Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Research in Personality 37 (2003) 205–223
JOURNAL OF RESEARCH IN PERSONALITY
www.elsevier.com/locate/jrp

Self-deception and failure to modulate responses despite accruing evidence of error

Jordan B. Peterson,[a,*] Colin G. DeYoung,[a] Erin Driver-Linn,[b] Jean R. Séguin,[c] Daniel M. Higgins,[b] Louise Arseneault,[d] and Richard E. Tremblay [c]

[a] *Department of Psychology, University of Toronto, 100 St. George Street, Toronto, Ont. M5S 3G3, Canada*
[b] *Harvard University, MA, USA*
[c] *Université de Montréal, Montréal, Que., Canada*
[d] *University of London, London, UK*

Abstract

Two studies assessed performance on a gambling-type card playing task (Newman, Patterson, & Kosson, 1987) by males defined as high or low in self-deception. Monetary success in this task depends upon the ability to modulate reward-seeking responses, by attending to information indicative of task-failure. In Study 1, 28 13-year-old boys categorized as high in self-deception using Eysenck's Junior Lie Scale (Eysenck & Eysenck, 1975), played more cards and won significantly less money than 143 categorized as low in self-deception. Study 2 replicated these findings in a sample of 42 male Harvard undergraduates defined as high or low in self-deception using Eysenck's Lie scale (Eysenck, Eysenck, & Barrett, 1985) and the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991). Also, a higher proportion of high self-deceivers played until the end of the task in both samples, thereby losing all their money, despite the fact that 19 of the last 20 cards were losing. These findings support a model of self-deception as ignoring evidence of error and reinforce the argument that self-deception may be maladaptive.

Keywords: Self-deception; Social desirability; Response modulation; Learning; Perseveration

* Corresponding author. Fax: +416-978-4811.
E-mail address: jpeterson@psych.utoronto.ca (J.B. Peterson).

0092-6566/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0092-6566(02)00569-X

1. Introduction

The construct of self-deception has a long history in psychology. The notion that the mind contains more information than is consciously realized was of central importance at the dawn of personality psychology, in the work of Janet, Freud, and Jung. The assertion that some of this information may be voluntarily hidden from consciousness has, from that time, been a closely linked corollary in personality theory. Despite a century of speculation and argument, however, the nature of the processes by which self-deception might take place is still subject to debate (for a review, see Mele, 1997). The last 15 years, furthermore, have seen an ongoing debate over whether self-deception is problematic and maladaptive, as traditional accounts have generally maintained (Colvin & Block, 1994), or whether it is beneficial, even necessary, for normal psychological functioning (Taylor & Brown, 1988). In what follows, we attempt to offer insight into these issues using a new model of self-deception (Peterson, 1999; Peterson, Driver-Linn, & DeYoung, 2002) and a behavioral measure of maladaptive perseveration.

Peterson (1999) and Peterson et al. (2002) have described self-deception as failure to utilize evidence indicating that current expectations or beliefs are in error. Human behavior is largely goal-directed, and unexpected or undesired disruption of progress toward a goal indicates an error or insufficiency in one's currently operative plans and beliefs and provokes an affective response (Carver & Scheier, 1982; Gray, 1982, 1987; Miller, Galanter, & Pribram, 1960; Oatley & Johnson-Laird, 1987; Panksepp, 1999; Peterson, 1999). Like Damasio's (1994) "somatic marker," this affective response serves initially to delimit and direct voluntary attention. When initial attention indicates that nothing additionally unpredictable or otherwise threatening is immediately likely to occur, evidence of error also elicits curiosity and exploration—cognitive, behavioral, or both (Blanchard & Blanchard, 1989; Dollard & Miller, 1950; Gray, 1982, 1987; Peterson, 1999). Exploration ideally transforms undifferentiated evidence of error into detailed, explicit, pragmatically useful information. Such information allows for recalibration of beliefs and transformation of plans and goals, so that the environment is once again rendered predictable and productive. Self-deception, in this model, consists of ignoring affectively-marked evidence of error, rather than exploring its implications. In consequence, self-deceivers cling rigidly to their current conceptions, and their behavioral and cognitive responses become increasingly maladaptive, as plans and beliefs are not questioned and adjusted to correspond more closely to the real situation.

Traditional models of self-deception, like Sackeim and Gur's (1978, 1979), require that self-deceivers must hold two contradictory beliefs, while remaining unaware that one of them is held. This capacity has been deemed paradoxical, or even impossible, on logical, philosophical, and psychological grounds (e.g., Mele, 1997). Alternative theories, therefore, presume that the central mechanism of self-deception is the ignoring of evidence conflicting with current beliefs (Greenwald, 1988; Mele, 1997). The major difficulty with such alternative models has been their assumption that the ignored evidence is a self-evident, objective feature of the situation in question. Sackeim and Gur (1978) argued convincingly that possession of conflicting ev-

idence, conceived as objective information, is insufficient to determine self-deception because a given individual may simply be ignorant of the significance of that information (due to lack of intelligence, background knowledge, prior experience, etc.). Our specification of the conflicting evidence as affectively-marked, and thus inherently subjective, addresses this criticism. It is not any objective feature of the situation that the self-deceiver ignores so much as his or her affective response to the situation—that twinge of emotion indicating that something is not as one would like it to be. Self-deception thus produces a mental state in which the self-deceiver continues to harbor, unquestioningly, plans and beliefs that have been indicated as problematic by his or her own affective responses.

Peterson et al. (2002) demonstrated that individuals scoring high on a commonly used measure of self-deception (Self-Deceptive Enhancement; Paulhus, 1991) identified an anomalous visual stimulus significantly more slowly than low-scoring individuals, despite apparently equal initial affective response to the anomaly. In the face of a situation that violated expectations, self-deceivers were slower to adjust, supporting the idea that self-deception entails failure to utilize evidence of error. In that paradigm, however, participants were simply asked to identify stimuli, and there was no emphasis on performance; nor were there any consequences for taking longer to make correct identifications. The current studies, therefore, employed Newman's Card Playing Task (Newman et al., 1987), to examine a situation in which self-deception might be more obviously maladaptive, and to test the hypothesis that self-deceptive individuals would respond less readily to indications that their current strategy of action was not producing the desired outcome.

