The Yale Review of Undergraduate Research in Psychology

A Student Publication

Second Issue

About

The Yale Review of Undergraduate Research in Psychology is an annual journal that showcases the best and most original research in psychology conducted by undergraduates from around the world. Our goal is to contribute to the scientific advance by encouraging serious, quality research early on. We provide a platform for undergraduate scientists to share their findings, and aim to bring together a community of young psychologists from both the United States and abroad.

Acknowledgements

We extend our deepest gratitude to the following individuals, departments, and organizations, without whom this journal could not have been possible: Marcia Johnson, Laurie Santos, and the Yale Psychology Department; Tamar Gendler, Brian Scholl, and the Yale Cognitive Science Department; John Meeske and The Yale Undergraduate Organizations Funding Committee; Edgar Letriz and the Yale College Dean's Office; and Professor Peggy Clark.

The YRURP Staff

Staff

Contents

Incidental Exposure to No-Smoking Signs Primes Craving for Cigarettes: An Ironic Effect of Unconscious Semantic Processing?

Brian D. Earp, Brendan Dill, Jennifer L. Harris, Joshua M. Ackerman, & John A. Bargh

YALE UNIVERSITY

The present study tests whether incidental exposure to no-smoking signs may ironically boost craving for cigarettes in smokers. Smokers viewed photographs in which no-smoking signs were either inconspicuously embedded (prime) or edited out (control). Participants then used a joystick to make quick approach vs. avoid motions while viewing smoking-related and neutral stimuli on a computer screen (Chen & Bargh, 1999). We hypothesized that primed smokers, but not controls, would show an automatic reach bias toward the smoking-related stimuli. The data supported our prediction. Possible mechanisms for the effect are discussed, as well as implications for public health policy, negation-based social campaigns in general, and our understanding of the unconscious processing of semantic information.

INTRODUCTION

No-smoking signs are ubiquitous. Part of a wide-reaching social, legal, and public health movement meant to discourage smoking in public places, they can be found nearly everywhere people are: train stations and bus stops, book stores and coffee shops, restaurants, airplanes, schoolyards—and even some taverns and bars (Quenda, 2009). These typically unassuming icons are now so commonplace that it is easy to take them for granted. Fading into the background, they become just another part of our day-to-day environment.

This does not mean that they do not affect us, however. Decades of research on automatic priming effects have confirmed that our everyday physical environment is a rich source of influence on our unconscious minds, shaping everything from our mood to our social judgments to our very behavior, and often in surprising ways (cf. Bargh & Chartrand, 1999). Rain outside our window can make us feel dissatisfied with our whole lives (Schwarz & Clore, 1983). The mere presence of a briefcase in the room can boost our economic drive (Kay, Wheeler, Bargh, & Ross, 2004). Holding a warm cup of coffee makes us likelier to trust a stranger (Williams & Bargh, 2008a). Washing hands may assuage guilt (Zhong & Liljenquist, 2006). And the relative placement of dots on a grid may alter how distressed we are by violence on TV (Williams & Bargh, 2008b). In each of these cases, the physical stimulus from the immediate environment exerts its influence outside of our conscious awareness or control.

In addition to perceiving objects like a briefcase, or the tactile experience of holding a coffee cup, exposure to semantic content—usually in the form of words—can affect us subconsciously as well. For example, covertly exposing study participants to words that are stereotypic of the elderly causes them to walk slower and perform less well on memory tasks (Bargh, Chen, & Burrows, 1996; Dijksterhuis, Bargh, & Miedema, 2000). Priming the stereotype of a professor boosts performance on a general knowledge test, while semantically activating the idea of a soccer hooligan impairs performance (Dijksterhuis & Van Knippenberg, 1998). And administering a modified Stroop task with words related to drunkenness causes undergraduates to quaff more beer (Roehrich & Goldman, 1995). Many of these results can be explained by a direct and pervading link between perception—including unconscious semantic perception—and behavior: monkey see, monkey do (cf. Bargh et al., 1996).

But what does this tell us about no-

This research was funded by Grant RO1-MH60767 from the National Institute of Mental Health to JAB. Thanks are due to Molly Priedeman for research assistance. This research is an undergraduate thesis submitted in 2010 by the first author in partial fulfillment of the degree of B.A. in Cognitive Science, Yale University. The first author can be reached at brian.earp@gmail.com or at Oxford Uehiro Centre for Practical Ethics, Suite 8, Littlegate House, St. Ebbes St., Oxford OX1 1PT.

smoking signs? Since such signs are background fixtures of many public environments as well as physical objects with semantic content, they are just the sort of item one could expect to exert an influence on the unconscious mind. Drawing on the priming literature just cited, as well as research into the nature of unconscious semantic integration (discussed below), we hypothesized that incidental exposure to these signs will ironically *increase* smoking behavior in nicotine-addicted individuals. To understand why, we need to consider the seeming inability of the unconscious mind to factor *negation* into its semantic analyses (cf. Greenwald & Liu, 1985).

The Problem of Negation

In 1925, Sigmund Freud published an article entitled simply "Negation" (Freud, 1925). As an opening illustration, he asks his reader to imagine an exchange with a patient:

> "You ask who this person in the dream can have been. It was *not* my mother." We emend this: so it was his mother. In our interpretation we take the liberty of disregarding the negation and of simply picking out the subject-matter of the association. It is just as though the patient had said: "It is true that I thought of my mother in connection with this person, but I don't feel at all inclined to allow the association to count" (p. 367).

Freud's thesis, then, is that negation is not a faculty of the unconscious mind, but rather a tool of repression and inhibition exerted over it by the ego. "We never discover a 'No' in the unconscious," he writes (p. 371). If Freud is right, this fact would have obvious implications for our question about no-smoking signs. Specifically, it might mean that the unconscious mind would treat "*no* smoking"' on a par with its opposite—"smoking"—in much the same way that "It was *not* my mother" means (according to Freud) that it almost certainly was.

Modern psychological research seems to supports this view—i.e., that the unconscious mind does indeed show a deficit in handling negation. Deutsch, Gawronski, and Strack (2006), for instance, used an evaluative priming task to show that negated, positively-valenced prime words had the same effect on target judgments as affirmed primes: in both cases, the target evaluation showed a positive-valence bias. As far as the unconscious mind is concerned, then, "*not* good" means the same as its opposite; the negation is effectively dropped. Gawronski, Deutsch, Mbirkou, Seibt, and Strack (2008) extended these findings to anti-stereotype training regimens, showing that the desired outcome—namely a fully automatized positive stereotype in place of an original negative one—was attainable only through repeated affirmations of the counterstereotype, and not by constant negation of the original. The latter method perversely yielded the opposite effect. That is, repeated negations of a negative stereotype serve only to facilitate stereotype-consistent judgments.

These findings are consistent with a dual-process theory of mind according to which unconscious reasoning systems are computationally associative—and therefore incapable of rule-based manipulations like negation—whereas it is the *conscious* mind that is rule-based and systematic and thus able to inhibit or repress automatic associational outputs (e.g., Sloman, 1996). But even the conscious mind struggles with negation. In a classic demonstration, Wason (1959) showed that participants are slower and more error-prone when performing logical operations involving negation compared to

operations that are meaning-identical but framed affirmatively. Gilbert (1991) provides evidence that while "the acceptance of an idea is part of the automatic comprehension of that idea," the conscious negation of an idea "occurs subsequent to, and more effortfully than, its acceptance" (p. 107). And Fyodor Dostoevsky (1863/1997), as penetrating a psychologist as any, gives the following case study: "Try to pose for yourself this task," he writes in *Winter Notes on Summer Impressions*, "*not* to think of a polar bear—and you will see that the cursed thing will come to mind every minute" (p. 49, emphasis added). That is, even when we consciously *try* to negate, we often fail. The relevant concept, once introduced by the negation-based prompt, blithely remains activated despite our best efforts. This paradoxical nature of conscious thought suppression has been thoroughly confirmed in numerous laboratory settings over the past thirty years (for a review, see Wenszlaff & Wegner, 2000).

Both conscious and unconscious mental systems, then, seem to treat negations differently from affirmations—processing them more reluctantly in the first case or not at all in the second. Accordingly, we have some clear parameters for the sort of prediction we should make about the case at hand, namely the likely effect of incidental exposure to no-smoking signs in a typical public environment. Either the signs will be perceived unconsciously, in which case the "no" in "no-smoking" should be totally discounted; or else they will be processed with some degree of conscious awareness, in which case the negation will be handled more or less effectively depending upon the degree of awareness and the available cognitive resources. Either way, we should expect incidental exposure to the signs to induce (relatively) increased salience of the concept "smoking" (through relatively *decreased* processing of the "no"), and for individuals who are predisposed to evaluate smoking-related cues positively—i.e., habitual smokers—to exhibit corresponding evidence of boosted affect toward those cues.

THE PRESENT STUDY

The present study is our attempt to test this hypothesis. For the IV task, participants viewed a series of photographs on a computer screen, and were asked (as a cover) to judge whether the pictures were taken by a professional or amateur photographer. The photographs included images of everyday scenes in which no-smoking signs were either inconspicuously embedded (prime condition) or edited out (control condition). The "professional vs. amateur" distraction task was meant to conceal the nature of the prime as well as to preclude participants from focusing on any one aspect of the pictures, in particular the no-smoking signs themselves, to ensure that exposure was truly "incidental" and essentially consistent between participants.

For the DV task, participants used a joystick to make quick approach vs. avoid motions while viewing smoking-related and neutral stimuli on a computer screen. This task was adapted from Chen and Bargh (1999; see also Krieglmeyer, Deutsch, De Houwer, & De Raedt, 2010), who showed that moving a lever toward a stimulus is treated by the motor system as a "pushing away" or avoidance motion (participants are quicker to move the joystick in this direction in response to *negative* stimuli), while pulling a lever toward oneself activates muscle patterns associated with approach (i.e., pull-

ing the stimulus closer to oneself, which participants are quicker to do in response to *positive* stimuli). We hypothesized that smokers in the prime condition would be slower than controls to make avoid motions, and faster to make approach motions, in response to the smoking-related stimuli. Thus the joystick task functions as a craving or motivational measure without participants' awareness of what is being assessed.

Method

Participants

Participants were students and community members recruited on or near a college campus in the Northeastern United States. Recruitment consisted of a poster and email campaign asking, "Are you a smoker?" and which described the experiment as a "Health Attitudes" study. Compensation of $15 for approximately 20 minutes' participation was advertised. Of the total respondents, 20 men and 12 women ($n = 32$) met the pre-screen requirements and were invited to participate in the study. Participants were aged 18 to 49 years old ($M = 26.17$).

Materials and procedure

Prior to being invited to join the study, potential participants were sent an email in which they were given the following information:

> We're doing research on the health attitudes of different people, so we're looking at different groups: students, community members, smokers, non-smokers, people with different diet and exercise habits, and so on. If you're eligible for the study, you will come into the lab for just about 20-30 minutes and complete two brief computer tasks as well as fill out a short questionnaire. Compensation is $15.

The study was described in this way to minimize participants' ability to guess the particular theme or hypothesis of our research. That is, by framing participants' smoker identity as constituting just one of multiple demographic qualities of interest, we aimed to reduce vigilance for smoking cues on the day of actual study participation, with only minimal deception. After reading the above paragraph, potential participants filled out a brief health questionnaire, also by email. This questionnaire included an item asking about the recent (within 48 hours) consumption of a number of products, including our pre-screen item: a cigarette. Participants who indicated that they had consumed at least one cigarette within the past 48 hours were invited to join the study.

On the day of the study, participants were brought into the lab and seated at a large table outside the testing cubicle. Before signing the consent form, participants were once again given the "multiple demographics" and broad "health attitudes" cover story, but this time with absolutely no mention of the word "smoking." Again, this was meant to divert any suspicion that the study might have to do with smoking or craving behavior in particular, and to minimize the chance for any extraneous priming effects outside the IV task.

After the experimenter had delivered the cover story and collected the consent form, participants were led into an adjacent 3 × 3 meter testing room and seated in front of a 800 × 600 pixel Dell Plug and Play CRT Monitor running on a 2.8 GHz Pentium® 4 CPU. The monitor was positioned at the participant's eye level, approximately 60 cm from the face, and a standard computer key-

board was placed between the monitor and the participant.

Participants were then told that they would be shown a series of images, and that their task was to determine whether each image had been taken by a professional or an amateur photographer. Each image would appear for 1.5 seconds, and then the screen would show a prompt, at which point participants were to use the keyboard to press "p" for "professional" or "a" for "amateur." Participants were explicitly told that this first task was *not* a speed task, and that they should look at each image for the full 1.5 seconds and wait for the prompt before pressing any keys. The experimenter remained in the cubicle while the participant executed 5 practice trials on neutral images, confirmed that the instructions were fully understood, and then pressed a key to start the experimental trial before leaving the cubicle.

In the experimental trial, participants were shown 23 images of everyday scenes selected from a Google image search, one at a time, in randomized order (different for each participant). Images appeared on the screen for 1.5 seconds, after which time the screen refreshed to show the prompt. After participants registered their judgment, the next image appeared, again for 1.5 seconds, and so on through the rest of the images. In the prime condition, 13 of the 23 images included no-smoking signs. The control condition included all of the same images in randomized order, but with the no-smoking signs discretely edited out.

For the DV task, the experimenter brought out a Logitech® ATK3 gaming joystick that was connected to the computer's USB port, and instructed participants on its use. Participants were told the following:

> This task is different from the task you just did in a couple of ways. First of all, speed is really important – this is a reaction speed test, and your one goal is to react as quickly as you can. Various images will pop onto the screen, and your job is, as soon as you any image at all, whatever it is, to "knock it off the screen" by moving this joystick.

The DV task was divided into two blocks: a "forward" block and a "backward" block. Each participant completed both blocks, counterbalanced for order. Statistical analysis revealed no effect of order on the data (UNI-ANOVA of prime by order, $p = .802$), so this design feature will receive no further discussion. In the forward block, participants were instructed to move the lever *toward* the monitor in order to "knock the images" off the screen, which Chen and Bargh (1999) showed involves an implicit "pushing away" or avoid motion. In the backward block, participants were told to move the lever *away* from the monitor, an implicit "pulling toward" or approach motion. It is important to note that no actual mention was made of "pushing" or "pulling" the lever, as this might activate corresponding conscious motivations; only the goal-neutral words "forward" and "backward" were used in the task instructions.

After 5 practice trials (3 for the second block), the experimenter pressed a key to start the first block of experimental trials and left the cubicle. Participants were then shown a series of 25 stimuli, 21 of which were images of neutral objects like soccer balls and can openers, and 4 of which were "smoking-related" images like lit cigarettes. Each stimulus appeared after a variable interval ranging from 2 – 7 seconds, and disappeared in response to joystick movements in the appropriate direction depending on block. Response times for each image were

recorded by the computer. After completing the first block, participants were instructed to do the very same task once again, but to "knock the images off the screen" by moving the joystick in the opposite direction.

After completing both DV blocks, participants were led back into the main lab room and given a brief nicotine addiction questionnaire (FTND; Heatherton, Kozlowski, Frecker, & Fagerstrom, 1991). The experimenter then administered a careful funnel interview (Bargh & Chartrand, 2000) to test for awareness or suspicion of the hypothesis. Participants were then compensated and debriefed.

Results

Scores on the FTND ranged from 0 to 5 (M = 1.00, SD = 1.39) out of a possible 8 points, indicating relatively low levels of nicotine addiction among this group. Participants in the control and prime conditions did not differ in level of nicotine addiction (p = .81). Due to the positive skew in response time data, all response latencies were log transformed (Chen & Bargh, 1999), however, for ease of interpretation, we report original millisecond response times. For each participant, we subtracted the mean response time of approach trials from the mean response time of avoid trials for both smoking-related and neutral images. Because participants would be *faster* to make an approach motion in response to stimuli they are motivated to approach, and *slower* to make an avoid motion, a higher *avoid-minus-approach* score indicates a behavioral tendency to approach the stimulus. Conversely, a lower or negative avoid-minus-approach score indicates a less-pronounced approach tendency or a behavioral tendency to avoid the stimulus. We conducted an Analysis of Covariance (ANCOVA) with a between-subjects comparison of avoid-minus-approach scores for smoking stimuli (prime versus control condition). For this analysis we covaried avoid-minus-approach scores for neutral stimuli in order to control for individual differences in relative pushing versus pulling speed, and in chronic tendencies to approach or avoid stimuli in general.

Data from two participants were excluded from the final analysis: one for failing to follow task instructions, the other due to a computer malfunction that lost the participant's data. Overall, participants displayed a modest tendency to approach smoking-related stimuli (M = 8.6 ms, SD = 104) versus neutral stimuli (M = -14.0 ms, SD = 73), $t(29) = 1.63$, $p = .11$.

Results of the ANCOVA confirmed that exposure to no-smoking signs increased approach tendencies toward smoking-related stimuli, $F(1, 28) = 4.02$, $p = .055$, $\varphi = .13$. In addition, this approach tendency toward smoking-related stimuli was significantly higher after participants had been primed with no-smoking signs, $t(14) = 2.51$, $p = .03$, but not after neutral primes, $t(15) = .22$, $p = .83$ (see Figure 1). These findings support our hypothesis that incidental exposure to no-smoking signs increases automatic approach tendencies of smokers toward smoking-related stimuli.

Preliminary Discussion and Supplementary Analyses

We expected some variability in the level or type of priming (incidental exposure to the no-smoking signs) experienced by participants in the experimental condition. Since participants were free to cast their gaze over the IV photographs in whatever manner they saw fit, it is possible that some participants,

to a greater or lesser degree, consciously *noticed* the signs. Indeed, the funnel interview identified seven participants in the experimental condition who noticed the no-smoking signs in the photographs (but could not guess the experimental hypothesis).

Consistent with our earlier reasoning, we would expect that these participants, compared to those who did not notice the signs, would be more able to engage in conscious or rule-based processing and hence more able to account for the negation. This would predict that participants who noticed the signs would show *less* activation of the concept "smoking"—and therefore smaller avoid-minus-approach scores. To test for this, we ran another ANCOVA between-subjects analysis, controlling for responses to neutral stimuli as before. Surprisingly, there was no significant difference in these scores between participants who did and did not notice the signs, $F(1, 14) = .61$, $p = .45$.

Given our theoretical framework, there are at least two ways to account for this result. The first way has to do with the fact (noted earlier) that "even the conscious mind struggles with negation." On this account, "noticing" the no-smoking signs while engaged in a cognitively loading distracter task may involve insufficient processing resources, even conscious ones, to fully account for the "no" in "no-smoking." This would be an "up-stream" explanation saying in essence that the negation was never fully registered, even at the outset, and so failed to have an inhibitory effect on craving.

