of a meta-analysis of phallometric studies of rapists' and nonrapists' sexual preferences (Lalumière and Quinsey, 1994).

DISCUSSION

The results suggest that the semantic tracking task increased discriminant validity by interfering with rapists' deliberate or unintentional attempts to produce the pattern of responses characteristic of non-sex offenders. That is, the semantic tracking task appeared to make rapists' sexual responses more deviant. The semantic tracking task does not appear to cause an increase in deviant responding among men without histories of sexual aggression towards women. Most of the rapists had deviant indices even without the button-pressing task but more were deviant when the task was used. Among sexual aggressors with prior experience with phallometric assessment, the results suggest a large effect size in the relationship between accuracy and deviance, suggesting that semantic tracking has its greatest value with experienced assessees—those who tracked accurately yielded more deviant phallometric profiles.

That the semantic task differentially increased discriminant validity among offenders who had previously been assessed phallometrically is consistent with its only other reported test with sex offenders. Proulx *et al.* (1993) tested homosexual child molesters with and without the semantic tracking task. There was no difference found between testing conditions when subjects had no previous experience with phallometric assessments. Subjects exhibited more deviant responses, however, when tested with the semantic task if they had been tested before. The investigators interpreted this difference as a result of experienced subjects attempting to fake normal responding. Together with meta-analyses (Lalumière and Quinsey, 1994), these results suggest few sex offenders fake spontaneously and successfully in their first phallometric assessment, particularly when optimal methods are used (Harris *et al.*, 1992; Rice *et al.*, 1994). Semantic tracking may best be used with those who have received prior phallometric assessments and are motivated to dissimulate. Frequent retesting of sex offenders might be contraindicated because it increases opportunities to discover ways to fake.

The effects of the semantic tracking task on rapists' sexual responding, though statistically reliable, were modest. Because discriminant validity was already very high, there was little room for improvement. The present results probably provide a conservative estimate of the value of the semantic tracking task in phallometric assessment generally. Because the purpose of the semantic task was to make the brutality of rape unmistakable, the present results are consistent with conclusions expressed elsewhere that experimental manipulations that increase the salience of violence and victim suffering thereby increase discriminant validity (Chaplin *et al.*, 1995; Harris *et al.*, 1992; Lalumière and Harris, 1998; Harris and Rice, 1996; Lalumière and Quinsey, 1993, 1994; Rice *et al.*, 1994).

Finally, we distinguish between manipulations to foil attempted dissimulation and test results indicating that such attempts were made. Psychometricians usually tackle these separately, using indices of desirable (or undesirable) responding for the latter, and other means (very short latency responses, "physiological" responses, nonobvious items, etc.) for the former. The present results suggest that semantic tracking itself interferes with dissimulation while the accuracy of the tracking judgments provides an index of efforts to fake. We are currently pursuing these research questions further.

ACKNOWLEDGMENT

We thank Martin Lalumière for comments on an earlier version of this paper.

REFERENCES

- Abel, G. G., Barlow, D. H., Blanchard, E. B., and Mavissakalian, M. (1975). Measurement of sexual arousal in male homosexuals: Effects of instructions and stimulus modality. *Arch. Sex. Behav.* 4: 623–629.
- Adams, H. E., Motsinger, P., McNulty, R. D., and Moore, A. L. (1992). Voluntary control of penile tumescence among homosexual and heterosexual subjects. *Arch. Sex. Behav.* 21: 17–31.
- Bancroft, J. H., and Mathews, A. (1967). Penis plethysmography: Its physiological basis and clinical application. *Acta. Med. Psychosom.* 3: 8.
- Chaplin, T. C., Rice, M. E., and Harris, G. T. (1995). Salient victim suffering and the sexual responses of child molesters. *J. Consult. Clin. Psychol.* 63: 249–255.
- Freund, K. (1961). Laboratory differential diagnosis of homo- and heterosexuality—An experiment with faking. *Rev. Czech. Med.* pp. 20–31.
- Freund, K., Watson, R., and Rienzo, D. (1988). Signs of feigning in the phallometric test. *Behav. Res. Ther.* 26: 105–112.
- Hanley, J. A., and McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148: 839–843.
- Harris, G. T., and Rice, M. E. (1996). The science in phallometric testing of male sexual interest. *Curr. Direct. Psychol. Sci.* 5: 156–160.
- Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., and Earls, C. M. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychol. Assess.* 4: 502–511.
- Henson, D. E., and Rubin, H. B. (1971). Voluntary control of eroticism. *J. Appl. Behav. Anal.* 4: 37–44.