The Card Playing Task is a gambling task, employing principles of classical learning theory to test sensitivity to extinction. Participants are told, prior to beginning the task, that they may keep all the money they make in the task, and that they may stop playing at any time. Participants are rewarded (given money) on an almost continuous basis at the beginning of the task. As the task continues, however, the contingencies gradually switch. Toward the end of the task, participants are being punished (losing money) almost continuously. To perform well on the task requires "response modulation" (Newman & Lorenz, 2001), which has been described by Patterson and Newman (1993) as "temporary suspension of a dominant response set and a brief concurrent shift of attention from the organization and implementation of goal-directed responding to its evaluation" (p. 717). Poor performance on the task results from a failure of response modulation, leading to perseveration—continuing to play when the ratio of wins to losses is clearly no longer positive (Newman et al., 1987).

Self-deception, in our model, bears an obvious relation to failure of response modulation. Newman and Lorenz (2001) note that "given a deficit in response modulation, individuals are less likely to suspend approach behavior (i.e., pause), evaluate their response strategies (i.e., reflect), and learn from corrective feedback." Self-deceivers similarly fail to reflect on evidence that should prompt them to explore potential errors in their plans and beliefs. This is not to say that failure of response modulation is identical to self-deception—other processes than self-deception could certainly lead to failure of response modulation—but the similarities between the

two constructs suggest that self-deceivers are likely to fail to modulate their responses.

We predicted, therefore, that individuals high in self-deception (High-SDs) would be likely not to modify their game strategy in the Card Playing Task in response to feedback indicating increasing loss—in other words, that they would show greater perseveration than individuals low in self-deception (Low-SDs). We hypothesized that High-SDs would play more cards and earn less money than Low-SDs.[1] We further hypothesized that a greater proportion of High-SDs would lose all their money by playing their cards right to the end, despite the fact that 19 out of the last 20 cards were losing. Such behavior would support the assertion that self-deceivers are likely to persist in their plans and beliefs even when it is clearly to their own detriment.

1.1. Questionnaire measures of self-deception

In order to assess the level of self-deception of our participants we used two instruments, both of which fall under the broad heading of measures of socially desirable responding: the Balanced Inventory of Desirable Responding (BIDR), which consists of three subscales, Self-Deceptive Enhancement, Impression Management, and Self-Deceptive Denial (Paulhus, 1991; Paulhus, 1999) and Eysenck's Lie scale (standard and "Junior" versions; Eysenck et al., 1985; Eysenck & Eysenck, 1975). The BIDR was developed out of the earlier Self-Deception and Other-Deception Scales (Paulhus, 1984; Sackeim & Gur, 1979). The Self-Deceptive Enhancement and Denial scales were designed to assess self-deceptive tendencies, while the Impression Management scale was intended to assess the tendency to fake good—that is, consciously to exaggerate one's good qualities and play down one's bad qualities. The Lie scale contains items very similar to those included in Impression Management and was designed for the same purpose.

Despite the fact that the latter two scales were designed to control for consciously biased responses on personality questionnaires, evidence has accumulated that such scales do not adequately serve this purpose. Multiple researchers have demonstrated that controlling for socially desirable responding tends to decrease, rather than increase, the correlation between self-reports and other-ratings on personality questionnaires (Borkenau & Amelang, 1985; McCrae & Costa, 1983; Piedmont, McCrae, Riemann, & Angleitner, 2000). Similarly, controlling for socially desirable responding has not been found to improve criterion-related validities of personality predictors of job performance (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Ones, Viswesvaran, & Reiss, 1996). Remarkably, Eysenck and Eysenck (1975) argued that their own Lie scale probably did not, in fact, typically measure lying,

[1] Despite the fact that cards played and winnings are negatively correlated over the entire task, their relation is not linear, as winnings increase in the first third of the task, level out for the next third, and drop for the final third. Thus, both cards played and winnings must be examined to identify maladaptive perseveration. Even when cards played and winnings are highly correlated in an actual sample (as is especially true in our second study) because few or no participants dropped out in the early third of the task, the curvilinear relation of the two measures makes it statistically incorrect to combine them.

but rather some stable dimension of personality, the nature of which was not clear to them. The Lie scale has since been used to assess "repression" (e.g., Kline, Schwartz, Allen, & Dikman, 1998), a construct derived from the Freudian tradition, which is directly analogous to self-deception, as it involves hiding unpleasant information from conscious awareness. Our model assumes that any evidence that one's beliefs or plans are in error may be unpleasant enough to motivate self-deception in some individuals.

One of the most recent developments in understanding measures of socially desirable responding and self-deception stems from factor analyses that have identified two distinct types of content subsumed by the various scales (Paulhus & John, 1998; Raskin, Novacek, & Hogan, 1991). One is characterized by overconfidence—claims of heightened ability, especially social and intellectual (e.g., "I am a completely rational person," Self-Deceptive Enhancement); the other is characterized by claims of heightened conformity with societal moral norms and by the denial of deviant impulses (e.g., "I never read sexy books or magazines," Impression Management). These factors have been labeled "egoistic bias" and "moralistic bias," respectively, (Paulhus & John, 1998) and they appear to constitute a more viable classification system for measures of socially desirable responding than the distinction between whether a scale measures self-deception or conscious lying. According to Paulhus and John (1998), both factors may indicate self-deceptive tendencies; one can be egoistically self-deceptive or moralistically self-deceptive. Both are consistent with the notion of self-deception as ignoring evidence of error: Self-deceivers are likely to ignore evidence that their abilities are not as impressive as they had believed or that their moral standards are not as perfect as they had believed.

Of the four scales employed in the current studies, Self-Deceptive Enhancement loads on the egoistic factor, and the other three load on the moralistic factor (Paulhus, 1999; Paulhus & John, 1998). While Self-Deceptive Enhancement is the best validated and most commonly used questionnaire measure of self-deception (e.g., Fossum & Barrett, 2000; Johnson, Vincent, & Ross, 1997; Paulhus, 1991), it is of interest to examine scales that load on the moralistic factor, as well, to determine whether different types of self-deception scales have similar behavioral correlates.