The second way has to do with the general nature of unconscious priming effects, namely that they can occur even if the stimulus is perceived, so long as the *effect* of the stimulus on behavior takes place outside of awareness (Bargh & Chartrand, 2000). Importantly, none of our participants reported awareness of the nature of the DV task, namely that it was designed to assess crav-

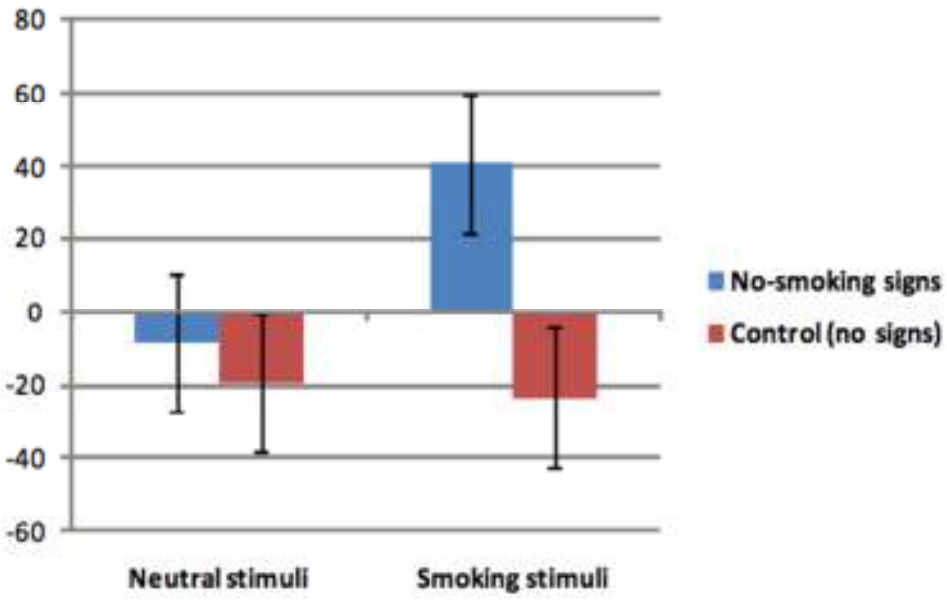


Figure 1. Differences in avoid-minus-approach scores for neutral versus smoking stimuli. Y-axis is response time in milliseconds.

ing, nor an understanding of the relationship between the IV and DV tasks. Thus it is possible that the "no" in "no-smoking" *was* fully registered by conscious mental systems, but failed to integrate with the rest of the sign's message somewhere "downstream" during unconscious semantic processing. This would be consistent with the work of Greenwald and Liu (1985), who demonstrated that for evaluative priming, two-word phrases such as "enemy fails" did not function as positive primes (as they should if the meanings of the individual words were integrated into a single concept), but rather as 'double negative' primes (in the above example, each of the words in isolation has a negative connotation).

Both of these explanations would count as "ironic" effects of unconscious semantic processing, involving the activation of a negated concept, and its subsequent effect on behavior. However, another conceivable explanation side-steps our unconscious/negation model completely. Perhaps nicotine addiction is so robust that *any* smoking-relevant cue—positive or negative, conscious or unconscious—is sufficient to prime craving. Harris, Pierce, and Bargh (in prep), for example, found that exposing smokers to televised anti-smoking ads actually increased smoking behavior, even when the ads were consciously remembered, and regardless of ad valence: even grotesque images of blackened, cancerous lungs caused participants to light up cigarettes during a break between study tasks. However, as nicotine addiction levels in our sample group were quite low, the "unconscious semantic processing" hypothesis seems better able to account for our reported results. Future research should investigate these explanatory avenues in more detail. Whatever the mechanism, we report no difference in cigarette-approach behavior between participants who consciously noticed the no-smoking signs and those who did not.

GENERAL DISCUSSION

We may summarize our findings as follows: incidental exposure to no-smoking signs appears to boost implicit motivation to approach smoking-related stimuli. This motivation is evidenced by increased stimulus-approach movements by self-identified smokers after exposure to the primes. Furthermore, this automatic approach bias seems to occur regardless of the level of conscious processing of the no-smoking signs.

What does this mean for real life? The question of ecological validity is an important one. Let us assume that our findings are valid, that is, that incidental exposure to no-smoking signs through a computer task can increase the rate of cigarette-approach movements by several milliseconds, outside of participant awareness. This, of course, does not mean that every smoker who walks past a no-smoking sign on her way to the gym will find herself reaching for a pack of smokes and lighting up. In other words, the craving measure used in the present study is potentially so sensitive, and our prime so robust (participants in the experimental condition were exposed to 13 separate no-smoking signs over a short period of time), that the effect we report might not translate to actual smoking behavior. Nevertheless, having taken a first, controlled step in establishing this effect, we hope to have opened a door for future researchers to try blunter, noisier, and more ecologically valid meth-

ods in teasing out the real-life limits of the phenomenon.

Even if our effect turns out to be weak, however, it may still have important public health implications. A weak influence on behavior is still enough to tip the balance when an individual is on the edge between one course of action and another. With millions of smokers worldwide passing by multiple no-smoking signs several times per day, there is surely a fair chance that even the smallest boost in craving could have compound consequences from year to year, including increased sales for tobacco companies—one cigarette at a time. If this turns out to be the case, it might be prudent to reevaluate the widespread posting of no-smoking signs as a measure aimed at reducing public smoking.

There are broader implications still. How often do we attempt to influence others' attitudes and behavior by recruiting a negative frame—*don't* believe this; *stop* doing that—instead of by affirming the target judgment or action? Once called to mind, a concept is liable to stick there—whether it's preceded by a negation or not. The present research lends credibility to affirmation-based persuasion strategies, whether the domain is parenting, education, or public health, and raises a caution against discouragement based on "no."

REFERENCES

Bargh, J., & Chartrand, T. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H.T. Reis and Judd, C.M. (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 253-285). New York: Cambridge University Press.

Bargh, J., & Chartrand, T. (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462-479.

Bargh, J., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype priming on action. *Journal of Personality and Social Psychology*, 71, 230-244.

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

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, 91, 385-405.

Dijksterhuis, A., Bargh, J., & Miedema, J. (2000). Of men and mackerels: Attention and automatic behavior. In H. Bless & Forgas, J. P. (Eds.), *Subjective Experience in Social Cognition and Behavior* (pp. 36-51). Philadelphia: Psychology Press.

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior or how to win a game of trivial pursuit. *Journal of Personality and Social Psychology*, 74, 865-877.

Dostoevsky, F. (1997). *Winter notes on summer impressions* (D. Patterson, Trans.). Illinois: Northwestern University Press. Original work published 1863.

Freud, S. (1925). Negation. *The International Journal of Psychoanalysis*, 6, 367- 371.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, 44, 370-377.

Gilbert, D. T. (1991) How mental systems believe. *American Psychologist*, 46, 107-119.

Greenwald, A.G., & Liu, T.J. (1985). Limited unconscious process of meaning. *Bulletin of the Psychonomic Society*, 23, 292.

Harris, J., Pierce, M., & Bargh, J. (in prep). The perverse effect of antismoking campaigns on smoking behavior.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K. O. (1991). The Fagerstrom Test for Nicotine Dependence: A revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addiction*, 86, 1119-1127.

Kay, A., Wheeler, C., Bargh, J., & Ross, L. (2004). Material priming: the influence of mundane physical objects on situational construal and competitive behavioral choice. *Organizational Behavior and Human Decision Processes*, 95, 83-96.

Krieglmeyer, R., Deutsch, R., De Houwer, J., & De Raedt, R. (2010). Being moved: Valence activates approach-avoidance behavior independently of evaluation and approach-avoidance intentions. *Psychological Science*, 21, 607-613.

Quenda, D. (2009, December 31). Blowing smoke at a ban. *New York Times.*

Roehrich L., & Goldman M. S. (1995). Implicit priming of alcohol expectancy memory processes and subsequent drinking behavior. *Experimental and Clinical Psychopharmacology*, 3, 402-410.

Schwarz, N., & Clore, G. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513-523.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22.

Wason, P.C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, *11*, 92-107.

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*, 59-91.

Williams, L. E., & Bargh, J. A. (2008a). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322*, 606-607.

Williams, L. E., & Bargh, J. A. (2008b). Keeping one's distance: The influence of spatial distance cues on affect and evaluation. *Psychological Science*, *19*, 302-308.

Zhong, C.B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, *313*, 1451-1452.

Not Just in the Eye of the Beholder: Beauty as a Status Characteristic in Mixed Sex Dyads

Rachel E. Fudman

EMORY UNIVERSITY

The physical attractiveness bias, or the tendency to attribute positive characteristics to people perceived as attractive, offers attractive individuals a range of advantages in the social world. Incorporating concepts from research on attractiveness bias and expectation states theory, this study examined the communication behaviors of pairs working on a decision-making task in order to measure the influence of attractiveness and gender on social status hierarchy. Sixty-eight undergraduate students were separated into pairs that varied by gender and attractiveness rating and were videotaped while performing an interactive task. Researchers measured the dominant (interruptions, gestures, total talk time, speech initiation) and submissive (affirmations, head nods, smiling) communication behaviors exhibited by participants and used this information to infer status hierarchies for each pair. Attractive individuals displayed a range of dominant behaviors. In line with these findings, when participants evaluated their partners after the task, those who were described as attractive were also described as possessing other positive traits, such as intelligence and thoughtfulness. Results overall suggest that gender and physical attractiveness operate as status characteristics and, specifically, that physical attractiveness may elevate social status in face-to-face interaction.

INTRODUCTION

The physical attractiveness bias, or the tendency to attribute positive characteristics to attractive people, is ubiquitous in the social world and gives widespread advantages to attractive individuals. Not only are attractive people often perceived as possessing unrelated positive traits (Webster & Driskell, 1983), but they are also known to receive social (Benson, Karabenick, & Lerner, 1976), academic (Ritts, Patterson, & Tubbs, 1992), legal (Castellow & Wuensch, 1990), political (Lewis & Bierly, 1990), occupational (Hosoda, Stone-Romero, & Coats, 2003), and economic (Loh, 1993) advantages over less attractive people.

These benefits represent forms of status and power, whose processes social psychologists have modeled in a variety of ways, including social-role theory (Eagly, 1987) and communication accommodation theory (Giles, Taylor, & Bourhis, 1973). Expectation states theory (EST) has been described as a "leading explanation of social influence" and "the most systematic and empirically well-documented theory of status processes in groups currently available" (Kalkhoff & Thye, 2006, p. 219; Ridgeway, 2006, p. 347). Joseph Berger and colleagues developed EST to explain how status operates in small, task-oriented groups (Berger, Cohen, & Zelditch, 1972). While these groups emerge naturally in a range of contexts, they are of particular importance in the workplace, where teamwork is often required to achieve specific objectives. In light of the well-documented application of the attractiveness bias in this context (Vo, 2001), EST can be used to help discern the impact of physical attractiveness on group communication. Within this framework, attractiveness serves as a diffuse status characteristic that signals task competence in small groups (Webster & Driskell, 1983).

The Physical Attractiveness Bias

In a seminal study by Dion, Berscheid, and Walster (1972), participants attributed more socially desirable personality traits, such as friendliness and interestingness, to more attractive individuals. Participants also expected attractive targets to lead happier and more successful lives compared to unattractive targets. Research from the past several decades reveals the pervasiveness of this bias in many facets of social life. Several meta-analyses have reviewed such investigations and considered the range and intensity of the bias across different domains. In one review, Feingold (1992) found that attractive males and females were judged as more sociable, dominant, sexually warm, mentally healthy, and socially skilled but were not judged as more intelligent or competent. Eagly, Ashmore, Makhijani, and Longo (1991) found that attractiveness has a strong impact on perceptions of social competency and extraversion and a moderate impact on

The author can be reached at ref2130@columbia.edu or at Social-Organizational Psychology Department, Teachers College, Columbia University, 525 W 120th St., New York, NY 10027.

perceptions of intelligence and authority, but little impact on perceptions of integrity and concern for others. Jackson, Hunter, and Hodge (1995) argue that the bias is more diffuse, documenting its effects on diverse measures of competence. Though opinion varies on the strength and scope of this phenomenon, these reviews converge on a bias that moderates perceptions of traits not directly related to outward appearance.

Physical attractiveness can also impact decisions and behaviors. Research suggests that helping behavior increases when the recipient of the help is attractive. Benson, Karabenick, and Lerner (1976) found that people were more willing to help mail a graduate admissions application for more attractive individuals, and West and Brown (1975) found that attractive individuals received more donations in emergency conditions. Similarly, more room is given to attractive people on sidewalks (Dabbs & Stokes, 1975). These findings extend to evaluative judgments. Teachers judge physically attractive students as more intelligent, as having higher academic ability, and as better adjusted than unattractive students (Ritts, Patterson, & Tubbs, 1992; Lerner & Lerner, 1977). Evaluations of vocal performance, peer essays, and college admission interviews are positively associated with attractiveness (Landy & Sigall, 1974; Shahani, Dipboye, & Gherlein, 1993; Wapnick, Darrow, Kovacs, & Dalrymple, 1997). Preference and perceptions of competence of political candidates have been correlated with candidate physical attractiveness (Adams, 1977; Lewis & Bierly, 1990). Attractive people are less likely to be asked for identification when purchasing alcohol (McCall & Nattrass, 2001). Research on evaluations of legal proceedings using undergraduate and graduate student participants reveals a strong attractiveness bias in the courtroom as well. Unattractive victims are considered to be more responsible for attacks than attractive victims in mock rape trials, defendants are less likely to be judged guilty if they are attractive or if the victim is unattractive, and attractive defendants are given more lenient sentencing (Thornton & Ryckman, 1983; Jacobson, 1981). Similarly, research on sexual harassment suggests that a jury is least likely to vote a defendant guilty when the defendant is attractive and the plaintiff is unattractive (Castellow & Wuensch, 1990).

Moreover, researchers have documented biases in favor of attractive people on several job-related outcomes such as ranking, hiring decisions, promotions, predicted success, employment potential, and performance evaluations (Hosoda, Stone-Romero, & Coats, 2003). Less attractive female job applicants are less likely to be hired regardless of qualifications, and employees' attractiveness increases wage levels and wage growth (Marlowe, Schneider, & Nelson, 1996; Loh, 1993). And both professional and student raters believed that attractive candidates had personalities best suited for the job, would outperform others, and had a better chance of getting the job (Gilmore, Beehr, & Love, 1986).

While most research confirms the advantageousness of being perceived as physically attractive, a handful of studies reveal exceptions to the rule. This "dark side" of attractiveness operates largely through perceptions of vanity (Eagly et al., 1991). Attractive people are perceived as vainer than less attractive people, more egotistical, more materialistic, and more likely to have failed marriages as a result of extramarital affairs

(Dermer & Thiel, 1975). Other exceptions involve behaviors resulting directly from this bias. Sigall and Ostrove (1975) found that in a mock trial with undergraduate jurors, attractive defendants received harsher sentences on crimes related to attractiveness, such as swindling. Additionally, attractive women were not hired for stereotypically masculine jobs, such as managerial positions (Cash & Jonda, 1984). While these findings may seem to reveal an equally disadvantageous side of the attractiveness bias, research generally points to physical attractiveness as an advantage in social life (Eagly et al., 1991).

Expectation States Theory

Despite the pervasiveness of attractiveness bias, few researchers have addressed physical attractiveness as a status characteristic (Jackson et al., 1995; Webster & Driskell 1983). It is possible that attractiveness gives advantages in everyday interaction and across the lifespan similar to those of race, gender, education, and age. In turn, attractiveness could operate as a status characteristic as outlined by expectation states theory in the same way that being male or Caucasian operates (Webster & Driskell, 1983). While previous reviews have examined this bias from alternate theoretical frameworks, including implicit personality theory (Eagly et al., 1991), expectancy theory (Feingold, 1992), and accommodation theory (Haas & Gregory, 2005), the present research uses an expectation states perspective to explore physical attractiveness as an index of social status. We propose that the attractiveness bias can be best understood through status hierarchies within small groups, as expectation states theory (EST) "accounts for a broader range of attractiveness effects" than other theories in the social psychology field (Jackson et al., 1995).

EST emerged from a research program developed by Joseph Berger and colleagues testing status differences in social interaction (Berger et al., 1972). According to EST, members of task groups form performance expectations about each other unconsciously when there are no cues about who will be most competent at the task. These performance expectations, or expectation states, form and maintain power hierarchies within groups (Berger et al., 1972). Status Characteristics Theory (SCT) is a branch of EST in which performance expectations are based on culturally determined status characteristics (known as diffuse status characteristics) that provide cues about how successful each group member will be at the task. Diffuse status characteristics are those in which (a) one status group or "state" is valued more than other states and (b) those with the more valued state are deemed more competent on a variety of tasks (Berger et al., 1972; Correll & Ridgeway, 2003). Gender and race are the most robust examples of diffuse status characteristics (Webster & Hysom, 1998). For example, men are stereotypically believed to be more competent at a variety of tasks, even those not explicitly, or stereotypically, related to gender. Within the framework of SCT, when there is no obvious indication of who will be most successful at a task in an unstructured group, expectation states will be higher for men because they hold the more culturally valued gender status.

SCT is defined by a set of principles that characterize the relationship between status beliefs and behaviors (Correll & Ridgeway, 2003; Kalkhoff & Thye, 2006). According to the principle of *salience*, a significant status

characteristic must differentiate group members, or must be relevant to the task. Next, the *burden of proof* rests on the disadvantaged group member, who must prove that a salient status characteristic should not be considered when expectations are formed. In other words, a female in a group of mostly males will have to prove that her low gender status is not relevant to the specific task. The principle of *aggregation* explains that all salient status characteristics combine to form an aggregate expectation of each group member. Therefore, if both race and gender are salient in the situation, an African-American woman will have a lower expectation state than an African-American man. Lastly, these aggregated performance expectations create a social structure that informs group interaction, maintaining and perpetuating the social order. Group members with high performance expectations are given more opportunities to participate in tasks, participate more often, are evaluated positively by the group, and influence the group. The social hierarchy outlined by SCT only applies if all members of the group are *task-oriented* and *collectively oriented*, the two scope conditions of the theory. Group members are task-oriented if they are motivated to successfully complete the task and collectively oriented if they believe that it is necessary to take each other's opinions into account when performing the task (Correll & Ridgeway, 2003). Extensive empirical evaluation of EST, typically using graphical representations to map out the relative performance expectations and status expectations of participants, has produced significant evidence supporting the theory (Correll & Ridgeway, 2003).

The previously discussed pattern of bias suggests that physical attractiveness is consistent with status characteristics like gender, race, age, or social class in influencing perceptions and behavior in small groups. There has been only one other attempt to evaluate physical attractiveness as a status characteristic. Webster and Driskell (1983) manipulated the attractiveness of student photographs and had participants gauge their comparative expectations of the pictured individuals. Participants completed a questionnaire describing their expectations for an attractive student compared to an unattractive student of the same sex. The authors found that raters had high expectations of attractive people, both specific (competent at flying a plane) and general (competent in most situations). These findings qualify physical attractiveness as a status characteristic according to EST.

The current study departs from methodology typically employed in EST research in two significant ways. First, we expand on Webster and Driskell's (1983) research on same sex targets by evaluating mixed sex targets. The aggregation assumption says that all status characteristics influence performance expectations. Therefore, if gender and attractiveness are salient during a task, attractive males should possess the highest overall expectation advantage and unattractive females should possess the lowest expectation advantage. Second, like most EST experiments, Webster and Driskell's (1983) study used an experimental setting in which participants formed expectations of target individuals without ever interacting with them in person. By contrast, the current study is interested in how these expectation states manifest themselves in face-to-face interaction.