- Julien, E., and Over, R. (1988). Male sexual arousal across five modes of erotic stimulation. *Arch. Sex. Behav.* 17: 131–143.
- Lalumière, M. L., and Earls, C. M. (1992). Voluntary control of penile responses as a function of stimulus duration and instructions. *Behav. Assess.* 14: 121–132.
- Lalumière, M., and Harris, G. T. (1998). Common questions about phallometric assessment. *Sex. Abuse* 10: 227–237.
- Lalumière, M. L., and Quinsey, V. L. (1993). The sensitivity of phallometric measures with rapists. *Ann. Sex Res.* 6: 123–138.
- Lalumière, M. L., and Quinsey, V. L. (1994). The discriminability of rapists from non-sex offenders using phallometric measures: A meta-analysis. *Crim. Justice Behav.* 21: 150–175.
- Laws, D. R., and Rubin, H. B. (1969). Instructional control of an autonomic sexual response. *J. Appl. Behav. Anal.* 2: 93–99.
- Malcolm, P. B., Davidson, P. R., and Marshall, W. L. (1985). Control of penile tumescence: The effects of arousal level and stimulus type. *Behav. Res. Ther.* 23: 273–280.
- McGraw, K. O., and Wong, S. P. (1992). A common language effect size statistic. *Psychol. Bull.* 111: 361–365.
- Proulx, J., Côté, G., and Achille, P. A. (1993). Prevention of voluntary control of penile response in homosexual pedophiles during phallometric testing. *J. Sex Res.* 30: 140–147.
- Quinsey, V. L., and Bergersen, S. G. (1976). Instructional control of penile circumference. *Behav. Ther.* 7: 489–493.
- Quinsey, V. L., and Carrigan, W. F. (1978). Penile responses to visual stimuli: Instructional control with and without auditory sexual fantasy correlates. *Crim. Justice Behav.* 5: 333–342.
- Quinsey, V. L., and Chaplin, T. C. (1988). Preventing faking in phallometric assessments of sexual preference. In Prentky, R. A., and Quinsey, V. L. (eds.), *Human Sexual Aggression: Current Perspectives*, 528th ed., New York Academy of Sciences, New York, pp. 49–58.
- Quinsey, V. L., Chaplin, T. C., and Varney, G. W. (1981). A comparison of rapists' and non-sex offenders' sexual preferences for mutually consenting sex, rape, and physical abuse of women. *Behav. Assess.* 3: 127–135.
- Rice, M. E., Chaplin, T. C., Harris, G. T., and Coutts, J. (1994). Empathy for the victim and sexual arousal among rapists and nonrapists. *J. Interpers. Violence* 9: 435–449.
- Rice, M. E., and Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *J. Consult. Clin. Psychol.* 63: 737–748.
- Rosen, R. C., Shapiro, D., and Schwartz, G. E. (1975). Voluntary control of penile tumescence. *Psychosom. Med.* 37: 479–483.
- Rubin, H. B., and Henson, D. E. (1975). Voluntary enhancement of penile erection. *Bull. Psychonom. Soc.* 6: 158–160.
- Smith, D., and Over, R. (1987). Correlates of fantasy-induced and film-induced male sexual arousal. *Arch. Sex. Behav.* 16: 395–409.
- Wormith, J. S., Bradford, J. M. W., Pawlak, A., Borzecki, M., and Zohar, A. (1988). The assessment of deviant sexual arousal as a function of intelligence, instructional set and alcohol ingestion. *Can. J. Psychiat.* 33: 800–808.
- Wydra, A., Marshall, W. L., Earls, C. M., and Barbaree, H. E. (1983). Identification of cues and control of sexual arousal by rapists. *Behav. Res. Ther.* 21: 469–476.