2. Study 1

2.1. Method

Participants in this study were part of an ongoing longitudinal study of the etiology and correlates of aggressive and/or anxious behavior in boys (Nagin & Tremblay, 2001; Tremblay, Pihl, Vitaro, & Dobkin, 1994). All 6-year old French-speaking Caucasian boys attending kindergarten ($N = 1037$) in the 53 lowest socioeconomic status schools of the Catholic School Commission of Montreal were identified and selected for an initial wave of school-based longitudinal assessments. At age 13, a subset of these boys were selected for a laboratory study, for reasons not related to the present analysis; physical aggression and negative affectivity scores on teacher

rating scales at ages 6, 10, 11, and 12, were used to select equal numbers of stably aggressive, unstably aggressive, and non-aggressive boys, excluding any who scored between the 50th and 70th percentile on negative affectivity, to enhance detection of potential interactions between physical aggression and negative affect (Séguin, Arseneault, Boulerice, Harden, & Tremblay, 2002; Séguin, Pihl, Harden, Tremblay, & Boulerice, 1995). This subset of thirteen-year-old boys ($N = 197$) participated in laboratory studies and were administered the Junior Eysenck Personality Questionnaire at that time (JEPQ; Eysenck & Eysenck, 1975).

The JEPQ contains 97 items, each answered "yes" or "no," broken down into four subscales: Neuroticism, Extraversion, Psychoticism, and Lie. The Junior Lie scale contains items similar in content to the standard Lie scale, but worded to apply to children or adolescents. The following year, at age 14, 171 of these boys also completed a short version of the Junior Lie scale. To increase reliability, we based our categorization of High and Low self-deceivers on both sets of Lie scale scores. (The BIDR, which is not appropriate for this age group, was not administered.)

From the sample of 171 boys, we selected as High-SDs those individuals who scored in the upper quartile on the Lie scale at age 13, but excluded any of these whose scores on the short form Lie scale at age 14 crossed to the lower side of the median. An extreme-groups approach was used because only extreme high scores on measures of self-deception seem likely to be associated with maladaptive behavior. This is implied by the finding, reviewed above, of the two content factors in the questionnaires: Some degree of "egoism" or confidence in one's abilities is presumably harmless or even beneficial; only *over*confidence should be considered problematic and probably self-deceptive. Similarly, some degree of "moralism" or conformity may be a natural component of socialization and only in the extreme does it seem likely to be problematic. We therefore compared our High-SD group to the rest of the sample, whom we assumed to be in the non-self-deceptive range and designated as the Low-SD group. (This approach is consistent with the methodology of research on repression, in which scores on the Lie scale and similar measures are typically cut at the top third or quarter to identify repressors; Kline et al., 1998.) Because of the degree of arbitrariness inherent in determining cutoff scores, which precludes us from being certain that there are no High-SDs who fall in the group designated as Low-SD, we also performed the analyses comparing High-SDs to Extreme Low-SDs, applying the same criteria described above for High-SDs to the lower end of the scales. Our cutoffs produced 28 High-SDs and 143 Low-SDs (of whom 30 were Extreme Low-SDs).

All participants had completed the Card Playing Task, among other questionnaires and tasks, at age 13. It is a computerized task in which participants are presented with a depiction of a deck of cards, with a question mark on the back, and two buttons, one of which turns over the next card and one of which quits the game. Money is won or lost depending upon the value of the card turned over. Participants were told beforehand that face cards were winners and all other cards were losers, that the deck was not a normal 52-card deck, that they could quit whenever they wished, and that they could keep their winnings. The deck consisted of one hundred cards. With each successive block of ten cards, the ratio of wins to losses decreased.

Within the first block, 9 out of the 10 cards were winning. In the second block, the win ratio was 8/10, in the third, 7/10, and so on. Thus, the probability of wins decreased linearly over the task. Maximum winnings ($1.55) were gained if a participant chose to stop about half way through the 100 cards. The boys were given 50 cents to start with and won or lost 5 cents on every card.

Because Séguin et al. (2002) found that physical aggression predicted performance on the Card Playing Task in this sample,[2] we controlled for physical aggression in our analyses by using it as a covariate in ANOVA, to determine whether self-deception would predict performance over and above aggression. The physical aggression scale consisted of three items: ''Fights with other children''; ''Kicks, bites, hits other children''; and ''Bullies other children.'' This scale has been found to be highly reliable ($\alpha = .86$; test–retest reliability after two months $= .83$) and convergent validity has been demonstrated by correlations with peer-rated aggression (Tremblay et al., 1991). A single physical aggression score was calculated by averaging standardized scores from each of the four assessments between ages 6 and 12.

2.2. Results and discussion

Correlations between all JEPQ scales, physical aggression, and winnings and total cards played in the Card Playing Task are presented in Table 1, demonstrating that no significant relation appears between the Lie scales and performance on the task when the data are treated continuously.

Table 2 shows the results of the ANOVAs. High-SDs earned significantly less money than Low-SDs. They also played more cards than Low-SDs. The effect sizes were even larger when High-SDs were compared to Extreme Low-SDs. These results are over and above the contribution of physical aggression, which, as expected, was a significant predictor, or approached significance, as a covariate in the ANOVAs.[3] Furthermore, a greater proportion of High-SDs played all the way to the end of the task, thereby losing all their money (Table 3). This is a particularly notable failure, as 19 out of the last 20 cards are losing. The increase in effect size found when using the Extreme Low-SD group suggests that some individuals genuinely high in self-deception fall outside the cutoffs for the High-SD group, thereby weakening the effect when the High-SD group is compared to the rest of the sample.

We have argued that self-deceptive individuals may perform worse on the Card Playing Task because they are likely to ignore evidence of error. An alternative hypothesis would be that High-SDs are simply less sensitive to negative affect than Low-SDs, meaning that they might experience losing money in the Card Playing

[2] Portions of the age-13 data reported here have previously been reported by Séguin et al. (2002), who found that aggression was associated with increased perseveration on the Card Playing Task. In addition to aggression scores, Séguin et al. (2002) reported the JEPQ scores for Extraversion, Neuroticism, and Psychoticism, which we utilized in the current study, but they did not present any data on the Lie scale.

[3] Physical aggression was a significant predictor of performance in the total sample (Winnings: $F = 10.59,\ p < .01$; Cards Played: $F = 7.82,\ p < .01$), and approached significance in the comparison of extreme self-deception groups (Winnings: $F = 3.71,\ p < .06$; Cards Played: $F = 2.80,\ p = .10$).