Status Differences in Communication Behaviors

Several studies have used EST to explore verbal and nonverbal communication behaviors that signify status differences within task groups. Findings from studies that directly observe behavior suggest that communication cues serve to reinforce preconceived performance expectations based on status characteristics (Ridgeway, Berger & Smith, 1985). Dovidio, Heltman, Brown, and Ellyson (1988) found that men displayed more power in gender-neutral tasks compared to women and attributed this outcome to men's conventionally higher gender status. Additional research on communication behaviors reveals that individuals in subordinate roles exhibit more hesitant and supporting behaviors, such as questions, affirmations, looking while listening, head nods, and smiling. Conversely, high-status individuals exhibit dominant behaviors, such as interruptions, directives, talking more, talking first, looking while speaking, hand gestures, and chin thrusts (Athenstaedt, Haas, & Schwab, 2004; Dovidio et al., 1988; Helweg-Larson, Cunningham, Carrico, & Pergram, 2004; Karakowsky, McBey, & Miller, 2004; Ridgeway et al., 1985).

It is important to note that these behaviors are also linked to gender-specific behaviors that exist part-and-parcel of perceived power differences between the sexes. Some exceptions seem to be a result of gender socialization as opposed to social status *per se* (though these concepts are interrelated insofar masculinity conventionally signals high status; Athenstaedt et al., 2004). For example, many scholars suggest that smiling is a gender-related behavior signaling warmth and positivity, not an indicator of low status or submissiveness (Athenstaedt et al., 2004; Dovidio et al., 1988; Hecht & LaFrance, 1998; Helweg-Larson, 2004). Ridgeway et al. (1985), on the other hand, argue that in task settings, nonverbal behavior between members of the opposite sex represents disparities in status and power independent of sex roles. Manipulating both physical attractiveness and gender could contribute to knowledge about the derivations of these status and sex differences in interaction.

THE PRESENT STUDY

The current study examined several verbal and nonverbal communication behaviors known to be indicative of either high or low status. High-status behaviors included length of time talking, speech initiation (first person to initiate speech), interruptions, and nonverbal gestures. Low-status behaviors included affirmations, smiling, and head nodding. These behaviors were selected for their variety, their ease of observation with video footage, and their presence in the literature.

Incorporating concepts from the physical attractiveness bias, expectation states theory, and communication behaviors, we examined whether communication behaviors between individuals varied by sex and attractiveness reflect status hierarchies. We hypothesize the following:

Hypothesis 1: Participants' attractiveness will be generalized to other positive traits.

Hypothesis 2: Attractiveness differences will be present for each communication behavior.

(a) Attractive participants will speak first, interrupt more, gesture more, and speak more overall compared to unattractive participants.

(b) Unattractive participants will affirm, head nod, smile, and write more compared to attractive participants.

Hypothesis 3: Gender differences will be present for each communication behavior.

(a) Male participants will speak first, interrupt more, gesture more, and speak more overall compared to female participants.

(b) Female participants will affirm, head nod, smile, and write more compared to male participants.

Hypothesis 4: Attractiveness differences for communication behaviors will vary by gender.

(a) Attractive males will use the most dominant communication behaviors.

(b) Unattractive females will use the most submissive communication behaviors.

Together, these hypotheses suggest that individuals possessing both forms of high status under examination (male and attractive) will communicate in a manner that represents their elevated status. Conversely, individuals possessing lower status (female and unattractive) will project low status. When attractiveness is held constant, gender will be the only salient status characteristic.

Methods

Participants and research staff

Sixty-eight volunteer participants were recruited through flyers and an online classifieds forum from the student population of a moderately sized southeastern liberal arts university and were arranged into 34 mixed sex dyads. Participants' ages ranged from 18 to 24 years, with a mean of 20.1. Thirty-three participants described themselves as Caucasian, 24 as Asian or Indian, six as African American, and five as Hispanic. No participants reported previous knowledge of their partner beyond recognition from a university course.

Fifteen undergraduate students (11 females and 4 males) served as attractiveness raters. The majority of these individuals were advanced students in the psychology honors program at the university. No raters were assigned participants whom they knew personally, and all raters signed a confidentiality agreement protecting the identity of participants.

Procedure

We implemented a between-group design in order to compare the verbal and nonverbal communication behaviors of college aged men and women working together on a task. So as not to give away the purpose of the study, volunteers were led to believe that they were participating in two sessions of research examining the relationship between personal values and ethical decision-making. Sessions were conducted in a laboratory room at the university.

In the initial research session, participants completed the Rokeach Values Survey (see Appendix 1). This questionnaire requires the respondent to rank order two different sets of 18 values according to person-

al importance. Examples of values included in the survey are "a comfortable life," "true friendship," "honesty," and "wisdom." Photographs of participants were taken in a standardized fashion and cropped below the shoulders. The researcher informed participants that the photograph would help maintain organization of the study files. Raters reported their subjective judgments of the physical attractiveness of participant photographs on a likert scale of 1 – 6, with 1 representing the lowest rating and 6 representing the highest rating. Each participant was rated by at least four raters, of whom at least one was male. The 33 participants with an average rating of 1 – 3.49 were designated a low attractiveness status and the remaining 35 with an average rating of 3.5 – 6 were designated a high attractiveness status.Because research shows that status is more salient in mixed sex settings, participants were placed into mixed sex pairs (Athenstaedt et al., 2004). Each pair was assigned to one of four groups:

Group 1: Attractive Male; Attractive Female (8 pairs)

Group 2: Unattractive Male; Unattractive Female (9 pairs)

Group 3: Attractive Male; Unattractive Female (9 pairs)

Group 4: Unattractive Female; Attractive Male (8 pairs)

Participants returned for a second research session at the same time as their assigned partner. They were instructed to sit next to each other and complete an ethical decision making exercise for 12 minutes (see Appendix 2). This videotaped task included descriptions of five ethical dilemmas adapted from Victor Grassian's book, *Moral Reasoning: Ethical Theory and Some Contemporary Moral Problems*. One example is entitled "A Poisonous Cup of Coffee":

> Tom, hating his wife and wanting her dead, puts poison in her coffee, thereby killing her. Joe also hates his wife and would like her dead. One day, Joe's wife accidentally puts poison in her coffee, thinking it's cream. Joe has the antidote, but he does not give it to her. Knowing that he is the only one who can save her, he lets her die. Is Joe's failure to act as bad as Tom's action? Why?

Participants were provided with only one copy of task instructions and dilemmas. They were able to move on to the next dilemma after arriving at an agreement or compromise about the previous one. As stated previously, the scope conditions of EST are task orientation, or motivation to complete the task successfully, and collective orientation, or the understanding that the opinion of others is important for the task. The ethical dilemma exercise used in this research meets both of these conditions, as both partners are led to believe that the task is an important element of the research and know that they must collaborate in order to proceed through the task.

Immediately following the discussion, participants completed a questionnaire about their personal performance on the task, the performance of their partner, and their perceptions of their partner (see Appendix 3). Responses were reported on a 1 – 5 scale from "Strongly Disagree" to "Strongly Agree." Examples of these statements include, "I found this task to be difficult," "My partner performed well on this task," "My partner was open to new ideas," and "My partner was attractive." During debrief, par-

Table 1. Means and standard deviations for communication behaviors.

			Verbal			Nonverbal		
			Dominant		Submissive	Dominant	Submissive	
			Talk Time	Interruptions	Affirmations	Gestures	Head Nods	Smiles
Male	Attractive	Mean	129.8	2.2	9.5	5.9	2.8	5.3
		SD	53.1	2.1	6.7	4.0	3.9	4.5
	Unattractive	Mean	136.1	2.6	10.6	8	4.8	5.7
		SD	47.6	1.9	7.8	6.1	5.8	4.0
	Total	Mean	133	2.4	10	6.9	3.8	5.5
		SD	49.8	2.0	7.2	5.1	5.0	4.2
Female	Attractive	Mean	119.1	3.5	9.7	6.3	5	7.6
		SD	51.1	3.3	5.5	5.5	54.3	5.8
	Unattractive	Mean	103.2	2.6	11.3	4.3	3.4	7.7
		SD	49.8	2.4	7.3	4.4	4.5	4.8
	Total	Mean	110.7	3	10.6	5.2	4.6	7.7
		SD	50.3	2.9	6.5	5.0	5.0	5.2
Total	Attractive	Mean	124.6	2.8	9.6	6.1	4.4	6.4
		SD	51.6	2.8	6	4.7	4.8	5.3
	Unattractive	Mean	119.2	2.6	11	6.0	4.1	6.7
		SD	50.9	2.2	7.4	5.5	5.1	4.5
Total	Total	Mean	121.8	2.7	10.3	6.0	4.2	6.6
		SD	50.9	2.5	6.8	5.0	5.0	4.8

ticipants were not informed of the existence or results of their attractiveness rating. Instead, they were told that this research focuses on the communication behaviors both partners exhibited as well as the influence of reported perceptions on their interaction. As an attractiveness check, participants' perceptions of their partners' physical attractiveness correlated significantly with the attractiveness rating assigned previously,

Table 2. Group means and standard deviations by gender for head nods.

Group	Gender	Mean	SD
Attractive Male - Attractive Female	Male	3.3	5.4
	Female	7	6.4
Unattractive Male - Unattractive Female	Male	5	5.9
	Female	4.9	6.1
Attractive Male - Unattractive Female	Male	2.4	2.3
	Female	1.9	1.3
Unattractive Male - Attractive Female	Male	4.5	6.0
	Female	5.0	4.0

Table 3. Group means and standard deviations by gender for smiles.

Group	Gender	Mean	SD
Attractive Male - Attractive Female	Male	4.1	2.0
	Female	7.6	7.4
Unattractive Male - Unattractive Female	Male	4.0	2.9
	Female	8.0	4.0
Attractive Male - Unattractive Female	Male	6.3	5.9
	Female	7.4	5.7
Unattractive Male - Attractive Female	Male	7.6	4.3
	Female	7.6	4.3

Pearson's $r(68) = .31$, $p < .05$.

Four undergraduate research assistants coded the middle eight minutes of video footage for each pair. The middle section was taken to avoid using footage prior to task orientation when participants were introducing themselves or after task orientation in the event that a pair completed the task early. Athenstaedt et al. (2004) also used a middle section of footage for analyses of gender differences. Assistants were provided with definitions of each behavior and participated in a training session on correct coding procedures. The verbal behaviors coded were speech initiation, total talk time in seconds, frequency of interruptions (dominant behaviors), and affirmations (submissive behavior). Nonverbal behaviors coded were gestures (dominant behavior), head nods, and smiling (submissive behaviors). Research assistants also recorded which participant of the pair elected to write the pair's answers on the assignment sheet. The primary researcher coded speech initiation (the first person to initiate speech). All research assistants coded the same footage of two pairs of participants in order to establish inter-rater reliability. Intraclass correlation coefficients were calculated for affirmations ($ICC = .421$, $p < .05$), gestures ($ICC = .961$, $p < .05$), head nods ($ICC = .543$, $p < .05$), smiles ($ICC = .969$, $p < .05$), and talk time ($ICC = .971$, $p < .05$).

Results

The first hypothesis, that the attractiveness of participants would be generalized to other positive traits, was confirmed. Participants who described their partners as attractive also described them as intelligent ($r(68) = .375$, $p < .05$), open to new ideas ($r(68) = .253$, $p <. 05$), capable ($r(68) = .271$, $p < .05$), thoughtful ($r(68) = .357$, $p < .05$), articulate ($r(68) = .459$, $p < .05$), warm ($r(68) = .316$, $p < .05$), engaging ($r(68) = .218$, $p < .05$) and funny ($r(68) = .32$; $p < .05$), but not significantly as helpful ($p > .05$) or ethical ($p > .05$).

The second, third, and fourth hypotheses—that physical attractiveness would im-

Table 4. Group means and standard deviations by gender for gestures.

Group	Gender	Mean	SD
Attractive Male - Attractive Female	Male	5.4	3.6
	Female	6.0	4.4
Unattractive Male - Unattractive Female	Male	9.1	7.1
	Female	4.3	4.6
Attractive Male - Unattractive Female	Male	6.3	4.4
	Female	4.2	4.5
Unattractive Male - Attractive Female	Male	6.9	5.1
	Female	6.5	6.7

pact communication behaviors and that this effect would vary by gender—were partially supported. We chose to use a confidence interval of 90% due to the decrease in sample size to 68 participants over all four groups. Square root log transformations for interruptions, affirmations, gestures, and head nods were performed to correct positive skew. We performed a 2 (attractive, unattractive) × 2 (male, female) analysis of variance (ANOVA) for each communication behavior in order to determine the influences of physical attractiveness and gender on communication between partners. First, gender main effects were examined exclusively. Smiles ($F(1, 68) = 3.457$, $p < .10$, $\eta p^2 = .051$), talk time ($F(1, 68) = 3.163$, $p < .10$, $\eta p^2 = .047$), and writer ($F(1, 68) = 22.74$, $p < .10$, $\eta p^2 = .263$) had marginally significant main effects. Females smiled more and talked less compared to males (see Table 1). They also served the role of "secretary" more often than men, writing down the pair's answers to the task. Only speech initiation approached significance for attractiveness ($F(1, 68) = 16.34$, $p < .10$, $\eta p^2 = .203$); attractive participants were usually the first to initiate conversation with their partner.

Next, the interactions of the ANOVAs were examined in order to establish that the effect of physical attractiveness on communication varies by gender. Speech initiation ($F(1, 68) = 3.72$, $p < .10$, $\eta p^2 = .055$) and head nods ($F(1, 68) = 5.17$, $p < .10$, $\eta p^2 = .075$) had significant interactions.

Means and standard deviations for all numeric communication behaviors are listed in Table 1. Notably, unattractive men ($M = 4.8$, $SD = 5.8$) head-nodded more than attractive men ($M = 2.8$, $SD = 3.9$). Attractive females ($M = 6.3$, $SD = 5.5$) gestured more than unattractive females ($M = 4.3$, $SD = 4.4$). Attractive females ($M = 119.1$, $SD = 51.1$) spoke more than unattractive females ($M = 103.2$, $SD = 49.8$).

Chi-square tests were performed as a factorial test for the categorical variables of writer ($\chi^2 (1) = 14.435$, $p < .05$) and speech initiation ($\chi^2 (1) = 3.29$, $p < .05$). M-H estimates were 8.57 for the writer and 6.06 for speech initiation.

Next, we examined group means in order to discern patterns of communication across attractiveness and gender. Tables 2-4 show group means for behaviors that display trends consistent with the hypothesis. As shown in Table 2, attractive females head-nod less when paired with unattractive males ($M_{Female} = 5$, $SD = 4$) than with attractive males ($M_{Female} = 7$ $SD = 6.4$). While the factorial ANOVA for smiling revealed only a main effect for sex, unattractive males smile almost twice as much when paired with attractive females ($M_{Male} = 7.6$, $SD = 4.3$) than with unattractive females ($M_{Male} = 4$, $SD = 2.9$; see Table 3). As shown in Table 4, attractive males gesture approximately the same amount as attractive females ($M_{Male} = 5.4$, $SD = 3.6$, $M_{Female} = 6$, $SD = 4.4$) but attractive males gesture more than unattractive females ($M_{Male} = 6.3$, $SD = 4.4$, $M_{Female} = 4.2$, $SD = 4.5$).

DISCUSSION

The primary objective of this investigation was to determine whether attractiveness influences communication behaviors in face-to-face interaction, while also replicating previous findings that attractiveness, like gender, operates as a status characteristic. More generally, we were interested in whether participants' perceptions of their partners provide further insight into the

cognitive bias for physical attractiveness. If communication between partners is an expression of relative status, then broad physical attractiveness bias can be understood through the expectation states perspective.

Hypothesis 1

Participants evaluated attractive individuals more positively, establishing the presence of an attractiveness bias among pairs. This finding reinforces previous research showing that attractive individuals are perceived favorably in both social (e.g., open, thoughtful, articulate, warm, engaging, funny) and intellectual (e.g., intelligent, capable) domains (Eagly et al., 1991; Jackson et al., 1995). Furthermore, the finding that attractive individuals were not assumed to be helpful or ethical aligns with Eagly's findings that the attractiveness bias has little to do with integrity or concern for others (Eagly et al., 1991). This pattern of attractiveness bias has a clear and consistent effect throughout the literature. According to expectation states theory, group members and partners evaluate individuals with the more valued characteristic positively. In this case, attractive individuals received approving feedback about their social and cognitive abilities based on their task performance, suggesting that attractiveness possesses the evaluative quality of a status characteristic.

Hypotheses 2-4

The hypothesis that attractiveness affects communication behaviors was supported by several observed behaviors, suggesting that attractiveness is a status characteristic operative in casual conversation and task performance. Because an interaction between attractiveness and gender was only found for two behaviors, the relationship between these characteristics and the relative strength of each remains unclear.

This investigation successfully replicated previous research that identified gender as an overt status characteristic in communication. Males talked more than females, verbally expressing their power by speaking more overall during the task (Dovidio et al., 1988). Females smiled more than males, which is consistent with suggestions that smiling is a gender-specific behavior and not necessarily a behavior exhibited by other low status traits as an indication of power (Athenstaedt et al., 2004; Dovidio et al., 1988; Hecht & LaFrance, 1998; Helweg-Larson, 2004). Females also tended to take on note-taking duties during the task. While this could be a submissive act, it could also result from a separate assumption that college-aged females have more legible handwriting than males.

The behaviors of speech initiation and head nods had significant interactions with participants' attractiveness and gender. Speech initiation was measured as the order in which partners spoke, recording which participant initiated speech. Attractive males spoke first more often than unattractive males, and attractive females spoke first more often than unattractive females. Overall, males spoke first more often than females. These results confirm previous findings that individuals who initiate speech have higher status and are likely to participate more frequently (Dovidio et al., 1988, Ridgeway et al., 1985).

The submissive behavior of head nodding also yielded a significant attractiveness-by-gender interaction. Unattractive males nodded their heads more than attractive males, signifying the influence of attrac-

tiveness on status among males. Furthermore, attractive females nodded their heads more toward attractive males than toward unattractive males, suggesting that the status difference between men and women was moderated by attractiveness. In light of previous research showing that head nods are a form of submission (Helweg-Larson et al., 2004), the head-nodding behavior of attractive females suggests that they are less willing to express submission to unattractive males than to attractive males.

With the exception of significant results for speech initiation, most visible trends for attractiveness were found in nonverbal communication behaviors. No significant effects were found for verbal affirmations or interruptions, which is inconsistent with literature depicting these behaviors as indicators of dominance and power in verbal communication (Athenstaedt et al., 2004, Karakowsky, 2004). It is possible that the experimental context prevented either of these behaviors from generating significant results. Partners were introduced for the first time immediately before the ethical decision making task began, so they likely wanted to interact in a socially desirable and supportive manner in order to make a good impression. This would include encouraging their partner through affirmations and abstaining from interruptions. Additionally, the experimental design primed participants with notions of morals and ethics, which may have influenced them to act accordingly. The research design required that partners were collectively oriented (i.e., interested in one another's opinions) in accordance with the scope conditions of expectation states theory. In this case, they were required to come to an agreement about their answer for each ethical situation's question in order to proceed with the task. This personal investment in the other participant's ideas probably facilitated a supportive climate with multiple affirmations and infrequent interruptions, regardless of status characteristics. In fact, participants of the current study used as many affirmations in eight minutes as the participants in Athenstaedt et al.'s (2004) study used in twelve minutes. Athenstaedt et al. (2004) found sex differences for interruptions with mixed sex couples who were familiar with each another. Similarly, Karakowsky et al. (2004) found interruption differences for discussion groups when indicators of participants' level of task competence were present. In light of the present research, these findings suggest that familiarity with a partner and clear indicators of ability influence status hierarchies as reflected by some verbal behaviors.

