PSYCHOLOGICAL SCIENCE

Research Article

Why Good Guys Wear White

Automatic Inferences About Stimulus Valence Based on Brightness

Brian P. Meier,[1] Michael D. Robinson,[1] and Gerald L. Clore[2]

[1]North Dakota State University and [2]University of Virginia

ABSTRACT—*Affect is a somewhat abstract concept that is frequently linked to physical metaphor. For example, good is often depicted as light (rather than dark), up (rather than down), and moving forward (rather than backward). The purpose of our studies was to examine whether the association between stimulus brightness and affect is optional or obligatory. In a series of three studies, participants categorized words as negative or positive. The valence of the words and the brightness of the letters were varied orthogonally. In Studies 1, 2, and 3, we found that categorization was inhibited when there was a mismatch between stimulus brightness (e.g., light) and word valence (e.g., negative). Studies 4 and 5 reveal boundary conditions for the effect. The studies suggest that, when making evaluations, people automatically assume that bright objects are good, whereas dark objects are bad.*

The association of brightness with affect is ubiquitous in popular culture. In American film, characters dressed in black are frequently evil, whereas those dressed in white are frequently good (e.g., in science fiction, *Star Wars*; in fantasy, *The Wizard of Oz*; in Westerns, Roy Rogers). More broadly, darkness is often associated with evil and death, whereas light is often associated with goodness and life. Such connections are very common in the Bible. For example, Jesus is the "light of the world," whereas Satan is the "prince of darkness." In the Hellenistic, Roman, and Christian traditions more generally, white is the color of joy, innocence, and purity, whereas black has opposite connotations (Eliade, 1996). For example, the prophet Zoroaster characterized the fight between good and evil as the fight between light and darkness, and Plato likened darkness to imprisonment and ignorance and light to freedom and knowledge. Although these connections might suggest that the pairing of brightness and affect is primarily a Western phenomenon, its scope appears to be broader. In Buddhist writings, truth is characterized as a light or a lamp, with seekers of truth shining brightly. In the Hindu Upanishads, light is equated with truth and immortality, whereas darkness is equated with delusion and death. And in the Koran, Allah is equated with light, and His message is a lamp and a star.[1] Indeed, Adams and Osgood (1973) found that observers from 20 countries tended to view light colors as good and dark colors as bad.

Why is affect, a relatively abstract concept, so often linked to relatively concrete qualities like brightness? One perspective on such perceptual-conceptual linkages comes from developmental research. According to Piaget and Inhelder (1969), in particular, cognitive development begins with sensorimotor representations. As children grow older, they develop the ability to think in more abstract terms. However, such abstract thought requires, and is built on, prior sensorimotor representations. More specific to affect is Johnson's (1997) theory of conflation, according to which aspects of the physical environment are paired with subjective experiences. For example, being held causes warmth (a sensory experience) as well as happiness (an affective experience). These early perceptual-conceptual pairings continue into adulthood, finding their most obvious manifestation in metaphor (e.g., a warm person is a pleasant person).

According to Lakoff and Johnson (1999), metaphors are more than convenient; they are essential. Lakoff and Johnson proposed the radical thesis that conceptual thought is always, or at least nearly always, built on physical metaphor (see also Barsalou, 1999; Gibbs, 1992). Indeed, they suggested that the neural networks responsible for abstract thought are intimately bound with the neural networks responsible for representing sensory experiences. If this is true, the ability to represent abstract experiences (e.g., affect) must be built on the ability to represent concrete experiences (e.g., brightness). In related research, participants have been shown to be faster to "approach" a stimulus if it is positive and "avoid" a stimulus if it is negative (Chen & Bargh, 1999; for an overview of related research, see Neumann & Strack, 2000). This interaction has suggested to some researchers that representations of the affect of stimuli automatically elicit inclinations to approach (in the case of positive stimuli) or avoid (in the case of negative stimuli).

Address correspondence to Brian P. Meier, Department of Psychology, North Dakota State University, Fargo, ND 58105; e-mail: brian.p.meier@ndsu.nodak.edu.

---

[1] We thank Louise Sundarajan for locating passages about light and dark from ancient texts, including the Buddhist texts Mahaparinibbana Sutta and Dhammapada, the Hindu Brihadaranyaka Upanishad, and the Muslim Koran.