Table 1
Study 1: Correlations for JEPQ, aggression, and performance on the card playing task

	1	2	3	4	1	6	7
1. JEPQ Extraversion	—						
2. JEPQ Neuroticism	−.04	—					
3. JEPQ Psychoticism	−.04	.32**	—				
4. JEPQ Lie (full)	−.05	−.09	−.16*	—			
5. JEPQ Lie (short)	−.27**	−.12	−.12	.50	—		
6. Physical aggression	−.05	.11	.29**	−.09	−.01	—	
7. Cards played	−.07	.03	.03	.11	.06	.19*	—
8. Winnings	.07	−.06	.01	−.11	−.10	−.23**	−.88**

Note. $N = 171$. JEPQ, Junior Eysenck Personality Questionnaire.
* $p < .05$.
** $p < .01$ (2-tailed).

Table 2
Study 1: Means (with standard deviations) for performance of High and Low self-deceivers on the card playing task, controlling for physical aggression

Comparison	Low-SD	High-SD	*F*	*d*
Total sample				
Cards played	81.12 (25.00)	90.25 (18.47)	5.29*	.42
Winnings	$0.54 ($0.58)	$0.31 ($0.50)	4.58*	−.42
Extreme groups				
Cards played	75.53 (27.10)	90.25 (18.47)	8.21**	.63
Winnings	$0.67 ($0.55)	$0.31 ($0.50)	6.64*	−.68

Note. SD, Self-deception; High-SD, $N = 28$; Low-SD, $N = 143$; Extreme Low-SD, $N = 30$.
* $p < .05$.
** $p < .01$.

Table 3
Study 1: Number of High and Low self-deceivers playing to the end of the card playing task

	Cards played (winnings)	
	<100 ($0.05–$1.55)	100 ($0.00)
High	9	19
Low	76	67
Ex-Low	22	8

Note. High/Low comparison $\chi^2(1, N = 171) = 4.13$, $p < .05$; High/Ex-Low comparison $\chi^2(1, N = 58) = 9.88$, $p < .01$; Ex-Low = Extreme Low self-deceivers.

Task as less negative. The fact that social desirability measures typically correlate negatively with various measures of negative affect, including Neuroticism (e.g., DeYoung, Peterson, & Higgins, 2002; Sackeim & Gur, 1979; Taylor & Brown, 1988), offers potential support for this hypothesis. Also, participants in this study were selected on the basis of relatively high or low negative affectivity, which could amplify this effect. Despite the fact that Neuroticism was not correlated with self-deception in

Table 4
Study 1: Mean JEPQ scores (with standard deviations) for all participants who played to the end of the card playing task compared to those who stopped before the end, controlling for aggression

	Cards played		F	d
	<100 ($N = 85$)	100 ($N = 86$)		
JEPQ Extraversion	18.01 (3.25)	17.74 (3.49)	.16	.08
JEPQ Neuroticism	9.00 (4.24)	9.41 (5.27)	.07	−.08
JEPQ Psychoticism	4.62 (3.13)	4.87 (3.02)	.09	−.08
JEPQ Lie (full)	9.06 (4.06)	10.45 (4.37)	6.22*	−.33
JEPQ Lie (short)	1.67 (1.58)	2.07 (1.49)	3.09†	−.26

Note. JEPQ, Junior Eysenck Personality Questionnaire.
* $p < .05$.
† $p < .10$.

this sample (Table 1), we decided to repeat our ANOVA controlling for Neuroticism as a covariate, in an attempt to offer some evidence that the link between self-deception and perseveration is not likely to be mediated by divergent affective responses.

Extraversion also has potential for relevance, as it has been convincingly argued that the central feature of Extraversion is incentive motivation—that is, sensitivity to the possibility of reward (Depue & Collins, 2000). Two contradictory hypotheses can be framed regarding reward: Individuals who perseverate on the Card Playing Task might be overly sensitive to the possibility of reward, leading them to respond impulsively. Alternatively, sensitivity to potential monetary rewards might motivate individuals to play the game *more* carefully. Extraversion was negatively correlated with one measure of self-deception in this sample (Table 1), which might seem to support the second hypothesis (i.e., if Low-SDs are more extraverted, they might be more motivated to win money). We therefore performed a second set of ANOVAs, identical to those described above except for the inclusion of Extraversion and Neuroticism as additional covariates. The differences between High-SDs and Low-SDs for cards played and winnings remained significant ($p < .05$ in all ANOVAs), and neither Extraversion nor Neuroticism was ever a significant predictor of performance ($p > .10$ in all ANOVAs).

Because of the desirability of analyzing continuous variables continuously, given that cutoffs are necessarily somewhat arbitrary, we also analyzed the data in reverse, using the Card Playing Task as a criterion and comparing individuals who played to the end with those who did not, once again controlling for physical aggression. Table 4 shows that the 86 boys who played all 100 cards scored significantly higher on the Lie scale than the 85 who did not, though they did not differ on Extraversion, Neuroticism, or Psychoticism.

3. Study 2

A second study was organized to replicate the findings of Study 1 in a different population and to include a broader range of self-deception measures. The boys

who participated in Study 1 had a fairly specific demographic profile and were selected on the basis of physical aggression and negative affect scores that rendered the sample a slightly biased representation of the population, even within that demographic profile. For the second study, therefore, we recruited undergraduates, a population not without its own characteristic demographics, but more comparable with most other psychological research. Employing young adults instead of 13-year-olds also allowed us to include the BIDR as well as the Lie scale, providing not only a more widely used measure of self-deception, but also the means to examine both egoistic and moralistic self-deception in relation to performance on the Card Playing Task. In this study, we hypothesized that High-SDs in the undergraduate population would tend to perseverate on the Card Playing Task more than Low-SDs, just as they did in Study 1. We also hypothesized that both egoistic and moralistic self-deception would be associated with perseveration.

3.1. Method

Participants in this study were Harvard University undergraduates. A total of 156 male students were screened with the BIDR and the adult version of the Lie Scale. (Only males were chosen to participate in this study so as to render the results more easily interpretable in relation to Study 1 and to past research using the Card Playing Task, which had previously been tested on males almost exclusively [e.g., Giancola, Peterson, & Pihl, 1993; Newman et al., 1987].).