Limitations and Suggestions for Future Research

Several limitations of this investigation should be considered. First, the sample size could have limited statistical power. In the future, group sizes should exceed ten pairs each. A larger sample size could have yielded significant results for all observed trends. A large initial participant pool would also allow researchers to foreground individuals on either end of the attractiveness spectrum. Researchers could either eliminate individuals with average attractiveness ratings from participation or include them to examine the effects of an "average" status. By contrast, the present study dichotomized all ratings, even those in the average range.

Another limitation of this study was the inter-rater reliability of behavioral coding by research assistants. Intraclass correlation

coefficients for affirmations (.421) and head nods (.543) were not sufficient. While it is possible that more extensive reliability tests (requiring all assistants to code more than the footage of two pairs for reliability purposes) could have revealed satisfactory reliability, the coding of affirmations and head nods was inconsistent. In this case, future research should consider using fewer coders or implementing more extensive training for behavioral coding procedures.

Participant behavior may have been biased due to the Hawthorne effect, in which individuals tend to act differently because they know that they are being watched, or in this case because of the presence of a video camera. For example, feelings of nervousness could directly impact the communication behaviors in question; a timid participant may smile or talk more or less in front of a camera. In addition, although participants completed their post-test questionnaires in separate rooms, they were still present in the same workspace and often exited simultaneously. Close proximity to partners could have influenced participants to evaluate their partner positively.

Future research on the cognitive bias of physical attractiveness should continue to use an expectation states perspective to explain attractiveness as a function of status in society. Specifically, in the context of group interaction, it should examine same sex pairs in addition to mixed sex pairs. Holding sex constant could reveal important patterns about attractiveness as a status characteristic within either males or females. Future research could also examine the interaction between attractiveness and other status characteristics besides sex, such as race or age. Furthermore, in addition to the communication behaviors observed here, other advantages of status characteristics outlined by expectation states theory should be measured, such as persuasive ability, participation rate, and opportunity to participate.

Some scholars have suggested that the scope conditions of expectation states theory are expandable. Studies have demonstrated the operation of diffuse status characteristics even in settings that are not both collectively oriented and task-oriented (Correll & Ridgeway, 2003; Foschi, Lai, & Sigerson, 1994). While further research in this area is needed, this evolution of EST would be helpful for considering the specific settings in which physical attractiveness operates as a status characteristic. Without collective orientation, interruption and affirmation behaviors could increase. Continuing research on attractiveness as a status characteristic should consider expanding scope conditions beyond the small work group to individually evaluative tasks (e.g., standardized testing) or to unobtrusive research designs.

Overall, this investigation contributes to research on the attractiveness bias, confirming a cognitive halo effect for physical attractiveness vis-à-vis other culturally positive traits. Our results are consistent with previous findings suggesting that gender and attractiveness are status characteristics which construct and maintain status hierarchies within small group interactions. Specifically, we extended previous work on attractiveness bias with same sex dyads to mixed sex dyads and examined status hierarchies in both verbal and nonverbal communication behaviors.

REFERENCES

Adams, G. R. (1977). Physical attractiveness research: Toward a developmental social psychology of beauty. *Human Development, 20*, 217-239.

Athenstaedt, U., Haas, E., & Schwab, S. (2004). Gender role self-concept and gender-typed communication behavior in mixed sex and same sex dyads. *Sex Roles, 50*, 37-52.

Benson, P. L., Karabenick, S. A., & Lerner, R. M. (1976). Pretty pleases: The effects of physical attractiveness, race, and sex on receiving help. *Journal of Experimental Social Psychology, 12*, 409-415.

Berger, J., Cohen, B., & Zelditch, M. (1972). Status characteristics and social interaction. *American Sociological Review, 37*, 241-255.

Cash, T. F., & Janda, L. H. (1984). The eye of the beholder. *Psychology Today*, 46-52.

Correll, S. J., & Ridgeway, C. L. (2003). Expectation states theory. In J. DeLameter (Ed.), *Handbook of Social Psychology* (pp. 29-51). New York: Springer.

Castellow, W. A., Wuensch, K. L., & Moore, C. H. (1990). Effects of physical attractiveness of the plaintiff and defendant in sexual harassment judgments. *Journal of Social Behavior and Personality, 5*, 547-562.

Dabbs, Jr., M., & Stokes, N. A. (1975). Beauty is power: The use of space on the sidewalk. *Sociometry, 38*, 551-557.

Dermer, M., & Thiel, D. L. (1975). When beauty may fail. *Journal of Personality and Social Psychology, 46*, 991-1004.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*, 285-290.

Dovidio, J. F., Brown, C. E., Heltman, K., & Ellyson, S. L. (1988). Power displays between women and men in discussions of gender-linked tasks: A multichannel study. *Journal of Personality and Social Psychology, 55*, 580-587.

Eagly, A. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L.C. What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110*, 109-128.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111*, 304-341.

Foschi, M., Lai, L., & Sigerson, K. (1994). Gender and double standards in the assessment of job applicants. *Social Psychology Quarterly, 57*, 326-339.

Giles, H., Taylor, D.M., & Bourhis, R.Y. (1973). Towards a theory of interpersonal accommodation through speech: Some Canadian data. *Language in Society, 2*, 177-192.

Gilmore, D.C., Beehr, T.A., Love, K.G. (1986). Effects of applicant sex, applicant physical attractiveness, type of rater, and type of job on interview decisions. *Journal of Occupational Psychology, 59*, 103-109.

Haas, A., & Gregory, Jr., S.W. (2005). The impact of physical attractiveness on women's social status and interactional power. *Sociological Forum, 20*, 449-470.

Hecht, M. A., & LaFrance. M. (1998). License or obligation to smile: The effect of power and sex on amount and type of smiling. *Personality and Social Psychology Bulletin, 24*, 1332-1342.

Helweg-Larson, M., Cunningham, S. J., Carrico, A., & Pergram, A. M. (2004). To nod or not to nod: An observational study of nonverbal communication and status in female and male college students. *Psychology of Women Quarterly, 28*, 358-361.

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel Psychology, 56*, 431-462.

Jackson, L. A., Hunter, J. E., Hodge, C. N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quarterly, 58*, 109-122.

Jacobson, M. B. (1981). Effects of victim's and defendant's physical attractiveness on subjects' judgments in a rape case. *Sex Roles, 7(3)*, 247-255.

Kalkhoff, W., & Shane, T. R. (2006). Expectation states theory and research: New observations from meta-analysis. *Sociological Methods & Research, 35*, 219-249.

Karakowsky, L., McBey, K., & Miller, D. L. (2004). Gender, perceived competence, and power displays: Examining verbal interruptions in a group context. *Small Group Research, 35*, 407-439.

Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29*, 299-304.

Lerner, R. M., & Lerner, J. V. (1977). Effects of age, sex, and physical attractiveness on child-peer relations, academic performance, and elementary school adjustment. *Developmental Psychology, 13*, 585-590.

Lewis, K. E., & Bierly, M. (1990). Toward a profile of the female voter: Sex differences in perceived physical attractiveness and competence of political candidates. *Sex Roles, 22*, 1-12.

Loh, E. S. (1993). The economic effects of physical appearance. *Social Science Quarterly, 74*, 420-438.

McCall, M., & Nattrass, K. (2001). Carding for the purchase of alcohol: I'm tougher than other clerks are! *Journal of Applied Social Psychology, 31*, 2184-2194.

Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology, 81*, 11-21.

Ridgeway, C. L., Berger, J., & Smith, L. (1985). Nonverbal cues and status: An expectation states approach. *The American Journal of Sociology, 90*, 955-978.

Ridgeway, C. (2006). Expectation states theory and emotion. In Stets, J. E., & Turner, J. H. (Ed.), *Handbook of the Sociology of Emotions* (pp. 347-367). New York: Springer.

Ritts, V., Patterson, M. L., & Tubbs, M. E. (1992). Expectations, impressions, and judgments of physically attractive students: A review. *Review of Educational Research, 62*, 413-426.

Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgment. *Journal of Personality and Social Psychology, 31*, 410-414.

Shahani, C., Dipboye, R. L., & Gherlein, T. M. (1993). Attractiveness bias in the interview: Exploring the boundaries of an effect. *Basic and Applies Psychology, 14*, 317-328.

Thornton, B., & Ryckman, R. M. (1983). The influence of a rape victim's physical attractiveness on observers' attributions of responsibility. *Human Relations, 36*, 549-561.

Vo, L. T. (2001). A more attractive look at physical appearance-based discrimination: Filling the gap in appearance-based anti-discrimination law. *Southern Illinois University Law Journal, 26*, 339-360.

Wapnick, J., Darrow, A. A., Kovacs, J., & Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education, 45*, 470-479.

Webster, M., & Driskell, J. E. (1983). Beauty as status. *The American Journal of Sociology, 89*, 140-165.

Webster, M., & Hysom, S. J. (1998). Creating status characteristics. *American Sociological Review, 63*, 351-378.

West, S. G., & Brown, T. J. (1975). Physical attractiveness, the severity of the emergency and helping: A field experiment and interpersonal simulation. *Journal of Experimental Social Psychology, 11*, 531-538.

APPENDIX 1

Values Questionnaire

On this page are 18 values listed in alphabetical order. Your task is to arrange them in order of their importance to YOU, as guiding principles in YOUR life. Study the list carefully and pick out the value which is most important for you. Put this value on line 1.Then pick out the value which is second most important for you. Put this value on line 2. Then do the same for each of the remaining values. The value that is least important should be placed on line 18. Work slowly and think carefully. If you change your mind, feel free to change your answers. The end result should truly show how you really feel.

1_____	A COMFORTABLE LIFE
2_____	AN EXCITING LIFE
3_____	A SENSE OFACCOMPLISHMENT
4_____	A WORLD AT PEACE
5_____	A WORLD OF BEAUTY
6_____	EQUALITY
7_____	FAMILY SECURITY
8_____	FREEDOM
9_____	HAPPINESS
10_____	INNER HARMONY
11_____	MATURE LOVE
12_____	NATIONAL SECURITY
13_____	PLEASURE
14_____	SALVATION
15_____	SELF-RESPECT
16_____	SOCIAL RECOGNITION
17_____	TRUE FRIENDSHIP
18_____	WISDOM

WHEN YOU HAVE FINISHED, GO TO THE NEXT PAGE

Below is another list of 18 values. Arrange them in order of importance, the same as before.

1_____	AMBITIOUS
2_____	BROADMINDED
3_____	CAPABLE
4_____	CHEERFUL
5_____	CLEAN

6_____ COURAGEOUS
7_____ FORGIVING
8_____ HELPFUL
9_____ HONEST
10_____ IMAGINATIVE
11_____ INDEPENDENT
12_____ INTELLECTUAL
13_____ LOGICAL
14_____ LOVING
15_____ OBEDIENT
16_____ POLITE
17_____ RESPONSIBLE
18_____ SELF-CONTROLLED

APPENDIX 2

TASK DIRECTIONS:

Below is a description of 5 ethical dilemmas adopted from Victor Grassian's book, *Moral Reasoning: Ethical Theory and Some Contemporary Moral Problems*. Please read each dilemma carefully with your partner. Discuss the questions that the dilemmas pose and any relevant moral issues that apply.

You and your partner must come to an agreement about each dilemma. Once you have agreed, please explain your answer in the box provided below each dilemma. You may not proceed to the next dilemma until you and your partner agree on an answer to put down.

You have 15 minutes for this task. Please take your time. You are not required to address all 5 dilemmas.

1. A Poisonous Cup of Coffee

Tom, hating his wife and wanting her dead, puts poison in her coffee, thereby killing her. Joe also hates his wife and would like her dead. One day, Joe's wife accidentally puts poison in her coffee, thinking its cream. Joe has the antidote, but he does not give it to her. Knowing that he is the only one who can save her, he lets her die. Is Joe's failure to act as bad as Tom's action? Why?

2. The Partiality of Friendship

Jim has the responsibility of filling a position in his firm. His friend Paul has applied and is qualified, but someone else seems even more qualified. Jim wants to give the job to Paul, but he feels guilty, believing that he ought to be impartial. That's the essence of morality, he initially tells himself. This belief is, however, rejected, as Jim resolves that friendship has a moral importance that permits, and perhaps even requires, partiality in some circumstances. So he gives the job to Paul.

Was he right?

3. The Overcrowded Lifeboat

In 1842, a ship struck an iceberg and more than 30 survivors were crowded into a lifeboat intended to hold 7. As a storm threatened, it became obvious that the lifeboat would have to be lightened if anyone were to survive. The captain reasoned that the right thing to do in this situation was to force some individuals to go over the side and drown. Such an action, he reasoned, was not unjust to those thrown overboard, for they would have drowned anyway. If he did nothing, however, he would be responsible for the deaths of those whom he could have saved. Some people opposed the captain's decision. They claimed that if nothing were done and everyone died as a result, no one would be responsible for these deaths. On the other hand, if the captain attempted to save some, he could do so only by killing others and their deaths would be his responsibility; this would be worse than doing nothing and letting all die. The captain rejected this reasoning. Since the only possibility for rescue required great efforts of rowing, the captain decided that the weakest would have to be sacrificed. In this situation it would be absurd, he thought, to decide by drawing lots who should be thrown overboard.

As it turned out, after days of hard rowing, the survivors were rescued and the captain was tried for his action. If you had been on the jury, how would you have decided? Why?

4. A Callous Passerby

Roger Smith, a quite competent swimmer, is out for a leisurely stroll. During the course of his walk he passes by a deserted pier from which a teenage boy who apparently cannot swim has fallen into the water. The boy is screaming for help. Smith recognizes that there is absolutely no danger to himself if he jumps in to save the boy; he could easily succeed if he tried. Nevertheless, he chooses to ignore the boy's cries. The water is cold and he is afraid of catching a cold – he doesn't want to get his good clothes wet either. "Why should I inconvenience myself for this kid," Smith says to himself, and passes on.

Does Smith have a moral obligation to save the boy? If so, should he have a legal obligation as well? Why?

5. The Torture of the Mad Bomber

A madman who has threatened to explode several bombs in crowded areas has been apprehended. Unfortunately, he has already planted the bombs and they are scheduled to go off in a short time. It is possible that hundreds of people may die. The authorities cannot make him divulge the location of the bombs by conventional methods. He refuses to say anything and requests a lawyer to protect his fifth amendment right against self-incrimination. In exasperation, some high-level official suggests torture. This would be illegal, of course, but the official thinks that it is nevertheless the right thing to do in this desperate situation.

Do you agree? If you do, would it also be morally justifiable to torture the mad bomber's innocent wife if that is the only way to make him talk? Why?

APPENDIX 3

Please read each item carefully and circle the one answer that works best. Because this process involved both you and your partner, your perceptions of your partner are important. Describe your experience with the ethical discussion honestly, and state your opinions as accurately as possible. Please make sure your answer is marked in the correctly numbered space.

[1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree]

1. I found this task to be difficult.
2. I think my partner found this task to be difficult.
3. I performed well on this task.
4. My partner performed well on this task.
5. My partner and I agreed on most issues we discussed.
6. My partner was intelligent.
7. Based on this task, I believe my partner and I have similar values.
8. My partner was open to new ideas.
9. My partner was capable.
10. My partner was thoughtful.
11. My partner was articulate.
12. My partner was attractive.
13. My partner was warm.
14. My partner was ethical.
15. My partner was engaging.
16. My partner was funny.
17. My partner was helpful.
18. My partner and I got along.
19. My partner appreciated my input.
20. I appreciated my partner's input.

21. Did you know your partner before participating in this task? If so, how well?
22. Task partners often disagree. If this happened in your discussion, how did you and your partner try to come to an agreement?

Beyond Pointing: Development of the "Showing" Gesture in Children with Autism Spectrum Disorder

Caitlin Clements & Katarzyna Chawarska

YALE UNIVERSITY

While research on early indicators of autism spectrum disorder (ASD) has focused predominantly on infants' inability to point, other social gestures such as showing may also help to distinguish ASD from non-autistic language delays in high-risk populations. Parents of infants at high-risk (HR) for ASD (n = 28) and typically developing infants at low-risk (TD-LR; n = 12) reported their child's gesture use at nine and twelve months. Infants were followed until 24 months, at which time 8 high-risk children showed symptoms of ASD (ASD-HR), 14 showed symptoms of language delay (LD-HR), and 6 were developing typically (TD-HR). Results indicate a lack of pointing in both ASD-HR and LD-HR groups at twelve months, but decreased frequency of showing only in the ASD-HR group. These results suggest that the showing gesture might be a better indicator of ASD than pointing among infant siblings of children with ASD. Implications for early identification are discussed.

INTRODUCTION

Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a group of neurodevelopmental disorders characterized by profound deficits in communication, social functioning, and the presence of repetitive and stereotyped behaviors or interests (American Psychiatric Association, 2004). It affects 1 in 110 children in the US (CDC, 2009) and in most cases impairs the child's ability to function independently (Howlin, 2005). Because there is currently no known cure, to date the best means of mitigating the negative effects of ASD are early identification. This way, children can begin an early behavioral intervention program (Klin, Chawarska, Rubin, & Volkmar, 2004).

ASD and nonverbal communication

Parents and clinicians use early indicators, such as delayed speech and language, to recognize ASD during the second year of life (Chawarska et al., 2007). Though communication is often impaired in ASD (American Psychiatric Association [*DSM-IV-TR*], 2004), identification during the first year of life cannot depend on verbal language alone. Even in typical development, children do not speak their first words until around twelve months of age and do not combine words into phrases until around 18 months (Tager-Flusberg, Paul, & Lord, 2005). Therefore, early recognition and diagnosis must rely on indicators of nonverbal communication skills that typically emerge before twelve months of age.

Nonverbal communication

Nonverbal communication skills develop before verbal language and therefore can play an important role in identification of ASD during the first year of life (Baron-Cohen, Cox, Baird, Sweettenham, & Nightingale, 1996; Chawarska et al., 2007; Lord, 1995; Mundy, Sigman, Ungerer, & Sherman, 1986; Paul, 2007; Stone, Ousley, Yoder, Hogan, & Hepburn, 1997; Wetherby, 1986). Nonverbal communication behaviors such as pointing, waving goodbye, and smiling convey a message to another person without language (Tager-Flusberg et al., 2005). Nonverbal communication is particularly relevant to early ASD diagnosis because children use hand gestures such as reaching, pointing, and giving to express their desires before they can do so with verbal language. In typical development, gestures emerge around eight to ten months and are used frequently until around 18 months, when their use begins to fade as verbal language develops to replace them (Bates & Dick, 2002; Capone & McGregor, 2004).