AFFECT AND BRIGHTNESS

As we have stated, numerous metaphors are consistent with the association between affect and brightness. Empirical studies are also consistent with such associations. Stabler and Johnson (1972) found that children tended to associate good objects and positive self-statements with a white box; by contrast, they tended to associate bad objects and negative self-statements with a black box. In another study, Frank and Gilovich (1988) asked participants to rate pictures of professional football and hockey players dressed in their home uniforms. Players in black uniforms were rated as more malevolent than players in white uniforms. These studies are consistent with the idea that light objects are automatically assumed to be good, whereas dark objects are automatically assumed to be bad.

These studies do not establish that brightness-affect inferences are obligatory, however. Uncovering an obligatory relation between brightness and affect is important for supporting the idea that abstract thoughts are built on perceptual experiences. Lakoff and Johnson (1999) suggested that perceptual-conceptual links (e.g., between brightness and affect) are automated. According to this account, physical metaphor guides thought and behavior, but does so "like a 'hidden hand' that shapes how we conceptualize all aspects of our experiences" (p. 13). If metaphor is an important basis of conceptual thought, then physical cues should influence categorization performance even when irrelevant to the task at hand. The brightness (lightness vs. darkness) of a stimulus word should produce facilitation or interference depending on the valence (good vs. bad) of the word in question. In the current studies, we sought to determine if the brightness-affect inference is indeed obligatory.

OVERVIEW OF STUDIES

We used gray scale as the relevant chromatic dimension. With gray scale, light is manifested in the color white (100% gray scale), and dark is manifested in the color black (0% gray scale). Therefore, in all studies, stimuli were presented in black (dark) or white (light) on a light gray background (50% gray scale).

All studies used a 2 (valence: positive vs. negative) × 2 (color: black vs. white) repeated measures design. Positive and negative words were presented in a random order and randomly assigned to be displayed in black or white. Within Studies 1 through 3, we conducted original and replication substudies. In this sense, Studies 1 through 3 actually involved six participant samples. In these studies, participants' task was to evaluate words as positive or negative. In Study 4, we were interested in whether the association between brightness and affect is unidirectional or bidirectional. Accordingly, we presented the same stimuli, but asked participants to classify the font color of each word as black or white. In Study 5, we sought to determine if the same stimuli bias lexical decisions.

In Studies 1a and 1b, we emphasized both speed and accuracy, with particular emphasis on accuracy. We expected that negative words would be evaluated faster when presented in a black font than when presented in a white font; we expected the opposite pattern for positive words. In Studies 2a and 2b, we sought to determine if the same interaction characterizes accuracy when speed is emphasized. Finally, in Studies 3a and 3b, we forced participants to respond faster than they otherwise might to determine if the interaction between word valence and stimulus brightness occurs under conditions that require extremely quick responding (Draine & Greenwald, 1998).

Studies 4 and 5 were conducted to determine the scope of the Valence × Color interaction. We suspected that the interaction would be asymmetric, such that color would interact with valence judgments (Studies 1, 2, and 3), whereas valence would not interact with color judgments (Study 4). Such an asymmetric pattern would be consistent with the idea that in the race between physical cues and valence, physical cues win. In Study 5, participants performed the lexical decision task with the same stimuli as in the prior studies. We did not expect to find the Valence × Color interaction in Study 5. Such null results would suggest that this interaction involves stimulus-response compatibility mechanisms (e.g., from an irrelevant light color to a positive response tendency) rather than stimulus-stimulus compatibility mechanisms (e.g., from an irrelevant light color to a positive word like *angel*). The lexical decision task (Study 5) and the evaluation task (Studies 1, 2, and 3) shared the same stimulus-stimulus manipulations. The only difference between the tasks was the involvement of an irrelevant stimulus-response compatibility mechanism in the evaluation task but not the lexical decision task (De Houwer, 2003; Kornblum, Hasbroucq, & Osman, 1990).

To summarize, all the studies involved positive or negative words presented in a black or white font. Studies 1, 2, and 3 used an evaluation task. In Study 4, participants categorized font color rather than the valence of the words. In Study 5, participants categorized each stimulus as a word or nonword.