Volunteers completed the screening questionnaires at several Harvard College common spaces. To decrease potential social desirability effects, only individuals seated alone were approached. A cover sheet provided with the questionnaires assured volunteers of anonymity. This sheet also indicated that there was a 50% chance that volunteers would be telephoned to schedule participation in an experiment paying $5.00 per hour plus winnings on a computerized gambling-type card game. All screened participants provided their names and phone numbers on the cover page, so that they could be contacted, but were told explicitly that their names would not be associated with their questionnaire responses. Instead, an ID number, written in the corner of each sheet, was used for this purpose. While participants watched, the experimenter removed the cover page from the questionnaire and placed it randomly among others in a large envelope. To ensure that researchers were blind to the participants' levels of self-deception, different experimenters entered the data, scheduled participants, and ran participants through the protocol.

As described in Section 1, the BIDR consists of three subscales of 20 items each: Self-Deceptive Enhancement, Impression Management, and Self-Deceptive Denial. Self-Deceptive Denial has not typically been included in the BIDR because high correlations with Impression Management make it somewhat redundant (Paulhus, 1999). We felt it to be worth including here, however, because it loads on the moralistic factor of social desirability and, unlike the Lie and Impression Management scales, was specifically designed as a measure of self-deception. Responses to the BIDR are given on a seven-point Likert scale ranging from *not true* to *very true*, with only extreme high responses (6 and 7, or 1 and 2 for reversed items) being scored

(Paulhus, 1991). The BIDR has a high degree of internal consistency (Cronbach's $\alpha = .83$) and adequate test–retest reliability (.65–.69) (Paulhus, 1991). The Lie scale includes 12 statements similar to Impression Management items, but with each item rated categorically (*Yes* or *No*). Internal reliability scores on this scale range from .73 to .82 (Eysenck et al., 1985).

Individuals were selected from the pool of screened participants if their combined standardized BIDR and Lie scores fell in the top or bottom quartile of the population. (Based on the difference in effect size in Study 1 when High-SDs were compared to all other participants versus to Extreme Low-SDs, this procedure seemed likely to enhance the possibility of detecting an effect, while conserving resources). Forty-seven men who met these criteria proved willing to report to the laboratory, where they completed the full Eysenck Personality Questionnaire (EPQ; Eysenck et al., 1985) and a computerized version of the BIDR. The following analyses employ BIDR and Lie scale scores obtained in the laboratory administration, as we assume the controlled setting rendered them more reliable than the screening data.

For the Card Playing Task, participants were seated in front of a computer and told they would be playing a simple "gambling-type game." Each participant was given $1.00 to start. They were instructed, both verbally and in writing presented on-screen, to play by turning the cards (using the mouse to click a button labelled, "Play again") until they wished to stop, at which point they could click a button labelled "Quit." Before starting, they were also informed that they could keep their winnings. Winning cards resulted in 10 cents being added to the total and were accompanied by a pleasant "win" sound. Losing cards resulted in 10 cents being subtracted from the total and were accompanied by a buzzing "loss" sound. The probability of picking a winning card decreased by 10% after every block of 10 cards, as in Study 1. When participants finished the task, they were debriefed and paid ($5.00 per hour plus winnings, which in this study reached a maximum of $3.00).

Five participants were dropped from the analysis, one due to a malfunction of the computer program and four because they did not play the Card Playing Task long enough to receive any negative feedback about their performance (that is, long enough in the first block of ten cards to receive the first losing card),[4] meaning that their response to negative feedback could not be assessed. Because we employed the Card Playing Task specifically to gauge responses to negative feedback, these participants can not be considered relevant to our hypothesis.

3.2. Results and discussion

Correlations between the EPQ, BIDR, and winnings and total cards played in the Card Playing Task are presented in Table 5. Correlations indicated that one measure

[4] One possible reason that these four may have quit so soon is that participants in this study were not told that the deck of cards was not a normal 52-card deck. If one assumed that it were, the odds of picking face cards would seem so unfavorable that quitting almost immediately would be logical. After playing a few cards, however, it becomes obvious that the deck is not normal, as 19 of the first 20 cards are face cards, some of which are duplicates. Of the participants other than these four, none played less than 40 cards.

Table 5
Study 2: Correlations for EPQ, BIDR, and performance on the card playing task

	1	2	3	4	5	6	7	8
1. EPQ Extraversion	—							
2. EPQ Neuroticism	−.24	—						
3. EPQ Psychoticism	−.10	−.01	—					
4. EPQ Lie	−.09	−.52**	.11	—				
5. BIDR SDE	.11	−.47**	.25	.52**	—			
6. BIDR IM	.00	−.60**	−.06	.72**	.65**	—		
7. BIDR SDD	.09	−.54**	.17	.66**	.66**	.78**	—	
8. Cards played	−.14	−.08	.13	.37*	.30†	.16	.19	—
9. Winnings	.15	.10	−.15	−.40**	−.30†	−.17	−.20	−.97**

Note. $N = 42$. EPQ, Eysenck personality questionnaire; BIDR, balanced inventory of desirable responding; SDE, self-deceptive enhancement; IM, impression management; SDD, self-deceptive denial.

* $p < .05$.

** $p < .01$.

† $p < .06$ (2-tailed).

from the egoistic factor (Self-Deceptive Enhancement) and one measure from the moralistic factor (the Lie scale) were related to performance (Table 5) in the predicted direction. Because participants were preselected for extreme scores on these scales, however, it is potentially misleading to treat them as continuous variables. We therefore made two groupings of High and Low self-deceivers, by median splits of the Self-Deceptive Enhancement and Lie scale scores obtained in the laboratory.

High-SDs on Self-Deceptive Enhancement ($N = 22$) won less money ($M = \$1.20$, $SD = \$1.26$) than Low-SDs ($N = 20$; $M = \$2.01$, $SD = \$1.14$; $F = 4.68$, $p < .05$, $d = -.67$). They also played more cards ($M = 82.95$, $SD = 19.12$) than Low-SDs ($M = 69.00$, $SD = 20.16$; $F = 5.30$, $p < .05$, $d = .71$). In this sample, all of the self-deception measures were strongly negatively correlated with Neuroticism, and none were significantly correlated with Extraversion (Table 5). Controlling for Extraversion and Neuroticism as covariates in ANOVA did not eliminate the significant differences between High-SDs and Low-SDs on winnings ($F = 4.41$; $p < .05$) and cards played ($F = 5.24$; $p < .05$). Neither Extraversion nor Neuroticism was a significant predictor ($p > .20$ in both ANOVAs). A greater proportion of High-SDs also played all the cards, thereby losing all their money (Table 6).