Pointing

The vast majority of studies on nonverbal communication in ASD have focused on

The authors would like to give special thanks to Jessie Bradshaw, Suzanne Macari, and Fred Shic.
The first author can be reached at cclements2@partners.org or at Massachusetts General Hospital, Simches Research Bldg., 6th floor, 185 Cambridge St., Boston, MA 02114.

pointing gestures. Pointing emerges around eight to ten months in typically developing children (Bates & Dick 2004; Capone & McGregor, 2004), and accounts for the majority (60%) of gestures by twelve months (Kita, 2003). There are two types of pointing: protodeclarative and protoimperative (Paul, 2007; Tomasello, 2007). *Protoimperative* points represent desire for an object, e.g., pointing to a cookie. *Protodeclarative* points indicate the desire to share an experience with another person, e.g., a child pointing to an airplane overhead. Protodeclarative pointing requires joint attention, or the ability to share experiences with others by attracting or following their attention by looking or pointing (Paul, 2007; Tomasello, 2007; Loveland & Landry, 1986). Joint attention is an early emerging social cognitive skill and is impaired in autism (Dawson et al., 2004; Loveland & Landry, 1986).

Pointing in Autism Spectrum Disorders

In autism, protodeclarative pointing is impaired while protoimperative pointing remains relatively intact, especially in older children (Baron-Cohen, 1989; Camaioni, Perucchini, Muratori, & Milone, 1997; Camaioni, Perucchini, Muratori, Parrinini, & Cesari, 2003; Itoh, 2000). Because protodeclarative pointing requires social interest, its deficit is considered one of the critical ASD risk indicators on a number of screening instruments, including the Checklist for Autism in Toddlers (CHAT; Baron-Cohen, Allen, & Gillberg, 1992) and Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001). Pointing deficits have proven to be an effective flag for autism as measured by the longitudinal stability of CHAT results. In a study that followed 16,000 children from 18 months to 3.5 years, Baron-Cohen et al. (1996) found that a deficit in protodeclarative pointing was one of the three best predictors of autism. (Gaze monitoring and pretend play also predicted autism.) An early deficit in protodeclarative pointing is consistent with the social impairments that characterize autism, and other highly social gestures are likely to be impaired early as well.

Other protodeclarative gestures: giving and showing

Though pointing is a crucial indicator of social functioning, other gestures such as showing and giving also reflect social skills and interest. *Showing* refers to a person's arm extending toward another person's face while holding an object (Wetherby, Yonclas, & Btyan, 1989; Stone et al., 1997). *Giving* refers to placing an object in another person's hand or pushing an object at least halfway toward another person (Wetherby et al., 1989; Stone et al., 1997). While many instances of giving serve protodeclarative purposes (e.g., giving an interesting object to a parent), others serve protoimperative purposes (e.g., handing a jar to a parent to get it opened) or occur in the context of sharing (e.g., giving a cracker or a toy to another person). Both showing and giving emerge around eight to ten months in typically developing children, concurrently with pointing (Capone & McGregor, 2004; Bates & Dick, 2004).

Giving and showing in Autism Spectrum Disorder

Little is known about the early development of giving and showing in children with ASD, but studies of older children with ASD have found deficits in both types of gestures. Children with autism at 52 months were less

likely to give, take, and show objects in response to adult cues compared to developmentally delayed children (Leekam & Ramsden, 2006). At 49 months, children with ASD showed similar differences in frequency of giving (DiLavore, Lord, & Rutter, 1995). Children with ASD demonstrated fewer "gestural joint attention skills" (defined as pointing and showing) at 45 months when compared to children with mental retardation matched on language, mental age, or IQ (Mundy, Sigman, & Kasari, 1990). Wetherby, Watt, Morgan, and Shumway (2007) found that a strong predictor of ASD between 24 and 36 months was "inventory of gesture," which assessed the children's use of eight gestures and included showing, pointing and giving. In an earlier prospective study, the authors found that limited showing was a red flag that differentiated children with ASD from both developmentally delayed and typically developing children. Frequency of pointing, however, did not differ between children with ASD and those with developmental delays. When Stone et al. (1997) measured giving and showing within the same age range, a significant difference was found for showing but not for giving between children with ASD and developmentally delayed and language-impaired children. These studies indicate that multiple aspects of nonverbal communication may be impaired in ASD and suggest that nonverbal gestures beyond pointing may also serve as early indicators of the disorder.

Importance of studying giving and showing in diagnosis

More data is needed to disambiguate ASD from other symptomologies, especially when diagnosed early. A longitudinal study found that of 469 Danish people diagnosed with a developmental language delay as children, 10 later carried a diagnosis of ASD in adulthood (Mouridsen & Hauschild, 2009). Deficits in social gestures may provide an extra indicator in these ambiguous situations. Wetherby and Prizant (2002) found that a deficit in showing differentiated children with ASD from both developmentally delayed and typically developing children. Protodeclarative pointing, however, did not distinguish children with ASD from developmentally delayed children. A similar problem was also noted in the Baron-Cohen et al. (1996) study on 18-month olds. These results suggest that other signs beyond pointing are necessary for distinguishing between these groups. Showing and giving, which depend heavily on social skills, may serve this purpose.

Diagnosis in high-risk populations

Distinguishing ASD from similar non-autistic disorders is of particular concern when diagnosing children at high risk for ASD. Children with an older sibling with ASD have increased genetic risk for developing the disorder (approximately 20%; Rogers, 2009), and a substantial proportion of the high-risk group (18-30%; Landa, Holman, & Garrett-Mayer, 2007) do not meet diagnostic criteria for ASD but do exhibit other symptoms such as linguistic disorders and impaired social, cognitive, and linguistic skills. The siblings without ASD often show early deficits similar to children with ASD, such as diminished pointing, social smiling, receptive language, and overall use of gestures and words (Toth, Dawson, Meltzoff, Greenson, & Fein, 2007). Since the population of infants at high risk is often closely monitored by clinicians for signs of

ASD, robust indicators are needed to differentiate those who develop non-autistic symptomology from those who develop traditional ASD. Studies have shown that pointing may not serve this purpose because deficits in pointing are present both in children with ASD and in those with developmental delays (Toth et al., 2007). Investigation of other protodeclarative gestures such as giving and showing may help disambiguate the two groups (Leekam & Ramsden, 2006; Stone et al., 1997; Wetherby et al., 2007).

Gestures tend not to be studied individually

Although gestures such as giving and showing may provide valuable diagnostic information, they have yet to be carefully studied. Studies that do include measurements of showing or giving usually analyze "gesturing" as a single construct or focus on pointing and do not analyze giving or showing independently. Charman, Drew, Baird, and Baird (2003) investigated summary scores on the MacArthur Communication Development Inventories (CDI) of "early" and "late" gestures but did not analyze differences in scores on specific items that assess showing and giving independently. Mitchell et al. (2006) replicated the same study with a younger sample and more control groups, but again focused on "early" and "late" gestures instead of the individual development of giving or showing. These studies yielded interesting results about global gesturing but did not explore profiles of specific gestures such as giving and showing. Data on the development of individual gestures could inform the dialog on very early indicators and aid in distinguishing language disorders from ASD.

STUDY OVERVIEW

The present study explores the development of pointing, showing, and giving at twelve months in children who were later diagnosed with ASD. The children were at high risk (approximately 20%; Rogers, 2009) for developing the disorder in comparison to the average American child (< 1%; CDC, 2009) because they had an older sibling with an ASD. Studying the high-risk population from birth yields insights into the early development of ASD (Zwaigenbaum et al., 2009). This study assessed early gesture development in infants who later showed symptoms of ASD or language delays at 24 months. The main objective was to investigate the individual profiles of giving, showing and pointing, and to compare the patterns of development in children with ASD to patterns among typically developing children as well as those with language delays.

We hypothesize the following:

(1) ASD infants will generally gesture less frequently than language delayed (LD) and typically developing (TD) infants.

(2) Pointing will be diminished in the ASD and LD groups compared to the TD group. Showing and giving will be diminished in the ASD group alone.

(3) Across development from nine to twelve months, children who are later diagnosed with ASD will exhibit fewer gains in gestures compared to children in the LD and TD groups.

Method

Participants

Forty participants were recruited from an ongoing prospective longitudinal study of early social-cognitive development. Twenty-eight of the infants had an older sibling with ASD and therefore were at increased genetic risk for developing it. Gesture proficiency was measured at nine and twelve months using the MacArthur Communication Development Inventories (CDI). Based on assessments at 24 months, infants in the high-risk for ASD group were divided into three sub-groups: those diagnosed with ASD (ASD-HR; n = 8); language and communication delay but not ASD (LD-HR; *n* = 14); and infants developing typically (TD-HR; *n* = 8). A fourth group included typically developing children with no familial history of ASD and thus considered at low-risk for developing the disorder (TD-LR; *n* = 12). Males represented 42% of the TD-LR group, compared to 33% of the TD-HR group, 93% of the LD-HR group, and 88% of the ASD group (see Table 1). Participants in the TD-LR, TD-HR, and LD-HR groups did not differ in level of development as measured by the five scales of the Mullen Scales of Early Learning (receptive language, expressive language, visual reception, fine motor skills, and gross motor skills), with the exception of visual reception (see Table 1). Participants in the ASD-HR and LD-HR groups also had comparable scores on the developmental test (see Table 1), but infants with ASD had lower scores than the TD-HR group on expressive language scales and lower scores than the TD-LR group on expressive language, visual reception, and fine motor scales. As expected, children in ASD-HR group had significantly higher total scores on the ADOS-T scale than the remaining diagnostic groups at 24 months (see Table 1).

ASD Inclusion Criteria

Eight participants at risk for ASD received a provisional diagnosis of ASD at 24 months. A team of two experienced clinicians evaluated each child over two days and assigned a provisional diagnosis based on DSM-IV criteria modified for children under the age of three (Chawarska & Volkmar, 2005) and performance on the Autism Diagnostic Observation Schedule–Toddler Module (Lord et al., 2000), Vineland Adaptive Behaviors Schedule-Expanded (Sparrow, Balla, & Cicchetti, 1984), and Mullen Scales of Early Learning (Mullen, 1995).

Language Delay Inclusion Criteria

The LD group included fourteen participants at risk for ASD who showed significant communicative delays at 24 month assessments but did not receive a diagnosis of ASD. Five participants received a clinical diagnosis of a language-related disorder (expressive language disorder, receptive language disorder, developmental delay, and two with mixed language disorder). In nine participants, expert clinicians identified impaired communicative functioning and social deficits in the subclinical range (Broader Autism Phenotype, BAP).

Procedure

Parents reported their infants' gesture use using the MacArthur Communication and Development Inventories (CDI), which provides a comprehensive profile of a child's communication skills (Fenson et al., 1993). The present study used responses on the Words and Gestures version in the subsection "First Communicative Gestures." This

Table 1. Group means (standard deviation) on Mullen Scales of Early Learning and ADOS-T tests at 24 months.

	Sample characterization				
Measure	ASD-HR	LD-HR	TD-HR	TD-LR	p-value
N	8	14	6	12	
% male	88%	93%	33%	42%	
Mullen (12 months)					
Visual Reception	51.1 (3.6)[a]	49.9 (11.7)[a]	59.9 (7.1)[ab]	62.0 (6.8)[b]	.004
Gross Motor	39.8 (7.0)	46.2 (11.5)	48.0 (12.5)	44.0 (11.9)	.510
Fine Motor	48.9 (8.3)[a]	59.1 (9.3)[ab]	60.5 (9.8)[ab]	60.6 (10.4)[b]	.047
Receptive Language	41.1 (11.2)	39.6 (8.9)	45.2 (9.7)	48.2 (6.8)	.104
Expressive Language	33.6 (7.0)[a]	42.0 (6.5)[ab]	49.9 (18.6)[b]	49.7 (10.6)[b]	.009
ADOS-T (12 months)					
Social Affect	14.9 (3.3)[a]	10.9 (3.7)[ab]	9.2 (5.9)[bc]	6.5 (3.0)[c]	< .001
Restricted & Repetitive Behavior	2.4 (1.9)	1.5 (1.6)	.8 (1.0)	7.5 (3.6)	.161
Total Score (SA + RRB)	17.3 (4.2)[a]	12.4 (4.4)[ab]	10.0 (6.5)[bc]	7.5 (3.6)[c]	< .001
ADOS-T (24 months)					
Social Affect	11.9 (5.7)[a]	7.3 (3.8)[ab]	3.0 (2.7)[bc]	2.3 (1.9)[c]	< .001
Restricted & Repetitive Behavior	2.1 (1.2)[a]	1.1 (1.1)[ab]	.6 (.6)[b]	.7 (.7)[b]	.019
Total Score (SA + RRB)	14.0 (6.4)[a]	8.3 (4.2)[b]	3.6 (3.2)[bc]	3.0 (2.1)[c]	< .001

Values within each row with different superscripts differ at $p < .05$ with Tukey's Honestly Significant Difference correction for multiple comparisons.

sub-section contains 12 items and asks parents to report how frequently their child uses particular gestures. Sample gestures assessed include showing, giving, pointing, waving goodbye, and nodding. Parents read a description of the gesture, such as "Extends arm to show you something he/she is holding" and then checked whether the child performed the gestures "often," "sometimes," or "not yet." For the purposes of this study, responses were coded according to whether the child performed the gesture proficiently, indicated by frequent use and a response of "often." This coding scheme was employed because the gestures of interest were those used functionally, not

Table 2. Proportion of children in ASD-HR, LD-HR, TD-HR, and TD-LR groups demonstrating gesture "often" based on parent report. Gestures that occurred infrequently were excluded from analysis.

	9 months				12 months			
	ASD-HR	LD-HR	TD-HR	TD-LR	ASD-HR	LD-HR	TD-HR	TD-LR
Show	13%	14%	0%	25%	38%	79%	67%	75%
Point	0%	0%	17%	25%	25%	36%	33%	58%
Give	0%	7%	0%	8%	50%	64%	67%	50%
Pick me up	25%	64%	50%	58%	63%	50%	50%	75%
Wave Bye	0%	7%	0%	0%	13%	21%	50%	58%
Shake Head 'No'	0%	7%	0%	0%	0%	7%	17%	17%
Nod Head 'Yes'	0%	0%	0%	0%	0%	7%	0%	0%
Hush' finger	0%	0%	0%	0%	0%	0%	0%	0%
Open/close hand	0%	0%	0%	8%	0%	21%	17%	0%
Blow kisses	0%	0%	0%	0%	0%	14%	0%	8%
Smack lips 'Yum'	13%	7%	0%	17%	13%	7%	0%	17%
Shrug 'All gone'	0%	0%	0%	0%	0%	0%	0%	8%

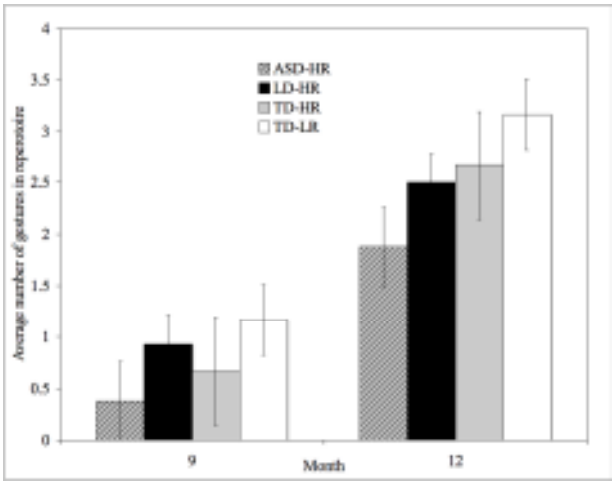


Figure 1. Average number of functional gestures (± 1 SE) in child's repertoire based on parent report for children at high risk for ASD and diagnosed with ASD, LD, or typically-developing and typically-developing children at low risk. Gestures that occurred infrequently (ie., exhibited often by fewer than 25% of all groups) were omitted.

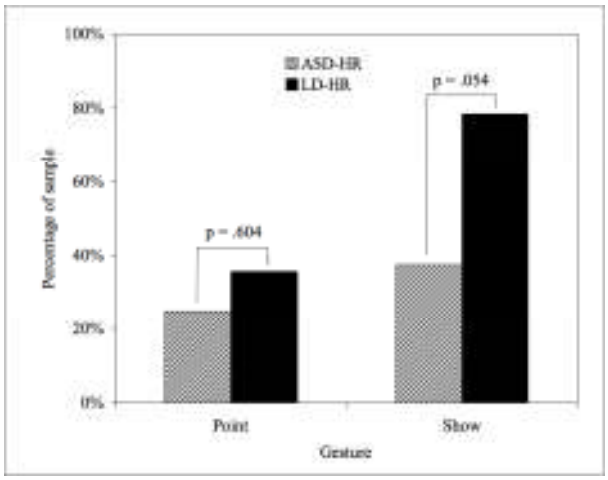


Figure 2. Proportion of children with ASD and children with LD who often point or show at 12 months. At twelve months, parents of children at risk for developing ASD reported how frequently their child exhibited showing or pointing. A larger proportion of children who were diagnosed with a language delay at 24 months often showed objects than children who were later diagnosed with ASD. There was less difference between groups in the proportion that frequently pointed at 9 months.

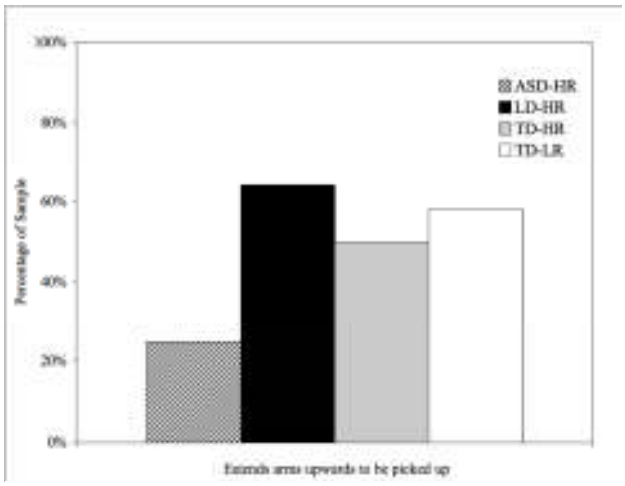


Figure 3. Proportion of ASD-HR, LD-HR, TD-HR, and TD-LR children who often reach to be picked up at 9 months.

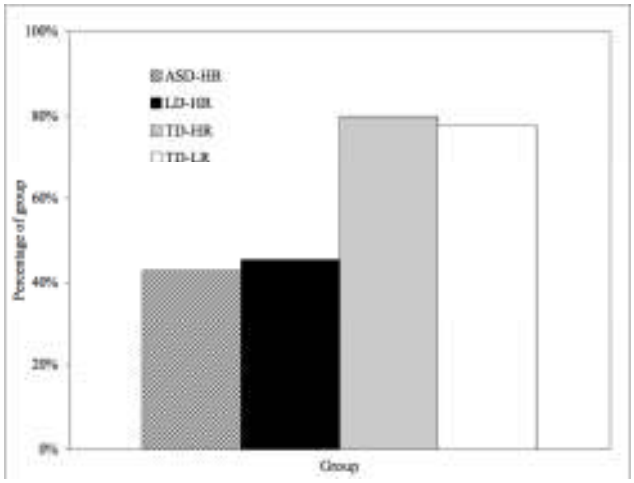


Figure 4. Proportion of ASD-HR, LD-HR, TD-HR, and TD-LR children who did not demonstrate functional use of any of the five gestures of interest (point, show, give, reach to be picked up, and waves goodbye) at 9 months and did develop the gesture by 12 months.

those still in development and used only sometimes. This study examined participants' scores on each individual gesture, which is notable because previous studies combined all individual gestures into a single summary score (Charman et al., 2003; Mitchell et al., 2006). Each parent completed the CDI as part of a larger battery of surveys mailed 2 to 3 weeks prior to their child's scheduled 9- and 12-month visits. Parents returned the surveys at the time of the visit, which was scheduled within a month to the date of the chronological age (with two exceptions when the child could not visit until 2 and 3 months following their birthday).