METHOD

Participants

Participants were 169 undergraduates (18 in Study 1a, 21 in Study 1b, 23 in Study 2a, 15 in Study 2b, 26 in Study 3a, 19 in Study 3b, 22 in Study 4, and 25 in Study 5) who received extra course credit.[2]

Stimuli and Procedures

We selected 100 words, half with positive meanings (e.g., *gentle*) and half with negative meanings (e.g., *devil*).[3] The number of letters was similar for positive and negative words, $F < 1$. To validate our a priori classification of words as positive or negative, we asked 7 participants to rate the valence of each word (1 = *extremely negative*, 5 = *neutral*, 9 = *extremely positive*). With word as the unit of analysis, the positive words ($M = 7.46$) were rated as significantly more positive than the negative words ($M = 2.42$), $F(1, 98) = 1{,}040.44$, $p = .000$. We also sought to determine whether positive and negative words were equal

---

[2] Given that we used unselected participants and that our participant pool was 95% Caucasian, we did not have adequate power to test for possible ethnic differences. Doing so in future research is recommended.

[3] Following are the positive words used in the studies: active, agile, ambitious, baby, brave, candy, champion, clean, cordially, devotion, dream, earnest, ethical, faith, festival, garden, generous, genius, gentle, gracious, heaven, hero, justice, kiss, leisure, love, loyal, mature, mercy, neat, nurse, polite, power, pretty, prompt, radiant, reliable, righteous, satisfying, sensible, sincere, sleep, studious, sweet, talented, trust, truthful, victory, wise, and witty. Following are the negative words used in the studies: aimless, argue, beggar, bitter, cancer, cheat, clumsy, crime, critical, crooked, crude, cruel, danger, dead, defeat, delay, devil, diseased, divorce, enemy, fickle, foolish, fraud, greedy, hostile, insane, insolent, liar, mediocre, mosquito, nasty, neurotic, obnoxious, poison, pompous, profane, rude, sarcastic, shallow, sloppy, sour, spider, steal, stingy, theft, touchy, ugly, unfair, vain, and vulgar.

in extremity. To determine this, we computed the absolute difference between the valence rating of each word and the neutral midpoint. The average difference scores for positive and negative words were equal, $F < 1$.

In most of the studies, the words were presented once each, for a total of 100 trials. In Studies 1a, 2b, and 3b, however, the words were presented twice each, once in white and once in black. In none of the latter studies did the repetition factor interact with valence or font color, *F*s < 1. For this reason, we collapsed across the repetition factor in reporting the results of these studies.

In Studies 1, 2, and 3, participants were told that they would see a word presented on the computer screen along with the category labels "NEGATIVE" and "POSITIVE." Their task was to determine whether the word had a negative ("1" key) or positive ("9" key) meaning. In Study 1, we asked participants to be both quick and accurate. However, we gave feedback related to accuracy, but not speed. Because we expected a high degree of accuracy in this study, our main interest was in response latencies. In Study 2, we emphasized speed over accuracy. Specifically, we told participants that speed was crucial whereas accuracy was less important. In this study, we gave feedback concerning speed, but not accuracy. Because we expected a high degree of speed in this study, our main interest was in the accuracy of their responses. In Study 3, we used a response-deadline procedure that required very quick responses (within 600 ms in Study 3a and 700 ms in Study 3b). If participants did not make a response during this interval, they were punished with a visual message telling them they were "TOO SLOW!" Because of the procedure in this study, our main interest was in accuracy. Studies 1, 2, and 3, therefore, systematically varied whether accuracy (Study 1) or speed (Studies 2 and 3) was the more important criterion, thereby controlling possible speed-accuracy trade-offs.

In Study 4, participants were told that they would see a word presented on the computer screen along with the category labels "BLACK" and "WHITE." Their task was to determine whether the word was presented in black ("1" key) or white ("9" key). They were instructed to make these classifications as quickly and accurately as possible. A computer-generated beep was presented for 300 ms after each inaccurate response. In Study 5, participants were told to decide quickly and accurately whether each letter string was a word ("9" key) or not ("1" key). In this study, we added 100 nonword letter strings, also presented in black or white. A visual error message was presented after each inaccurate response.

RESULTS

In all the studies, response latencies were analyzed in the following manner. First, trials with inaccurate responses were dropped. Second, we replaced times that were more than 2.5 *SD*s below or above the overall latency mean (i.e., across trials and participants) with the 2.5-*SD* value. Analyses involved these raw reaction times. In all the studies, accuracy rates were arcsine-transformed prior to analysis. This transformation stabilizes the variance of accuracy rates, rendering them more appropriate for analysis of variance (Bartlett, 1947). However, the means we report here were calculated from untransformed accuracy rates. All studies had a 2 (valence: positive vs. negative) × 2 (color: black vs. white) repeated measures design. Parallel analyses were performed on latencies and accuracy rates. *F* values, means, and standard deviations are presented in Table 1 (latencies) and Table 2 (accuracy rates). Main effects (for valence and color) were inconsistent across studies. For this reason, we do not discuss them further.