A similar pattern emerged for High- Low-SDs on the Lie scale, although 7 participants were excluded from the groups because their scores fell exactly on the median. High-SDs ($N = 19$) won less money ($M = \$1.33$, $SD = \$1.33$) than Low-SDs ($N = 16$; $M = \$2.08$, $SD = \$1.04$; $F = 3.40$, $p < .05$, 1-tailed, $d = -.63$). They also played more cards ($M = 80.26$, $SD = 19.61$) than Low-SDs ($M = 68.38$, $SD = 20.66$; $F = 3.01$, $p < .05$, 1-tailed, $d = .60$). When Neuroticism and Extraversion were covaried out in ANOVA, the differences between High-SDs and Low-SDs fell below significance (Winnings: $F = 1.32$, $p > .10$; Cards Played: $F = 1.38$, $p > .10$). However, neither Extraversion nor Neuroticism was a significant predictor ($p > .50$ in both ANOVAs), and partialling these two variables out of the correlations between the Lie scale and task performance scarcely diminished them (winnings: partial

Table 6
Study 2: Number of high and low self-deceivers playing to the end of the card playing task

		Cards played (winnings)	
		<100 ($0.10–$3.00)	100 ($0.00)
Self-deceptive enhancement	High	12	10
	Low	17	3
Lie	High	10	9
	Low	14	2

Note. Self-deceptive enhancement: $\chi^2(1, N = 42) = 4.55$, $p < .05$; Lie: $\chi^2(1, N = 35) = 4.90$, $p < .05$.

$r = -.39$; cards played: partial $r = .36$; $p < .05$ for both). Finally, once again, a greater proportion of High-SDs played to the end (Table 6).

Interestingly, there was not complete overlap in group membership for High-SDs and Low-SDs on Self-Deceptive Enhancement and the Lie scale. Four High-SDs on Self-Deceptive Enhancement were Low-SDs on the Lie scale, and 6 Low-SDs on Self-Deceptive Enhancement were High-SDs on the Lie scale. Both egoistic and moralistic self-deceptive tendencies therefore appear to be associated with failure to utilize evidence of error. Given the similarity of Impression Management and Lie scale items, and given that these two scales and Self-Deceptive Denial all load on the moralistic factor, it would be somewhat surprising if Self-Deceptive Enhancement were the only BIDR scale significantly associated with task performance. Impression Management and Self-Deceptive Denial do show non-significant correlations in the expected direction, and these effects might prove significant in a larger sample. It is informative, in this regard, to compare the 13 participants who played all the way to the end, and thus won nothing, with the 29 who did not. As can be seen in Table 7, these participants scored significantly higher on all three BIDR subscales, as well as the Lie scale, though they were not significantly different in Extraversion, Neuroticism, or Psychoticism.

Table 7
Study 2: Mean EPQ and BIDR scores (with standard deviations) for participants who played to the end of the card playing task compared to those who stopped before the end

	Cards played		*F*	*d*
	<100 ($N = 29$)	100 ($N = 13$)		
EPQ Extraversion	7.55 (3.73)	6.38 (3.78)	.87	.31
EPQ Neuroticism	4.86 (3.23)	3.54 (2.85)	1.62	.43
EPQ Psychoticism	2.79 (1.93)	3.92 (2.06)	2.95	−.57
EPQ Lie	2.38 (2.29)	5.38 (3.07)	12.48**	−1.10
BIDR SDE	4.55 (3.32)	8.08 (4.39)	8.26**	−.91
BIDR IM	4.24 (3.87)	6.92 (3.88)	4.30*	−.69
BIDR SDD	4.59 (2.93)	6.85 (3.39)	4.84*	−.71

Note. EPQ, Eysenck personality questionnaire; BIDR, balanced inventory of desirable responding; SDE, self-deceptive enhancement; IM, impression management; SDD, self-deceptive denial.
* $p < .05$.
** $p < .01$.

4. General discussion

We predicted that individuals high in self-deception would be less responsive to negative feedback—that they would perseverate, showing less behavioral change when contingencies unexpectedly changed. The only way to maximize winnings in the Card Playing Task is to respond to feedback indicative of failure with response modulation, interrupting one's current response set. High self-deceivers in two very different populations showed greater response perseveration than did low self-deceivers. These results lend credence to the idea that self-deception consists of ignoring evidence of error, and to the hypothesis that self-deception may frequently have maladaptive consequences.

Séguin et al. (2002) have noted that there is a point in the Card Playing Task around 75 cards at which the rate of quitting increases sharply. At this point in the task, total winnings have begun to decrease steadily, after remaining fairly constant for the middle third of the task. For many participants, this seems to be the point at which sufficient evidence has accrued to judge that continued playing will not be advantageous. Consistent with the notion that self-deceivers ignore such evidence, we found that Extreme Low-SDs in Study 1 and Low-SDs in Study 2 played on average about 75 cards or fewer, whereas High-SDs played an average of 90 cards in the first sample and 83 in the second. Even more telling are the analyses focusing on participants who never quit, playing until the end of the task and losing all their money; this group scored significantly higher on all self-deception measures, in both studies.

The finding, in Study 2, that both egoistic and moralistic self-deception were associated with failure of response modulation may be illustrative of an important commonality between the two content factors that Paulhus and John (1998) observed in scales measuring social desirability and self-deception. It is not necessarily obvious why measures relating to egoistic overconfidence should be correlated with measures relating to moralistic conformity (as was true in our results, and as has typically been found in other studies; e.g., Lajunen, Corry, Summala, & Hartley, 1997; Fossum & Barrett, 2000). While both suggest a bias of some kind, the two biases seem quite different, at face value. Our model of self-deception indicates that what both probably share is a rigid devotion to the rightness of one's current plans and beliefs, whether these be provided from within, in the case of egoism, or from without, in the case of adherence to societal morality. As this rigidity appears to stem from the failure to explore evidence that one might be in error (that is, from self-deception), egoistic and moralistic biases may tend to vary together because each emerges from the same underlying process. Our results in this study further suggest that both egoistic and moralistic self-deception are likely to be maladaptive.