Results

General Frequency of Gesturing

Frequencies of Gestures

At both nine and twelve months, seven of the twelve gestures were reported as occurring "not yet" or only "sometimes" in at least 89.8% of children at nine and twelve months across all groups. Therefore, these gestures were excluded from subsequent analyses. At nine months, the only gesture that occurred frequently enough by all groups to allow for analysis was "extending one's arms to be picked up." At twelve months, infants demonstrated five gestures frequently enough for analysis: pointing, showing, giving, waving goodbye, and extending one's arms to be picked up.

Limited frequency of gestures in children with ASD

A visual inspection of Figure 1 suggests that children with ASD generally gestured less than other children at both time points. A mixed models 4 (diagnostic group: ASD-HR, LD-HR, TD-HR, TD-LR) × 2 (time: 9 months, 12 months) ANOVA on the total number of gestures with repeated measures on the last factor indicated a significant effect of time, $F(1, 36) = 38.99$, $p < .001$, and no effect of diagnosis $F(3, 36) = 1.47$, $p < .238$, or diagnosis × time interaction $F(3, 36) = .24$, $p < .868$. To test our specific hypothesis regarding between-group differences, we followed up with a post-hoc test comparing the four groups. Infants with ASD had significantly fewer gestures than infants in the TD-LR group $t(36) = -2.09$, $p < .043$, $d = -.70$, but differences between other groups were not significant (ASD-HR versus LD-HR: $p < .230$; ASD-HR versus TD-HR: $p < .363$)).

Patterns in Pointing, Showing and Giving at Twelve Months

Pointing

Children with autism pointed less frequently than all other children at twelve months. In comparison to the TD-LR group, the group with ASD pointed less than half as frequently (58.3% and 25.0%, respectively), although the difference failed to reach significance, chi-square $(1, 20) = 2.16$, $p = .142$. Children in the ASD-HR and LD-HR groups pointed at similar frequencies (25.0% and 35.7%, respectively), chi-square $(1, 22) = .27$, $p = .604$.

Showing

In contrast to pointing, the frequency of showing differed significantly between the ASD-HR and LD-HR groups at twelve months. A chi-square test indicated a marginally significant difference between groups in the proportion of participants demonstrating proficiency in the showing

gesture (ASD-HR: 37.5%; LD-HR: 78.6%; chi-square (1, 22) = 3.71, p = .054). The frequency of showing was similar across the LD-HR, TD-LR (75.0%), and TD-HR (66.7%) groups.

A binary logistic regression was conducted to determine which gestures (i.e., pointing, showing, giving, waving goodbye, and extending arms to be picked up) at 12 months best predicted that a twelve-month infant with some communicative delays would develop ASD instead of a language delay. We fitted a series of models to the data and began by including all five gestures with high frequency at twelve months as predictors of diagnostic outcome. The best fit model included only the showing and pointing gestures, and showing was found to be a stronger predictor of diagnosis than pointing (showing: b = 2.30, p < .077; pointing: b = -.88, p < .522). The odds ratio, which indicates the likelihood of belonging to a group, indicated that showing, not pointing, was a stronger predictor of diagnosis (showing: e^b = 10.00; pointing: e^b = .42). The odds of belonging to the LD-HR group are 10 times greater than the odds of belonging to the ASD-HR group, given that the infant is proficient in showing. Thus, high-risk infants who engage in frequent showing at 12 months are less likely to be diagnosed with ASD at 24 months than high-risk infants who are not proficient in the showing gesture at twelve months. While the presence of either pointing or showing at 12 months increases the odds that the infant belongs to the LD group, those odds are much higher when the infants use functional showing (10.00) as opposed to functional pointing (.42).

Giving

The frequency of giving did not differ significantly between groups at twelve months. Parents reported the frequency "often" for 50.0% of children with ASD, 64.3% of LD-HR children, 66.7% of TD-HR children, and 50.0% of TD-LR children.

Exploratory Analyses of Gesturing at Nine Months

Nine-month old infants exhibited low frequencies of most gestures. The only notable result was the proportion of parents who reported that their child "often" extended his or her arms upward to signal a wish to be picked up. In the typically developing groups, 58.3% of parents of low-risk infants and 50.0% of parents of high-risk infants reported this frequency, as did 64.3% of parents in the LD-HR group. A chi-square analysis suggests a trend between these groups and the ASD-HR group, where the proportion was only 25.0%, chi-square (1, 40) = 3.03, p = .082.

Group Differences in Acquisition of Gestures

The children with ASD exhibited fewer gains in gestures between nine and twelve months compared to the two typically developing groups. Of the infants who did not exhibit any of the five gestures of interest (the most frequent gestures) at nine months, only 43% of the ASD-HR group developed pointing or showing by twelve months, in contrast to 80% of the TD-HR group and 78% of the TD-LR group. Although these differences between groups are not significant, they illustrate a pattern of slower gesture acquisition in children with ASD.

DISCUSSION

The results of this study reaffirm previous findings that frequency and repertoire of gestures can provide important insights for clinicians assessing children at risk for ASD in their first year. The most notable finding is the potential value of the showing gesture as a predictor of ASD in high-risk children. Due to the prospective nature of the study, the number of infants included in the analysis was low, which contributed to the failure of some observed effects to reach statistical significance. However, the study is ongoing and we are currently analyzing data from a larger sample. Thus, the analyses presented in this paper are necessarily preliminary. The results thus far indicate the following:

(1) Infants who are later diagnosed with ASD may gesture less frequently than high-risk infants who later develop language delays as well as typically developing infants from both high- and low-risk samples.

(2) At twelve months, a lack of showing may differentiate ASD-HR from LD-HR better than a lack of pointing.

(3) At nine months, the only gesture demonstrated by any group at a high frequency was extending one's arms upward to be picked up, which was diminished in the ASD-HR group compared to all other groups.

(4) Children with ASD may acquire fewer gestures between nine and twelve months compared to typically developing children.

Each of these results will now be addressed in turn.

Children with ASD gesture less frequently

Data suggest that children with ASD exhibit fewer gestures than typically developing children at low risk at both twelve and nine months. This pattern replicates the results of past studies that measured gestures with the CDI (Charman et al., 2003; Mitchell et al., 2006) and with other measures (Carpenter, Pennington, & Rogers, 2002; DiLavore et al., 1995; Leekam & Ramsden, 2006; Mundy et al., 1990; Stone et al., 1997; Wetherby et al., 2007). ASD impairs communication, so it follows that early nonverbal communicative forms such as gestures should also be impaired.

Value of the showing gesture

The frequency of showing might differentiate high-risk infants who later develop ASD from those who will develop language delays or mild social difficulties. At twelve months, both the ASD-HR and LD-HR groups lacked proficiency in pointing. However, only the ASD-HR group exhibited limited frequency of showing. In fact, in this study, if a parent reported at twelve months that a child with communication difficulties and at risk for ASD often showed objects and toys, the calculated odds that the child would be diagnosed with LD rather than ASD were ten times greater at 24 months. Proficient pointing, however, did not strongly predict LD over ASD. This difference between pointing and showing could prove useful for clinicians during the diagnostic process. Infants with increased genetic risk for developing ASD, such as those in this study, are often monitored closely for early signs of the disorder. The high-risk infants in the LD group did not receive a diagnosis of ASD by a team of expert clini-

cians at 24 months, and early behaviors that distinguish them from the high-risk infants who were diagnosed with ASD could help clinicians diagnose earlier and more effectively.

In addition, if confirmed in the final analysis, our finding on decreased frequency of showing in infants with ASD could contribute to public discussion about early signs. Public education about autism traditionally emphasizes pointing as the hallmark gesture that is absent in autism. However, ASD impairs pointing in a complex way that is difficult to communicate in a few bullet points on a waiting room pamphlet. Protoimperative pointing remains intact in ASD, and only protodeclarative pointing, or pointing to an object or event of shared interest, is impaired (Baron-Cohen, 1989; Camaioni et al., 1997; Camaioni et al., 2003; Itoh, 2000). Parents who are aware of the commonly accepted "first signs" and who often witness their child point protoimperatively may falsely discount ASD as an explanation for their child's social difficulties. The showing gesture may offer a clearer alternative to pointing as a warning sign for ASD, and the two gestures could complement each other in educational materials on early signs. Unlike pointing, showing is used only to initiate a shared social experience and not to indicate a need. A lack of showing may be easier for some parents to identify than a lack of pointing, which refers only to protodeclarative pointing and excludes protoimperative pointing. In contrast, pointing may be a more salient gesture because it is marked by a very specific hand posture. Publicity about the early signs of ASD may be more effective if both pointing and showing are used. While the former is more salient, the latter is more accurate.

Giving

Contrary to our hypothesis, the giving gesture did not uniquely characterize the ASD group. In fact, similar proportions of children across all groups demonstrated proficiency in giving. If parents report data that accurately reflects child behavior, it is remarkable that infants with ASD do not exhibit deficits in giving, as previous research has yielded contradictory results on the impairment of giving later in childhood (impaired: DiLavore et al., 1995; Mundy et al., 1986; not impaired: Stone et al., 1997). Our results may reflect the versatility of the giving gesture; children may give to share (protodeclarative) or to request something, such as help with opening a toy (protoimperative).

Reaching to be picked up

At nine months, few infants demonstrated many gestures proficiently, with the exception of extending their arms upward to be picked up. This result mirrors Kanner's (1943) original remarks about children with autism, whom he claimed could be identified in part by their failure to anticipate being picked up by their parents. It was the only gesture demonstrated frequently by more than a quarter of any group at nine months, and therefore, as Kanner indicated, it could be used to distinguish children with autism from typically developing children.

Children with ASD develop fewer gestures between nine and twelve months

The data suggest that children with ASD may develop fewer gestures between nine and twelve months compared to typically developing children. Previous studies indicate that gestures are impaired in 18-month old infants who later develop ASD (Baron-

Cohen et al., 1996; Mitchell et al., 2006), but little is understood about the development of gestures in ASD during this time period.

Summary and Future Directions

The data suggest that children with ASD exhibit fewer gestures overall at nine and twelve months. In addition, it suggests that high-risk infants whose parents report that they often show objects or toys at twelve months are less likely to be diagnosed with ASD, but might show less severe impairments involving language developing and mild social difficulties at 24 months. At nine months, children with ASD extended their arms to be picked up half as frequently as typically developing children and children with language delays, regardless of risk status. Data also suggest that children with ASD and LD may acquire fewer gestures between nine and twelve months than their typically developing peers.

Further investigation with a larger sample is needed to explore whether these trends reflect true differences between groups. The larger subject pool from which these participants were recruited is maturing, and the data is currently being analyzed for more children from this unique population.

Further research is also needed to explore the reliability of parent report data from the CDI. Parents were not trained and may vary in their interpretation of the CDI categories of frequency of "not yet," "sometimes," and "often." We attempted to compensate with a coding scheme that combined "not yet" and "sometimes" responses, which both indicate relatively low frequencies of the behavior; however, reliability issues likely persist. Clinician assessments of children using the CDI could be used to validate parent reports. In addition, during diagnostic evaluation, clinicians could assess the type, frequency, and communicative function of gestures used, which would help elucidate the ambiguities of gesture classification by parents*. However, clinicians lack the extensive perspective of parents, whose reports still contribute to the discussion of early signs of ASD. Parents, not trained observers or clinicians, daily witness their child play and communicate in a naturalistic setting and are in the best position to identify ASD early. Therefore, the publicly advertised "early signs of ASD" should be easy for parents to utilize, and therefore should be based on research about which signs were actually observed by parents of children with ASD.

One implication of these results concerns future studies of gesture in ASD. While studies often combine all gestures and report a composite score (Charman et al., 2003; Leekam, 2006; Mitchell et al., 2006; Mundy et al., 1990; Wetherby et al., 2007), the results of this study indicate that the data on individual gestures should not be overlooked. Trends differed between specific gestures such as pointing, showing, and giving, and a summary score would have ignored these differences. Examining

* Parents may struggle to interpret descriptions of gestures on the CDI. For example, the CDI describes the reaching gesture as "Requests something by extending arm and opening and closing hand." Few parents reported proficiency in this item at nine or twelve months, although reaching is known to emerge between ten and twelve months (Capone & McGregor, 2004). These results suggest that the questionnaire may not accurately capture infant gesturing through parent report.

the CDI data on gestures in a new way yielded a potential new warning sign for ASD.

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (revised 4th ed.). Washington, DC: Author.

Baron-Cohen, S. (1989). Perceptual role-taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology*, *7*, 113-127.

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., & Nightingale, N. (1996). Psychological markers in the detection of autism in infancy in a large population. *British Journal of Psychiatry, 168*, 158-163.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *British Journal of Psychiatry,161*, 839-843.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology, 40*, 293-310.

Capone, N., & McGregor, K. K. (2004). Gesture development: A review for clinical and research practices. Journal of Speech, Language, and Hearing Research, 47, 173-186

Camaioni, L., Perucchini, P., Muratori, F., Parrinini, B., & Cesari, A. (2003). The communicative use of pointing in autism: Developmental profile and factors related to change. *European Psychiatry, 18*, 6-12.

Camaioni, L., Perucchini, P., Muratori, F., & Milone, A. (1997). A longitudinal examination of the communicative gestures deficit in young children with autism. *Journal of Autism and Developmental Disorders, 27*, 715-725.

Carpenter, M., Pennington, B. R., & Rogers, S. J. (2002). Interrelations among social-cognitive skills in young children with autism. *Journal of Autism and Developmental Disorders, 32*, 91-106.

Centers for Disease Control and Prevention. (2009). Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, United States, 2006. *Morbidity and Mortality Weekly Report*, Dec. 18, 2009.

Chawarska, K., Paul, R., Klin, A., Hannigen, S., Dichtel, L.E., & Volkmar, F. (2007). Parental recognition of developmental problems in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 62-72.

Chawarska, K., & Volkmar, F. (2005). Autism in infants and toddlers. In F. Volkmar, Klin, A., & Paul, R. (Eds.), *Handbook of Autism and Pervasive Developmental Disorders*. New York: Wiley.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30*, 213-236.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology, 40*, 271-283.

DiLavore, P. C., Lord, C., & Rutter, M. (1995). The pre-linguistic autism diagnostic observation schedule. *Journal of Autism and Developmental Disorders*, 25, 355-379.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

Howlin, P. (2005). Outcomes in autism spectrum disorders. In F. Volkmar, Paul, R., Klin, A., & Cohen, D., (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (3rd ed.; vol. A, pp. 201-222). New York: Wiley.

Itoh, H. (2000). Development of pointing behavior in young autistic children and children with other developmental disorders. *Japanese Journal of Child and Adolescent Psychiatry, 41*, 57-70.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Kita, S. (2003). *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.

Klin, A., Chawarska, K., Rubin, E., & Volkmar, F. (2004). Clinical assessment of young children at risk for autism. In R. DelCarmen-Wiggins, & Carter, A. (Eds.), *Handbook of Infant, Toddler, and Preschool Mental Health Assessment* (pp. 311-336). New York: Oxford University Press.

Landa, R., Holman, K. C., & Garrett-Mayer, E. (2007). Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of General Psychiatry, 64*, 853-864.

Leekam, S. R., & Ramsden, C. A. H. (2006). Dyadic orienting and joint attention in preschool children with autism. *Journal of Autism and Developmental Disorders, 36*, 185-197.

Loveland, K., & Landry, S. (1986). Joint attention and language in autism and developmental language delay. *Journal of Autism and Developmental Disorders, 16*, 335-349.

Lord, C., Risi, S., & Pickles, A. (2004). Trajectory of language development in autism spectrum disorders. In M. Rice & Warren, S., (Eds.), *Developmental Language Disorders: From Phenotypes to Etiologies* (pp. 7-29). Mahwah, NJ: Erlbaum.

Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., & DiLavore, P.C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule–Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*, 205-223.

Luyster, R., Lopez, C., & Lord, C. (2007). Characterizing communicative development in children referred for Autism Spectrum Disorders using the MacArthur-Bates Communicative Development Inventory (CDI). *Journal of Child Language, 34*, 623-654.

Mitchell, S., Brian, J., Zwaigenbaum, L., Roberts, W., Szatmari, P., & Smith, I., & Bryson, S. (2006). Early language and communication development of infants later diagnosed with autism spectrum disorder. *Developmental and Behavioral Pediatrics, 27*, S69-S78.

Mouridsen, S. E., & Hauschild, K-M. (2009). A longitudinal study of autism spectrum disorders in individuals diagnosed with a developmental language disorder as children. *Child: Care, Health and Development, 35*, 691-697.

Mullen, E. (1995). *Mullen Scales of Early Learning* (AGS ed.). Circle Pines, MN: American Guidance Service.

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*, 115-128.

Mundy, P., Sigman, M. D., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of Child Psychology and Psychiatry, 27*, 657-669.

Paul, R. (2007). Communication and its development in autism spectrum disorders. In F. Volkmar (Ed.), *Autism and Pervasive Developmental Disorders* (2nd ed.). Cambridge: Cambridge University Press.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*, 131-144.

Roger, S. J. (2009). What are infant siblings teaching us about autism in infancy? *Autism Research, 2*, 125-137.

Sparrow, S.S., Balla, D.A., & Cicchetti, D.V. (1984). *Vineland Adaptive Behavior Scales: Interview edition, survey form manual*. Circle Pines, MN: American Guidance Service.

Stone, W. L., Ousley, O. Y, Yoder, P. J., Hogan, K. L., & Hepburn, S. L. (1997). Nonverbal communication in two- and three-year-old children in autism. *Journal of Autism and Developmental Disorders, 27*, 677-696.

Tager-Flusber, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, Paul, R., Klin, A., & Cohen, D.J. (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (pp. 335-364). New York: Wiley.

Tomasello, M. (2007). If they're so good at grammar, then why don't they talk? Hints from apes' and humans' use of gestures. *Language, Learning, and Development, 3*, 133-156.

Toth, K., Dawson, G., Meltzoff, A.N., Greenson, J., & Fein, D. (2007). Early social, imitation, play, and language abilities of young non-autistic siblings of children with autism. *Journal of Autism and Developmental Disorders, 37*, 145-157.

Volkmar, F., Klin, A., Paul, R., & Cohen, D. (2005). *Handbook of autism and pervasive developmental disorders.* New York: Wiley.

Wetherby, A. M. (1986). Ontogeny of communicative functions in autism. *Journal of Autism and Developmental Disorders, 16*, 295-316.

Wetherby, A. M., & Prizant, B. M. (2002). *CSBS DP Manual: Communication and Symbolic Behavior Scales Developmental Profile* (1st Normed Ed.). Baltimore: Paul H. Brookes Publishing.

Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders, 37*, 960-975.

Wetherby, A. M., Yonclas, D. G., & Bryan, A. A. (1989). Communicative profiles of preschool children with handicaps: Implications for early identification. *Journal of Speech and Hearing Disorders, 54*, 148-158.

Zwaigenbaum, L., Bryson, S., Lord, C., Rogers, S. J., Carter, A., Carver, L., Chawarska, K., Constantino, J., & Dawson, G., Dobkins, K., Fein, D., Iverson, J., Klin, A., Landa, R., Messinger, D., Ozonoff, S., Sigman, M., Stone, W., Tager-Flusberg, H., & Yirmiya, N. (2009). Clinical assessment and management of toddlers with suspected ASD: Insights from studies of high-risk infants. *Pediatrics, 123*, 1383-1391.