TABLE 1

Mean Latencies and Results of Analysis of Variance

	Analysis of variance			Mean (milliseconds)			
Study	Valence main effect	Color main effect	Color × Valence interaction	White font, negative valence	White font, positive valence	Black font, negative valence	Black font, positive valence
1a ($n = 18$)	$F = 30.04$	$F = 0.93$	$F = 7.78$	927	844	896	859
	$p = .000$	$p = .349$	$p = .013$	(111)	(107)	(125)	(109)
1b ($n = 21$)	$F = 0.68$	$F = 1.07$	$F = 4.58$	897	856	881	896
	$p = .420$	$p = .313$	$p = .045$	(181)	(185)	(202)	(208)
2a ($n = 23$)	$F = 1.16$	$F = 0.03$	$F = 0.46$	615	614	622	605
	$p = .292$	$p = .858$	$p = .505$	(92)	(103)	(106)	(97)
2b ($n = 15$)	$F = 27.58$	$F = 2.25$	$F = 0.09$	855	766	844	755
	$p = .000$	$p = .156$	$p = .769$	(191)	(163)	(171)	(181)
3a ($n = 26$)	$F = 5.96$	$F = 0.57$	$F = 0.43$	467	453	459	449
	$p = .022$	$p = .456$	$p = .518$	(54)	(41)	(41)	(46)
3b ($n = 19$)	$F = 4.07$	$F = 2.88$	$F = 0.00$	514	497	508	492
	$p = .059$	$p = .107$	$p = .981$	(73)	(63)	(66)	(66)
4 ($n = 22$)	$F = 0.74$	$F = 0.72$	$F = 0.32$	508	497	517	513
	$p = .399$	$p = .406$	$p = .577$	(90)	(92)	(99)	(105)
5 ($n = 25$)	$F = 32.96$	$F = 0.00$	$F = 0.53$	805	770	812	772
	$p = .000$	$p = .975$	$p = .472$	(161)	(141)	(163)	(171)

Note. Standard deviations are in parentheses.

TABLE 2

Mean Accuracy Rates and Results of Analysis of Variance

	Analysis of variance			Mean			
Study	Valence main effect	Color main effect	Color × Valence interaction	White font, negative valence	White font, positive valence	Black font, negative valence	Black font, positive valence
1a ($n = 18$)	$F = 0.03$	$F = 2.82$	$F = 12.78$	.938	.954	.966	.946
	$p = .862$	$p = .112$	$p = .002$	(.053)	(.034)	(.030)	(.036)
1b ($n = 21$)	$F = 0.77$	$F = 0.18$	$F = 0.45$	.980	.979	.981	.971
	$p = .392$	$p = .669$	$p = .508$	(.030)	(.028)	(.026)	(.041)
2a ($n = 23$)	$F = 1.44$	$F = 1.88$	$F = 10.79$	.857	.890	.890	.831
	$p = .242$	$p = .184$	$p = .003$	(.104)	(.087)	(.102)	(.107)
2b ($n = 15$)	$F = 13.91$	$F = 0.15$	$F = 5.34$	.831	.923	.860	.900
	$p = .002$	$p = .706$	$p = .037$	(.065)	(.049)	(.065)	(.061)
3a ($n = 26$)	$F = 9.92$	$F = 1.67$	$F = 4.85$	.498	.610	.558	.593
	$p = .044$	$p = .208$	$p = .037$	(.173)	(.223)	(.207)	(.196)
3b ($n = 19$)	$F = 6.94$	$F = 2.01$	$F = 4.42$	.521	.647	.592	.609
	$p = .017$	$p = .173$	$p = .050$	(.104)	(.122)	(.081)	(.150)
4 ($n = 22$)	$F = 0.25$	$F = 0.06$	$F = 0.09$	.976	.979	.979	.979
	$p = .626$	$p = .804$	$p = .768$	(.035)	(.025)	(.032)	(.025)
5 ($n = 25$)	$F = 4.99$	$F = 0.33$	$F = 1.14$	.937	.963	.953	.958
	$p = .035$	$p = .569$	$p = .297$	(.075)	(.036)	(.049)	(.042)

Note. Standard deviations are in parentheses.