The paradigm used in these studies did not directly demonstrate participants' subjective registration of anomaly or error, as Peterson et al. (2002) did. This leaves open the possibility that High self-deceivers may not have noticed that they began to lose more and more as the task went on, rendering ignorance rather than self-deception responsible for their poor performance. Such obliviousness seems unlikely, however, as the total amount of money accrued by the participant was always visi-

ble. Further, as noted above, a greater proportion of High-SDs in both studies played through all 100 cards, despite the fact that 19 out of the last 20 were losing. Under these circumstances, the most parsimonious explanation appears to be that individuals high in self-deception failed to modulate their responses, despite being aware that their actions were increasingly failing to produce the desired results. While not entirely conclusive, the finding that statistical control for Neuroticism (associated with negative affectivity; e.g., John & Srivastava, 1999) and Extraversion (associated with reward sensitivity; Depue & Collins, 2000) did little to diminish the association between self-deception and performance does suggest that differences in affective or motivational response to the task are unlikely to account for the results. In future studies, it may be possible to take measures of affective response (e.g., electrodermal activity) that could provide more concrete evidence that self-deceivers have an unimpaired affective response to negative feedback in this task. Even without complete proof that the mechanism driving self-deceivers' perseveration involved ignoring an affective response to the losses, our results in these studies nonetheless demonstrate an association between self-deception and failure of response modulation.

5. Conclusion

The last decade saw the emergence of a formal debate over the utility of self-deception, following Taylor and Brown's (1988) review, in which they made the startling claim that self-deception was not only normative, but beneficial or even necessary in mentally healthy individuals. Taylor and colleagues have continued to argue that self-deceivers exhibit the markers of successful life adjustment (Taylor, 1989; Taylor & Brown, 1994) and that they remain flexible enough to respond to corrective information (Taylor, Collins, Skokan, & Aspinwall, 1989). The performance of self-deceivers on the Card Playing Task contributes an additional piece of evidence to the body of literature questioning, or even refuting, these assertions (e.g., Colvin & Block, 1994; Colvin, Block, & Funder, 1995; Johnson et al., 1997; Paulhus, 1998; Peterson et al., 2002). To perseverate in playing cards when one has begun steadily to lose money, or worse, to continue until all the money is gone, does not make one appear appropriately sensitive to corrective information. Rather, self-deception, in our view, entails precisely this sort of failure to respond to corrective information. If self-deceivers do not utilize indications that their current plans and beliefs are in error, as the present studies suggest, they seem likely to fail to learn and adapt to changes in the surrounding environment.

While it is not a logical necessity that perseveration on the Card Playing Task must be associated with maladaptive behavior outside of the lab (as pressing on in the face of adversity may sometimes be a good strategy in real-life situations), our findings place self-deceivers in the company of psychopaths (Belmore & Quinsey, 1994; Newman et al., 1987) and aggressive, conduct-disordered, or delinquent children (Fonseca & Yule, 1995; Séguin et al., 2002; Shapiro, Quay, Hogan, & Schwartz, 1988), who also display a deficit in performance on this task. Even if entirely different

mechanisms prove to be responsible for similar deficits in self-deception and syndromes of disinhibition like psychopathy and aggression, the similarity is nonetheless suggestive of the potential dangers of self-deception. Other researchers, furthermore, have found links between self-deception and poor outcomes in more ecologically valid circumstances (Colvin et al., 1995; Martocchio & Judge, 1997; Paulhus, 1998), and their findings lend additional credence to the idea that self-deceivers' performance on the Card Playing Task is likely to be indicative of a maladaptive aspect of self-deception.

Why should there be such a rift in opinion over whether self-deception is beneficial or detrimental? Perhaps emphasis on different time-scales of analysis is to blame. The mainstay of Taylor and colleagues' argument for the healthiness of self-deception is the consistent finding that measures of self-deception are negatively correlated with measures of negative affect, anxiety, neuroticism, and depression. In opposition, Colvin and Block (1994) have argued, cogently, that this evidence is flawed because self-deceptive individuals are unlikely to provide completely accurate answers to questions about their negative emotional states. We have argued, however, that self-deceivers actually *may* experience less negative affect, from day to day, because failing to attend to evidence of error allows them to avoid the stress, worry, and corrective effort that usually follows the discovery that one has made a mistake (Peterson et al., 2002). Nonetheless, we noted that self-deception must be deemed problematic when considered over any reasonable stretch of time because it vastly increases the likelihood that one's plans and beliefs will be poorly adapted to one's actual situation. Paulhus (1998) offered evidence supporting a discrepancy between short- and long-term outcomes of self-deception, with the finding that individuals scoring high on Self-Deceptive Enhancement made positive first impressions on unfamiliar partners in social interactions, but were evaluated negatively by the same partners after more extended interactions. To live blithely in a world that is not amenable to one's plans and beliefs is asking for trouble, and when that trouble comes it seems likely to induce negative affect ample to make up for its prior absence. The self-deceptive individual might be considered an exemplar of the Biblical saying, "Pride goeth before a fall." Freedom from negative affect in the short term is unlikely to be worth the long-term cost attendant on failure to adapt to one's environment.

Acknowledgments

Preparation of this article was made possible by a Knox Fund Grant from Harvard University and by support from the Social Sciences and Humanities Research Council of Canada and the National Health Research and Development Program of Canada. Support for Study 1 was provided by the Conseil Québécois de la Recherche Sociale, the Fonds pour la Formation des Chercheurs et l'Aide à la Recherche funds, and the National Science and Engineering Research Council. Portions of Study 1 were presented at the 107th Convention of the American Psychological Association, Boston, August 1999 and have been published in Journal of Child Psy-

chology and Psychiatry (Séguin et al., 2002). We wish to thank the boys, their families, the Commission des Écoles Catholiques de Montréal for their longstanding commitment to the longitudinal study, and Marc Lavoie and Julie Brousseau for data collection.

References

Belmore, M. F., & Quinsey, V. L. (1994). Correlates of psychopathy in a noninstitutional sample. *Journal of Interpersonal Violence, 9*, 339–349.