The Economics of Alcohol Consumption: Exploring the Relationship between Sensitivity to the Price of Alcohol and Behavioral Consequences

Emily H. Harris

YALE UNIVERSITY

A common method of controlling alcohol consumption at the population level is the imposition of sin taxes, which are additional taxes on "sinful" consumer goods like cigarettes and alcohol. Despite evidence that individuals who are insensitive to increases in the price of alcohol are more likely to be heavier drinkers, little research attention has been given to understanding this relationship. The current study was designed (1) to evaluate the psychometric properties of a measure of price sensitivity, (2) to explore the relationships among price sensitivity, alcohol use and the experience of alcohol-related problems, and constructs related to alcohol use including alcohol expectancies and drinking motives, and (3) to examine alcohol expectancies and drinking motives as potential mediators of the relationship between price insensitivity and alcohol use/problems. The results suggest that both alcohol expectancies and drinking motives partially mediate the relationship between price insensitivity to alcohol and drinking/problems. Because insensitivity to price is related to increased consumption of alcohol, these findings may have important implications for government taxation policy, suggesting that rather than solely focusing on price, interventions may benefit from addressing people's ideas about what happens when they drink and their motives for drinking.

INTRODUCTION

Alcohol consumption is prevalent in the United States and across the globe, with over two billion people worldwide reporting use (WHO, 2007). While there is evidence to suggest that consuming alcohol in moderation is not harmful (Ashley, Ferrence, Room, & Single, 1994), excessive drinking, including heavy drinking (averaging > 1 drink per day for women or > 2 drinks for men) and binge drinking (consuming at least four drinks in a single sitting for women and five drinks for men), has been linked to serious health problems (e.g., liver cirrhosis, cancer), automobile and other accidents, and death (Centers for Disease Control, 2008). Ultimately, excessive alcohol use is the fifth most common cause of death in the world (World Health Organization, 2007) and is responsible for nearly 80,000 deaths each year in the United States alone (CDC, 2008). Despite the numerous negative consequences associated with excessive drinking, there is a prevalent perception of personal immunity from experiencing negative alcohol-related outcomes (e.g., Hansen, Raynor, & Wolkenstein, 1991).

Neither binge drinking nor heavy drinking is a rare occurrence. Recent surveys report that over 50% of American adults drank alcohol in the past 30 days, with 5.2% engaging in heavy drinking and 15.8% engaging in binge drinking (CDC, 2008). The prevalence of excessive drinking is even higher among young drinkers; over 40% of college students binge drink (Wechsler, Dowdall, Maenner, Gledhill-Hoyt, & Lee, 1998). When taken in concert, the widespread use of alcohol, the plethora of detrimental outcomes, and the sense of personal invulnerability to the negative consequences of consumption make excessive alcohol use a particularly challenging public health problem. As such, there are numerous national and international organizations (e.g., the CDC and the WHO) dedicated to exploring methods by which alcohol consumption can be controlled.

While many interventions designed to decrease alcohol consumption have been proposed and implemented, there has been strong and consistent support for the effectiveness of increasing the price of alcohol, typically achieved through taxation, as a large scale intervention to modulate alcohol use (CDC, 2009; Guide to Community, 2006). In the United States, the imposition of "sin taxes" on goods or services that are not illegal but are commonly viewed in society as vices (e.g., alcohol and cigarettes; Grossman, Sindelar, Mullahy, & Anderson, 1993) is a common practice with a lengthy history. For example, cigarettes have been taxed since the Civil War, and as early as 1791, a tax was levied on whiskey in an effort to decrease alcohol consumption and increase government revenue (Altman, 2009).

The author would like to give special thanks to Meghan E. Morean, Teresa A. Treat, and William R. Corbin. She can be reached at emily.h.harris@gmail.com.

By targeting the vices of a subsector of the population, sin taxes represent, for many, an easier and more "justifiable" way to bring in government revenue than angering the population at large through raising costs like income taxes. Thus, there has been speculation that economic motives may be at least as much a part of the attraction to sin taxes as public health concerns (Newman, 2003; Newman, 2010).

Irrespective of the true intent of sin taxes, research supports the notion that increasing taxes on alcohol does contribute to decreased alcohol consumption. A recent meta-analysis of over one hundred studies provided evidence for significant inverse relationships between price and consumption of beer (r = -.17), wine (r = -.30), and liquor (r = -.29; Wagenaar, Salois, & Komro, 2009). When the analyses were limited to a smaller subset of studies that only included heavy drinkers, the effect became negligible (r = -.01; Wagenaar et al., 2009). The results of the meta-analysis suggested that, at an aggregate (population) level, increasing alcohol price is one strategy that might be used to curtail drinking. However, the relationship between relative price insensitivity and heavy drinking suggested that, as an intervention, sin taxes may be least effective for those who are at greatest risk for negative alcohol consequences. Despite the importance of this finding, no explanations were offered for the relative insensitivity to alcohol price among heavy drinkers.

Deconstructing the relationship between insensitivity to increases in alcoholic beverage price and heavy drinking is a complex task, in part because the relationship represents a complicated marriage of psychology and economics. However, recent advances in applied behavioral economics may be particularly well-suited to helping psychologists address this relationship. Using a behavioral economics approach, Hursh (2000) found that the relative reinforcing value of alcohol can be determined through the cost an individual is willing to incur to consume alcoholic beverages. Although possible, assessing individuals' willingness to purchase alcoholic beverages at increasing costs proved quite difficult in real-world contexts. Fortunately, a series of studies conducted by Hursh and colleagues suggested that the amount of alcohol individuals are actually willing to consume is strongly related to the amount that they report being willing to consume when provided with a hypothetical drinking scenario (mean r^2 = .84; Hursh, 2000; Hursh, Raslear, Shurtleff, Bauman, & Simmons, 1998).

Having provided a sound rationale for circumventing the difficulties of assessing individuals' alcohol utility in real-world contexts, Hursh and colleagues paved the way for the development of the Alcohol Purchase Task (APT; Murphy & MacKillop, 2006). The APT assesses individuals' hypothetical willingness to purchase alcohol across a range of increasing prices. Respondents are asked to imagine a situation in which they are at a bar and must purchase any and all alcoholic beverages they intend to drink for the night. They are then asked to indicate how many drinks they would be willing to purchase across a range of increasing prices (e.g., $0.00 [free] to $9.00). Mirroring the results of the meta-analysis, studies that have assessed the relationship between price sensitivity and drinking status (i.e., heavy versus light) using the APT have shown that heavier drinkers are less sensitive to increases in price than lighter drinkers (Murphy & MacKillop 2006; Murphy & MacKillop 2007;

MacKillop et al., 2009; Wagenaar et al., 2009). Although the relationship between price insensitivity and heavier drinking has been replicated across several studies, no study to date has attempted to explain why heavy drinkers are relatively insensitive to changes in the cost of alcohol. As an important first step toward understanding this relationship, the current study examined how the APT might be related to constructs known to relate to drinking, such as alcohol expectancies and motives for drinking.

Alcohol expectancies are "if-then statements about alcohol effects that can be interpreted as a cognitive representation of one's past direct and/or indirect experiences with alcohol" (Leigh, 1999) and are typically categorized as either positive (e.g., happy, talkative) or negative (e.g., depressed, sick). Numerous studies have suggested that alcohol outcome expectancies are associated with alcohol use, usually highlighting a relationship between drinking behavior and the expectation that consuming alcohol will result in positive outcomes (e.g., Brown, Goldman, & Christiansen, 1985; Carey, 1995). However, holding beliefs that negative consequences are unlikely to result from drinking alcohol may also increase the risk for heavy drinking (e.g., Grube & Agostinelli, 1999). Differences in expectancies have been noted across gender, drinker status, and age, with younger males and heavier drinkers reporting more positive expectancies and fewer negative ones (Carey, 1995). In sum, evidence suggests that individuals with more positive (and possibly less negative) alcohol expectancies are more likely than individuals with more negative (and less positive) expectancies to be heavy drinkers.

Drinking motives, presumed to be more proximal to actual alcohol consumption than alcohol expectancies, have also been shown to impact alcohol consumption (Grant, Stewart, O'Connor, Blackwell, & Conrod, 2007). People drink to achieve certain outcomes (Cox & Klinger, 1988), and they engage in different drinking behaviors depending on what motivates their alcohol use (Cutter & O'Farrell, 1984). Blackwell & Conrod (2003) identified five primary reasons why individuals drink: for social interaction, to enhance experience, to cope with anxiety, to cope with depression, and to conform to the drinking practices of others. Grant, Stewart, and Mohr (2009) found that enhancement and coping motives are positively related to typical quantity and frequency of alcohol use and heavy drinking, social motives are positively related to frequency and quantity of alcohol consumption but not to heavy drinking, and conformity motives are negatively associated with the quantity and frequency of alcohol use and heavy drinking, but positively related to drinking problems.

Expanding upon previous research, the current study examined the relationship between the APT, alcohol expectancies, drinking motives, alcohol consumption or use (for the purposes of the current study, this is defined as quantity of drinks consumed divided by frequency of consumption), and the experience of alcohol-related problems. Also unique to this study was the inclusion of disposable income as a covariate in all analyses, because increased spending money was expected to be related to an increased willingness (via ability) to spend money on alcohol. Since it could be the case that individuals who are wealthier overall are willing to spend more money on drinks independ-

ent of their drinker status, connections established by prior work between variables of interest (e.g., expectancies, motives, consumption, etc.) may be misleading. The APT, alcohol expectancies, and drinking motives were expected to relate to one another and to both alcohol use and the experience of alcohol-related problems. Specifically, price insensitivity was expected to predict heavier alcohol use and increased alcohol-related problems. Important to establishing the utility of the APT and price sensitivity as a construct, scores on this measure were expected to evidence incremental validity in predicting alcohol use and the experience of alcohol-related problems, above and beyond that accounted for by alcohol expectancies and drinking motives. Considering that alcohol expectancies and drinking motives are related to heavy alcohol use and that insensitivity to the price of alcohol is related to heavy drinking, it may be that price insensitivity translates to heavy drinking, in part, through the expectations and motives individuals have. Therefore, preliminary analyses examined alcohol expectancies and drinking motives as potential mediators of the relationship between alcohol purchase price and alcohol use as well as the relationship between purchase price and alcohol-related problems. Individuals who reported an increased willingness to buy alcohol even at high prices were expected to be heavier drinkers and to experience more alcohol-related problems as a result of having positive expectancies and/or drinking for social, enhancement, and/or coping reasons.

THE PRESENT STUDY

Methods

Procedure

Upon arriving at the laboratory, all participants provided consent before completing an hour-long survey. The survey was comprised of a demographics questionnaire (i.e., sex, age, and spending money) and a series of additional measures designed to assess participants' engagement in a range of health-related behaviors including alcohol consumption, drug use, and eating and gambling behaviors. Participants were compensated $10.00 for their time. All procedures were approved by Yale University's Institutional Review Board.

Participants

The sample consisted of 210 participants who reported drinking at least one time in the past three months (90 males and 117 females; three people did not report a sex). Participants were recruited from Yale University and the broader New Haven community. The sample was diverse with respect to race/ethnicity: 48.9% of the participants were Caucasian, 23.8% were Asian or Asian-American, 10.2% were African American, 6% were Hispanic/Latino, and 11.1% reported being biracial or identified as "other." The mean age of the sample was 23.94 years (SD = 7.39; range 18-61). When asked to report on their drinking habits over the past three months, participants reported drinking an average of 3.93 drinks (SD = 2.86) on 2.31 days per week (SD = 1.68). The average maximum number of drinks participants reported consuming during a single drinking episode was 7.34 drinks (SD = 4.33). Participants' average monthly disposable

income ranged from $200 to $299 per month.

Measures

Alcohol Purchase Task (APT; Murphy & MacKillop, 2006). The APT is a self-report measure designed to assess participants' willingness to purchase alcoholic beverages at increasing prices. Participants are given a hypothetical situation in which they are at a bar and must purchase any and all alcoholic beverages they intend to drink for the night. While the original version of the measure assessed 14 price points ranging from $0.00 to $9.00, given the high price of alcoholic beverages in New England, the current study employed a modified version of the APT in which twenty price points were assessed, ranging from $0.00 to $15.00. From these data, the APT generates five subscales. "Intensity of demand," shortened in the current study to "intensity," represents the number of drinks an individual would consume if each drink were free. Intensity serves as a baseline indicator of drinking behaviors. "Breakpoint" represents the first price point at which consumption drops to zero (i.e., the first price point at which the participant declines to purchase alcohol). "Output Maximum," shortened to "OMAX," represents the number of drinks at which point an individuals' financial expenditure on alcohol is at its peak. "Price Maximum," or "PMAX," is the price at which expenditure is maximized. In other words, PMAX is the price point that corresponds to OMAX. Finally, "elasticity of demand," represents individuals' sensitivity to increases in the cost of alcohol. The overall elasticity of demand is the mean of each individual's price elasticity curve. Typically, each of the subscales is evaluated as an independent predictor of alcohol-related outcomes or factor analysis is used to create two factors representing price (elasticity, PMAX, and breakpoint) and consumption (intensity and OMAX). However, given the extremely strong relationships between price and consumption factor scores in the current study, performance on the APT was quantified with a single APT factor score, representing general willingness to purchase alcohol at increasing prices.

Brief Comprehensive Effects of Alcohol questionnaire (BCEOA; Ham, Stewart, Norton, & Hope, 2005). The BCEOA is a 15-item self-report measure that assesses individuals' positive and negative alcohol expectancies. It comprises four subscales: Liquid Courage (e.g., feeling brave and daring), Change in Self-Perception (e.g., feeling guilty or moody), Sexual Experiences (e.g., feeling like a better lover or enjoying sex more), and Tension Reduction (e.g., feeling peaceful or calm). Participants rate the likelihood of experiencing each of the possible alcohol effects using a five-point Likert scale that ranges from "Disagree" to "Agree." While shorter than its parent measure (the CEOA), the BCEOA has demonstrated comparable psychometric properties to the CEOA.

Modified Drinking Motives Questionnaire Revised (MDMQR; Blackwell & Conrod, 2003). The MDMQR is a self-report measure designed to assess individuals' motivations for consuming alcohol. Using a five-point Likert scale ranging from one (almost never/never) to five (almost always/always), participants are asked to indicate how often they drink for 28 different reasons. The MDMQR is based on the Drinking Motives Questionnaire Revised (DMQR; Cooper, 1994), which described a four-factor model of drinking motives. Adding to the predic-

tive validity of its parent measure (the Drinking Motives Questionnaire Revised, Cooper, 1994), the MDMQR differentiates between anxiety-related coping and depression-related coping motives (Grant, Stewart, O'Connor, Blackwell, & Conrod, 2007). In total, the modified questionnaire includes five motives: improvement of social interactions (e.g., because it is what most of my friends do when we get together), enhancement of experience (e.g., because it makes me feel good), coping with anxiety (e.g., to reduce my anxiety), coping with depression (e.g., to stop me from feeling so hopeless about the future), and conforming to the drinking practices of others (e.g., to fit in with a group I like). Social, enhancement, and coping (anxiety and depression) motives have consistently been associated with alcohol consumption, while conformity and coping have been shown to be related to the experience of drinking-related problems (e.g., Grant, Stewart, & Mohr, 2009).

Results

Overview of statistical analyses

Given concerns about multicollinearity, the APT variables were ultimately combined into a single APT factor, representing willingness to purchase alcohol at increasing prices. Next, a series of analyses were conducted to evaluate the convergent, predictive, and incremental validity of the APT. The convergent validity of the APT with alcohol expectancies, drinking motives, alcohol use, and the experience of alcohol-related problems was evaluated using bivariate correlations. Sex and spending money were significantly related to the APT, so they were included as covariates in all remaining analyses. Separate multiple regression models evaluated the predictive validity of the APT with respect to alcohol expectancies, drinking motives, alcohol use, and the experience of alcohol-related problems. Completing the validity analyses, the incremental validity of the APT was evaluated with respect to alcohol use and the experience of alcohol-related problems, controlling for the effects of alternative measures of alcohol expectancies and drinking motives.

After the basic validity of the single APT factor was evaluated, alcohol expectancies and drinking motives were examined as possible mediators of the relationships between the APT and drinking and alcohol-related problems, respectively. Mediation was tested following the procedures outlined by Baron and Kenny (1986).

Creating an APT factor

Each of the APT subscales—intensity, breakpoint, PMAX, OMAX, and elasticity—proved to be significantly correlated with the others, and in some cases the correlations raised serious concerns about multicollinearity (e.g., PMAX and breakpoint were correlated at greater than .80). Past research provided a precedent for reducing the APT variables into two factors, reflecting price ("persistence," which was composed of elasticity, PMAX, and breakpoint) and consumption ("amplitude," which was composed of Intensity and OMAX; MacKillop et al., 2009). However, the relationship between the price and consumption factors in the current study was extremely strong—several of the APT variables loaded onto both factors—and a two factor model was unable to converge. Given the correlations among the APT subscales, a factor analytic model was specified to create a single APT factor from the original subscales, which represented individuals' general willingness to consume

drinks at increasing prices. Each of the original subscales loaded positively and significantly onto the composite APT factor (α = .71). The APT factor was used in all further analyses.

Determining covariates

Bivariate correlations were used to determine which variables should be included as covariates in the regression models assessing the relationships between the APT and the relevant measures. It was expected that gender and spending money would be significant covariates. As anticipated, moderate correlations were found between the APT and participant sex (male or female, with males being more likely to consume more alcohol at higher prices; r = -.30, p < .001) and between the APT and participant spending money (r = .30, p < .001). Consequently, in all analyses, participant sex and disposable income (i.e., spending money—participants averaged approximately $250 per month) were used as covariates. To assess how much variance in participants' APT scores was accounted for by the covariates, the covariates were entered into a regression model predicting a person's APT score. Sex accounted for 8.7% (p < .001) of the variance in APT scores and spending money accounted for 8.4% (p < .001).

Establishing the convergent validity of the APT

Based on previous research, alcohol expectancies and drinking motives were expected to be strongly associated with consumption of alcohol and with the experience of alcohol-related problems. Given past evidence of a relationship between APT scores and drinking, APT scores were also expected to be strongly associated with alcohol consumption, expectancies, and motives. Extending previous research, the current study also examined the relationship between APT scores and alcohol-related problems; moderate correlations between the constructs were expected.

To test these hypotheses, bivariate correlations between the APT and alcohol expectancies, drinking motives, alcohol consumption (i.e., a composite score representing typical quantity of alcohol consumed by typical frequency of consumption), and the experience of alcohol-related problems were examined (see Tables 1 & 2). A strong correlation (r = .58, p < .001) was found between the APT and alcohol consumption. Moderate correlations were found between the APT and social interaction motives (r = .36, p < .001), coping with anxiety motives (r = .32, p < .001), enhancement of experience motives (r = .40, p < .001), and experience of alcohol-related problems (r = .31 p < .001). Small positive correlations were found between the APT and social enhancement expectancies (r = .17, p = .011), sex-related expectancies (r = .26, p < .001), tension-reduction expectancies (r = .16, p =.014), and coping with depression motives (r = .26, p < .001).