In Study 1, we emphasized accuracy. Therefore, we expected effects to show up primarily in latencies. As expected, the Valence × Color interaction was significant in Study 1a and Study 1b (see Table 1). In both substudies, participants were faster to categorize positive words when presented in a white (vs. black) font, whereas they were faster to categorize negative words when presented in a black (vs. white) font (see Table 1). In Study 1a, there was also a parallel interaction involving accuracy rates (see Table 2): Participants were more accurate in categorizing positive words when presented in a white font than when presented in a black font; by contrast, participants were more accurate in categorizing negative words when presented in a black font than when presented in a white font. This interaction involving accuracy rates did not occur in Study 1b (see Table 2).

In Study 2, we emphasized speed. Therefore, we expected effects to show up primarily in accuracy rates. As expected, the Valence × Color interaction was significant in Study 2a and Study 2b (see Table 2): Participants were more accurate evaluating negative words if they were presented in a black (vs. white) font, but were more accurate evaluating positive words if they were presented in a white (vs. black) font. In neither study was there a Valence × Color interaction involving latencies (see Table 1).

In Study 3, we forced participants to respond quickly. Therefore, we expected effects to show up primarily in accuracy rates. As expected, the Valence × Color interaction was significant in Study 3a and Study 3b (see Table 2): Participants were more accurate evaluating negative words if they were presented in a black (vs. white) font, but were more accurate evaluating positive words if they were presented in a white (vs. black) font.

In Studies 4 and 5, we obtained no Valence × Color interactions, either for latencies (Table 1) or for accuracy rates (Table 2).

DISCUSSION

By manipulating stimulus valence and stimulus color independently, we sought to determine whether the brightness of an object automatically suggests its valence. Across three studies involving six participant samples, we found that a mismatch (vs. match) between stimulus color and valence interfered with affective categorization. When we emphasized accuracy (Study 1), the interaction affected mean latency. When we emphasized speed (Studies 2 and 3), the interaction affected mean accuracy. The fact that the interaction occurred within the response-deadline procedure (Study 3), in particular, points to the obligatory nature of affective inferences based on stimulus brightness.

We hypothesized that affective inferences are influenced by physical cues, but not vice versa, so in Study 4 we asked participants to judge the font color (rather than valence) of each word. Because there was no hint of an interaction in the results of this study, they confirm that the association of brightness and valence is asymmetric. This is what one would expect given a race model in which physical cues (i.e., font color) are available before stimulus valence is. Finally, Study 5 demonstrated the task-dependent nature of the interaction. Font color influenced categorization times for affective judgments (Study 1), but not for lexical decision judgments (Study 5). The lack of an interaction in Study 5 suggests that the interaction occurs because colors prime response tendencies (dark = bad, light = good) that are apparent only when the task at hand involves affective categorization.

The results of our studies (particularly Study 5) suggest that the association between brightness and affect is Stroop-like. If so, it involves response-related mechanisms rather than spreading activation within long-term memory. As both the Stroop (MacLeod, 1991) and the flanker (Zhang, Zhang, & Kornblum, 1999) effects seem to be due to response conflict, we view it as likely that the interaction we observed is due to this factor as well. For example, a negative word presented in white would give rise to two response tendencies, one to respond "negative" (on the basis of stimulus valence) and one to respond "positive" (on the basis of stimulus color). The response would be delayed because of these co-occurring, but conflicting, response tendencies (Zhang et al., 1999). In the traditional Stroop task (as well as the traditional flanker task), however, the irrelevant stimulus feature (word meaning) is directly relevant to the response. For example, the word "yellow" also represents one of the responses. In our studies, by contrast, the irrelevant stimulus feature (font color) was not directly related to one of the responses ("negative" or "positive"). The fact that stimulus color still had an influence under these circumstances is impressive evidence for the automaticity of the color-to-valence translation.

The Physicality of Conception

It is one thing to propose that physical metaphors (e.g., light = good) are helpful in conceptualizing affect. It is another thing to propose that such associations are obligatory. Our findings suggest that people cannot conceptualize the affect of a stimulus without considering its physical features (e.g., color). In the studies reported here, stimulus color affected participants' responses despite the fact that stimulus color was totally irrelevant to the task at hand, which was to categorize words as negative or positive. Our results are therefore consistent with the ideas of Lakoff and Johnson (1999), who proposed that conception is built on physical metaphor (see also Barsalou, 1999; Gibbs, 1992). This said, we recognize that the data do not provide support for the strongest claims concerning the metaphorical basis of thought (i.e., that metaphors are required for abstract thought). It may be that metaphors are convenient, but not absolutely necessary, in structuring abstract thought (Glucksberg, 2001).