Blanchard, D. J., & Blanchard, D. C. (1989). Antipredator defensive behaviors in a visible burrow system. *Journal of Comparative Psychology, 103*, 70–82.

Borkenau, P., & Amelang, M. (1985). The control of social desirability in personality inventories: A study using the principal-factor deletion technique. *Journal of Research in Personality, 19*, 44–53.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality—social, clinical, and health psychology. *Psychological Bulletin, 92*, 111–135.

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin, 116*, 3–20.

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68*, 1152–1162.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason and the human brain*. New York: Avon Books.

Depue, R. A., & Collins, P. F. (2000). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences, 22*, 491–569.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences, 33*, 533–552.

Dollard, J. C., & Miller, N. E. (1950). *Personality and Psychotherapy*. New York: McGraw-Hill.

Eysenck, H. J., & Eysenck, S. B. (1975). *Manual of the eysenck personality questionnaire*. London: Hodder & Stoughton.

Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences, 6*, 121–129.

Fonseca, A. C., & Yule, W. (1995). Personality and antisocial behavior in children and adolescents: An enquiry into Eysenck's and Gray's theories. *Journal of Abnormal Child Psychology, 23*, 767–781.

Fossum, T. A., & Barrett, L. F. (2000). Distinguishing evaluation from description in the personality–emotion relationship. *Personality and Social Psychology Bulletin, 26*, 669–678.

Giancola, P. R., Peterson, J. B., & Pihl, R. O. (1993). Risk for alcoholism, antisocial behavior, and response perseveration. *Journal of Clinical Psychology, 49*, 423–428.

Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. Oxford: Oxford University Press.

Gray, J. A. (1987). *The psychology of fear and stress* (2nd ed.). Cambridge: Cambridge University Press.

Greenwald, A. G. (1988). Self-knowledge and self-deception. In J. S. Lockard & D. L. Paulhus (Eds.), *Self-deception: An adaptive mechanism?* (pp. 113–131). Englewood Cliffs, NJ: Prentice Hall.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581–595.

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press.

Johnson, E. A., Vincent, N., & Ross, L. (1997). Self-deception versus self-esteem in buffering the negative effects of failure. *Journal of Research in Personality, 31*, 385–405.

Kline, J. P., Schwartz, G. E., Allen, J. J. B., & Dikman, Z. V. (1998). Perceptual and electroencephalographic registration of masked emotional words in defensiveness: An exploratory study. *Personality and Individual Differences, 24*, 499–512.

Lajunen, T., Corry, A., Summala, H., & Hartley, L. (1997). Impression management and self-deception in traffic behavior inventories. *Personality and Individual Differences, 22*, 341–353.

Martocchio, J. J., & Judge, T. A. (1997). Relationship between conscientiousness and learning in employee training: Mediating influence of self-deception and self-efficacy. *Journal of Applied Psychology, 82*, 764–773.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882–888.

Mele, A. (1997). Real self-deception. *Behavioral & Brain Sciences, 20*, 91–136.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Adams, Bannister, Cox.

Nagin, D. S., & Tremblay, R. E. (2001). Parental and early childhood predictors of persistent physical aggression in boys from kindergarten to high school. *Archives of General Psychiatry, 58*, 389–394.

Newman, J. P., & Lorenz, A. (2001). Response modulation and emotion processing: Implications for psychopathy and other dysregulatory psychopathology. In R. J. Davidson, K. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences*. Oxford: Oxford University Press.

Newman, J. P., Patterson, C. M., & Kosson, D. S. (1987). Response perseveration in psychopaths. *Journal of Abnormal Psychology, 96*, 145–148.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotion. *Cognition and Emotion, 1*, 29–50.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679.

Panksepp, J. (1999). *Affective neuroscience*. Oxford: Oxford University Press.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. Wrightsman (Eds.), *Measures of personality and social–psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*, 1197–1208.

Paulhus, D. L. (1999). BIDR Self-Deceptive Denial (SDD). World Wide Web: http://www.psych.ubc.ca/~dpaulhus/del/research/sdd.htm.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1025–1060.

Peterson, J. B. (1999). *Maps of meaning: The architecture of belief*. New York: Routledge.

Peterson, J. B., Driver-Linn, E., & DeYoung, C. G. (2002). Self-deception and impaired categorization of anomaly. *Personality and Individual Differences, 33*, 327–340.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78*, 582–593.

Raskin, R., Novacek, J., & Hogan, R. (1991). Narcissism, self-esteem, and defensive self-enhancement. *Journal of Personality, 59*, 19–38.

Sackeim, H. A., & Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation, advances in research and theory* (Vol. 2, pp. 139–197). New York: Plenum Press.

Sackeim, H. A., & Gur, R. C. (1979). Self-deception, other deception, and self-reported psychopathology. *Journal of Consulting and Clinical Psychology, 47*, 213–215.

Séguin, J. R., Arseneault, L., Boulerice, B., Harden, P. W., & Tremblay, R. E. (2002). Response perseveration in adolescent boys with stable and unstable histories of physical aggression: The role of underlying processes. *Journal of Child Psychology and Psychiatry*.

Séguin, J. R., Pihl, R. O., Harden, P. W., Tremblay, R. E., & Boulerice, B. (1995). Cognitive and neuropsychological characteristics of physically aggressive boys. *Journal of Abnormal Psychology, 104*, 614–624.

Shapiro, S. K., Quay, H. C., Hogan, A. E., & Schwartz, K. P. (1988). Response perseveration and delayed responding in undersocialized aggressive conduct disorder. *Journal of Abnormal Psychology, 97*, 371–373.

Taylor, S. E. (1989). *Positive illusions: Creative self-deception and the healthy mind*. New York: Basic Books.

Taylor, S. E., & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193–210.

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*, 21–27.

Taylor, S. E., Collins, R. L., Skokan, L. A., & Aspinwall, L. G. (1989). Maintaining positive illusions in the face of negative information: Getting the facts without letting them get to you. *Journal of Social and Clinical Psychology, 8*, 114–129.

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivée, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology, 19*, 285–300.

Tremblay, R. E., Pihl, R. O., Vitaro, F., & Dobkin, P. L. (1994). Predicting early onset of male antisocial behavior from preschool behavior. *Archives of General Psychiatry, 51*, 732–739.