More broadly, it may be the case that amodal (i.e., purely abstract) representation is not possible (Barsalou, 1999). Rather, concepts may be "grounded" in physical metaphor and sensorimotor representation. Such an idea fits with what is known about the evolution of the human brain. Humans' most abstract representational abilities were added to earlier parts of the brain that are much more stimulus-bound in representation (MacLean, 1990). Within human development, too, the ability to represent concepts abstractly comes after the ability to represent objects in a sensorimotor manner (Piaget & Inhelder, 1969). The asymmetry of this sequence suggests that conception may have a perceptual basis (Barsalou, 1999; Lakoff & Johnson, 1999).

From a more applied perspective, our findings may have implications for racial stereotyping. If there really is an automatic tendency to relate stimulus color to stimulus valence, then people who are dark skinned may be at a disadvantage in interpersonal relations. In fact, recent work has shown an automatic association between skin color (e.g., Caucasian American vs. African American) and valence (e.g., Dasgupta, McGhee, Greenwald, & Banaji, 2000; Wittenbrink, Judd, & Park, 2001), at least for the samples studied. Our work suggests that the association between color and valence may be the more universal tendency (see also Adams & Osgood, 1973, in this regard), with prejudice perhaps a manifestation of this tendency. Consistent with this idea, past research has shown that people associate dark boxes (Stabler & Johnson, 1972) or uniforms (Frank & Gilovich, 1988) with negative connotation. Our studies reveal the obligatory nature of this connection.

Conclusion

Our results support the existence of an automatic association between brightness and affect. The association suggests that in making affect judgments, people attend to the physical features of an object in such a way that irrelevant physical metaphors (e.g., related to color) bias their responses. The findings are consistent with the idea that conceptual thought is embodied in nature (Barsalou, 1999; Lakoff & Johnson, 1999). It would be fruitful to see if other metaphors for affect (e.g., up = good, more = good, and close = good) bias affective judgments in a similar manner.

Acknowledgments—Research reported in this article was supported by funding from the National Science Foundation (Robinson: SBR 981764) and National Institute of Mental Health (Clore: MH 50074).

REFERENCES

Adams, F.M., & Osgood, C.E. (1973). A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology*, *4*, 135–156.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Bartlett, M.S. (1947). The use of transformations. *Biometrics*, *3*, 39–52.

Chen, M., & Bargh, J.A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*, 215–224.

Dasgupta, N., McGhee, D.E., Greenwald, A.G., & Banaji, M.R. (2000). Automatic preference for white Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*, 316–328.

De Houwer, J. (2003). A structural analysis of indirect measure of attitudes. In J. Musch & C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219–244). Mahwah, NJ: Erlbaum.

Draine, S.C., & Greenwald, A.G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*, 286–303.

Eliade, M. (1996). *Patterns in comparative religion*. Lincoln: University of Nebraska Press.

Frank, M.G., & Gilovich, T. (1988). The dark side of self- and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, *54*, 74–85.

Gibbs, R.W., Jr. (1992). Categorization and metaphor understanding. *Psychological Review*, *99*, 572–577.

Glucksberg, S. (2001). *Understanding figurative language*. New York: Oxford University Press.

Johnson, C. (1997). Learnability in the acquisition of multiple senses: Source reconsidered. In J. Moxley, J. Juge, & M. Juge (Eds.), *Proceedings of the Twenty-Second Annual Meeting of the Berkeley Linguistics Society* (pp. 469–480). Berkeley, CA: Berkeley Linguistics Society.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility: A model and taxonomy. *Psychological Review*, *97*, 253–270.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenges to western thought*. New York: Basic Books.

MacLean, P.D. (1990). *The triune brain in evolution: Role in paleocerebral functions*. New York: Plenum Publishing.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203.

Neumann, R., & Strack, F. (2000). Approach and avoidance: The influence of proprioceptive and exteroceptive cues on encoding of affective information. *Journal of Personality and Social Psychology*, *79*, 39–48.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books.

Stabler, J.R., & Johnson, E.E. (1972). The meaning of black and white to children. *International Journal of Symbology*, *3*, 11–21.

Wittenbrink, B., Judd, C.M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815–827.

Zhang, H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus-stimulus and stimulus-response compatibility. *Cognitive Psychology*, *38*, 386–432.

(Received 10/30/02; Revision accepted 2/25/03)