Establishing the predictive validity of the APT

In general, the bivariate correlations provided preliminary evidence that the APT was related to alcohol use, the experience of alcohol-related problems, alcohol expectancies, and drinking motives. Therefore, a series of univariate and multivariate regression models were used to examine the predictive validity of the APT with respect to these constructs. It was expected that the APT would account for significant variance in both alcohol use and alcohol-related problems.

Multiple regression models suggested that the APT was a significant predictor of alcohol consumption ($\eta p^2 = .13$, $p < .001$) and the experience of alcohol-related problems, although the effect was small ($\eta p^2 = .02$, $p = .015$), and after controlling for typical drinking, the APT no longer predicted alcohol-related problems.

Multivariate regression models were used to evaluate the ability of the APT to predict alcohol expectancies and drinking motives because each dependent variable was comprised of several subscales (see Table 2). There was a significant main effect of the APT on alcohol expectancies as measured by the CEOA ($\eta p^2 = .09$, $p < .001$). With respect to the specific expectancies subscales, the APT accounted for significant variance in social ($\eta p^2 = .03$, $p = .017$), sexual ($\eta p^2 = .07$, $p < .001$), and tension reduction ($\eta p^2 = .03$, $p = .011$) expectancies. A univariate regression model was used to examine the relationship between the APT and the single positive expectancy factor. As expected, the APT predicted positive alcohol expectancies ($\eta p^2 = .07$, $p < .001$). In regards to the Drinking Motives Questionnaire, the APT accounted for significant variance in drinking motives as assessed by the MDMQR ($\eta p^2 = .16$, $p < .001$). With respect to the subscales, the APT accounted for significant variance in social ($\eta p^2 = .09$, $p < .001$) and enhancement ($\eta p^2 = .14$, $p < .001$) motives as well as drinking to cope with anxiety ($\eta p^2 = .09$, $p < .001$) and depression ($\eta p^2 = .06$, $p < .01$).

Establishing the incremental validity of the APT

Given preliminary evidence of convergent and predictive validity, the incremental validity of the APT was evaluated to determine if it accounted for significant variance in alcohol use and problems above and beyond existing measures known to relate to these outcomes (i.e., expectancies and motives). For the model assessing alcohol use, the APT was entered simultaneously with the covariates, the four expectancy subscales, and the five drinking motives subscales. The APT predicted drinking above and beyond expectancies and motives ($\eta p^2 = .13$, $p < .001$). While the magnitude of the effect was much smaller, the APT also predicted the experience of alcohol-related problems ($\eta p^2 = .02$, $p = .015$). However, when typical alcohol use was controlled for, the APT composite variable no longer predicted problems above and beyond expectancies and motives. This suggests that the effects of APT scores on problems operating through levels of consumption are primarily indirect.

Testing the APT as a mediator of alcohol consumption and problems

Through the process of evaluating the validity of the APT, we found that it was related to positive expectancies and to drinking motives, and that it accounted for unique variance in alcohol use and problems above and beyond these constructs. Given our interest in further understanding the relationships between the APT and alcohol use, and the APT and problems, we assessed whether these could be understood, at least in part, as a function of individuals' alcohol expectancies and/or drinking motives. Using hierarchical multiple regression, we first examined whether alcohol expectancies and/or drinking motives mediated the relationship between the APT and alcohol consumption. We then checked whether expectancies and/or motives mediated the relationship between the APT and the experience of alco-

hol-related problems. Our hypothesis was that expectancies and drinking motives would mediate both of these relationships. The following steps designed by Baron and Kenny (1986) were used to test for mediation:

Step 1: Show that the independent variable (the APT) predicts the outcome (consumption of alcohol or experience of alcohol related problems). To establish the potential for mediation, run a regression model with the APT and the outcome variables.

Step 2: Show that the APT predicts the mediator (expectancies or motives for drinking). Test the ability of the APT to predict expectancies or motives by running a regression model with the mediators as the outcome variables*.

Step 3: Show that the mediators predict the outcome variable(s) when controlling for the APT. If expectancies and motives completely mediate the relationships between the APT and alcohol consumption or between the APT and drinking problems, the effect of the APT on the outcome variables when controlling for expectancies/motives would be zero.

Step 4: If full mediation is not present, there is still the potential for partial mediation. Partial mediation can be tested for using the Sobel test to evaluate whether there is a significant indirect effect. In this case, there may be a significant indirect effect of APT scores on drinking through alcohol expectancies.

Alcohol Use

Expectancies as mediators of the relationship between the APT and weekly drinking

Step 2: A significant relationship between the APT and alcohol expectancies was established (see Table 1). The APT predicted sex, social, and tension reduction alcohol expectancies. Because these three subscales constitute the positive side of alcohol expectancies, they were grouped together and are hereafter referred to as "positive expectancies."

Step 3: Positive expectancies were significantly related to alcohol consumption when controlling for the APT ($\eta p^2 = .046$, $p = .01$; see Table 2 and Figure 1 for more detail). As the effect of the APT on the outcome variables was not zero, expectancies did not fully mediate the relationship between APT scores and drinking.

Step 4: To evaluate whether expectancies served as partial mediators, the Sobel test was used to determine whether the indirect effect was significant. Positive expectancies significantly partially mediated the relationship between the APT and alcohol consumption ($z = 3.068$, $p = .002$; see Figure 1).

Motives as mediators of the relationship between APT and drinking

Step 2: A significant relationship between the APT and drinking motives was established (see Table 1). The APT predicted so-

* Both steps 1 and 2 were evaluated within the predictive validity analysis (see Tables 1 and 2). The results of Step 2 are presented in the following section as a reminder of which subscales were significant. However, to prevent repetition, results from Step 1 are not presented again.

cial, enhancement, and coping with depression and anxiety motives for drinking. Although some of these subscales were correlated with each other, the correlations were not strong enough to make multicollinearity a concern.

Step 3: Social (ηp^2 = .013, p = .005) and enhancement (ηp^2 = .017, p = .004) motives, but not coping motives, were significantly related to alcohol consumption when controlling for the APT and other motives (see Table 2 and Figure 2 for more detail). As the

Table 1. Correlations of APT with alcohol use, problems, expectancies, and motives.

Alcohol Use and the Experience of Alcohol-Related Problems				
Alcohol Use .575**			Problems .316**	
Expectancies				
Self-Perception -.039	Sex .260**	Social .17*	Tension Reduction .163*	
Drinking Motives				
Social	Coping with Anxiety	Coping with Depression	Enhancement	Conformity
0.355**	0.321**	0.259**	0.398**	0.109
Note: *p < .05. **p < .001.				

Table 2. Predictive and incremental validity of the APT for alcohol use, problems, expectancies, and motives.

Variable	R^2	*B*	*SE B*	β
Alcohol Use	.132**	.079	.009	.538
Problems	.019*	.011	.003	.271
Positive Expectancies	.089*	.247	.060	.294
Drinking Motives				
Social	.090*	.659	.153	.295
Coping (Depression)	.057*	.722	.239	.212
Enhancement	.135*	.958	.170	.384
Note: *p < .05. **p < .001.				

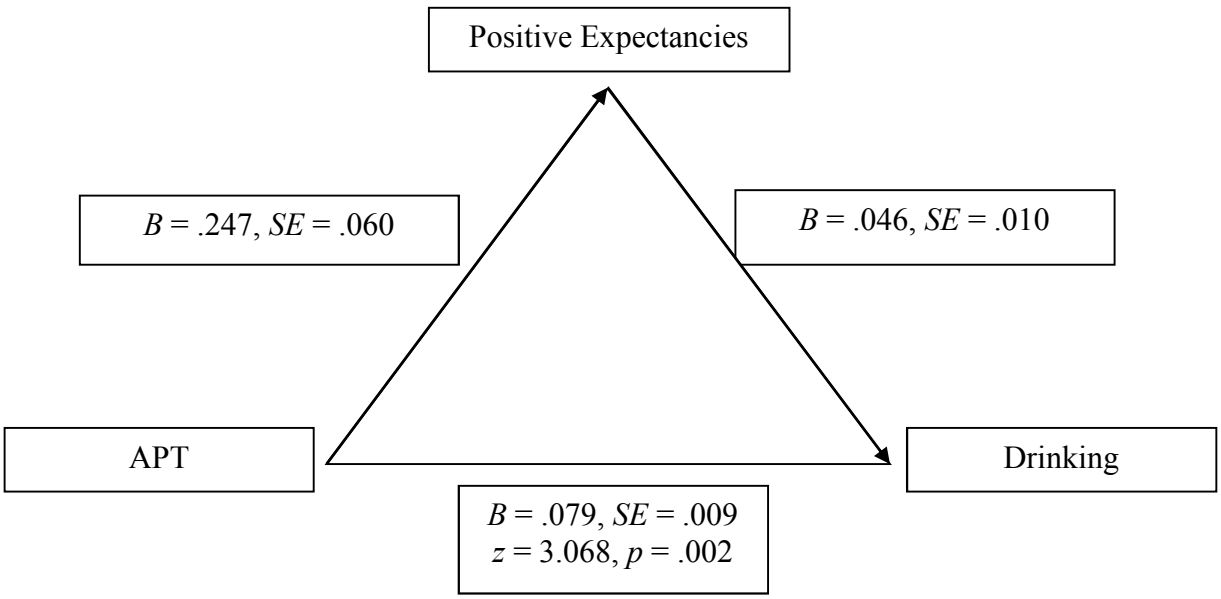


Figure 1. Positive expectancies as partial mediators of the relationship between price sensitivity (APT) and alcohol use.

effect of the APT on the outcome variables was not zero, motives did not fully mediate the relationship between APT scores and drinking.

Step 4: To evaluate whether motives served as partial mediators, the Sobel test was used to determine whether any of the indirect effects were significant. Social ($z = 3.14$, $p = .001$) and enhancement ($z = 3.42$, $p < .001$) motives significantly partially mediated the relationship between the APT and consumption of alcohol (see Figure 2).

The Experience of Alcohol-Related Problems

Expectancies as mediators of the relationship between APT and problems

Step 3: Positive expectancies were significantly related to alcohol-related problems when controlling for the APT ($\eta p^2 = .013$, $p = .003$; see Table 2 and Figure 3 for more detail). As the effect of the APT on the outcome variables was not zero, positive expectancies did not fully mediate the relationship between APT scores and problems.

Step 4: To evaluate partial mediation, the Sobel test was used to determine whether the indirect effect was significant. Positive expectancies ($z = 2.98$, $p = .002$) significantly partially mediated the relationship between the APT and experiencing alcohol-related problems (see Figure 3).

Motives as mediators of the relationship between APT and problems

Step 2: A significant relationship between the APT and drinking motives was established (see Table 1). The APT predicted social, enhancement, and coping with depression and anxiety motives for drinking.

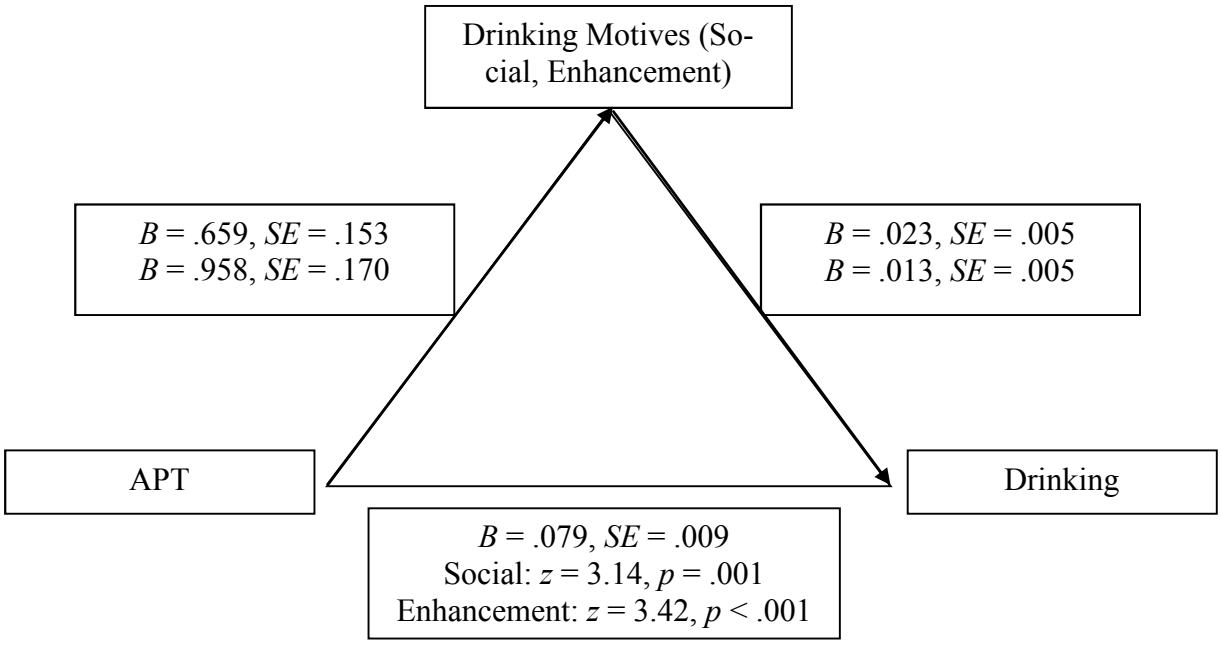


Figure 2. Drinking motives as partial mediators of the relationship between price sensitivity (APT) and alcohol use.

Step 3: Drinking to cope with depression motives was significantly related to alcohol-related problems when controlling for the APT and other motives ($\eta p^2 = .004$, $p = .001$; see Table 2 and Figure 4 for more detail). As the effect of the APT on the outcome variables was not zero, motives did not fully mediate the relationship between APT scores and problems.

Step 4: To evaluate partial mediation, the Sobel test was used to determine whether the indirect effect was significant. Coping with depression ($z = 2.41$, $p = .015$) significantly partially mediated the relationship between the APT and experiencing alcohol-related problems (see Figure 4).

DISCUSSION

The APT is a relatively new measure, and there is a paucity of research examining its psychometric properties. The current study was the first to extend analyses of the convergent, predictive, and incremental validity of the APT. As anticipated, the APT was significantly related to alcohol expectancies and drinking motives, two constructs that have been consistently identified as important predictors of drinking behavior. The strength of the correlations indicated that these constructs are related to one another but that they are also distinct. Improving upon previous research, the current study was the first to evaluate the relationship between the APT and spending money. Results indicated a moderate positive relationship between these constructs, such that as disposable income increased, so did participants' general willingness to purchase drinks at higher prices. Although it was not possible to assess why the relationship between spending money and the APT existed, a plausible explanation may be that individuals with more money might be willing to incur higher drink prices simply because they can afford to. Therefore, the results of previous studies could have been driven by a

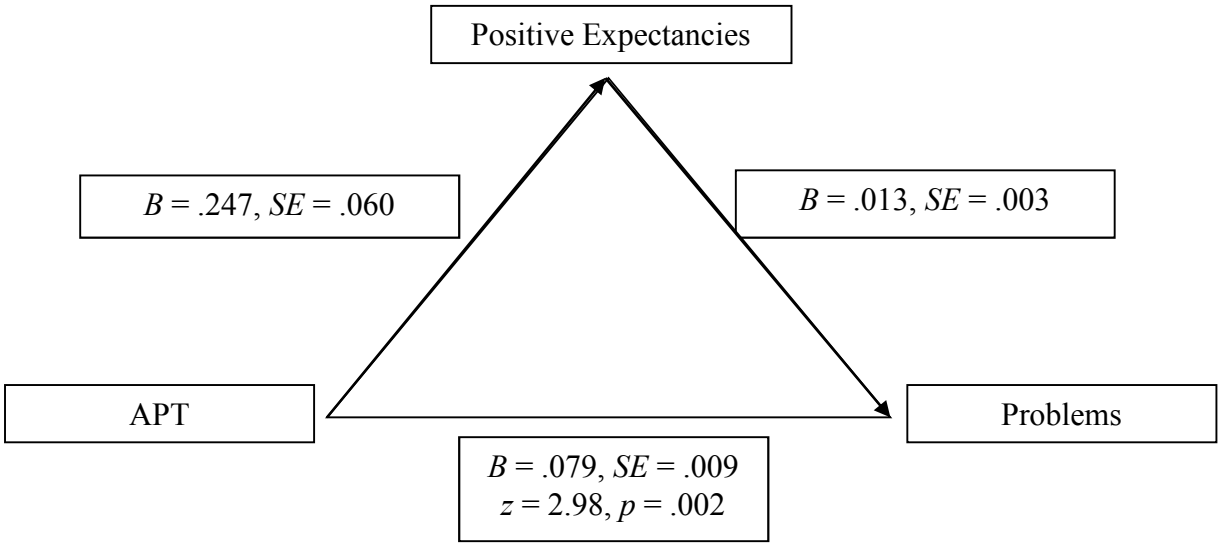


Figure 3. Positive expectancies as partial mediators of the relationship between price sensitivity (APT) and the experience of alcohol-related problems.

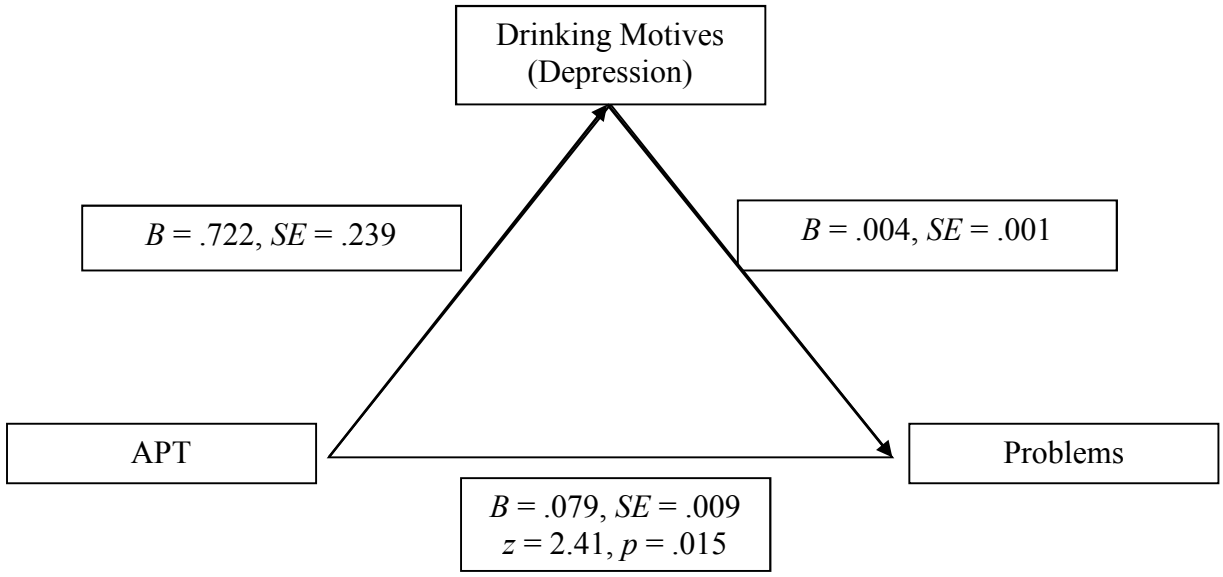


Figure 4. Drinking motives as partial mediators of the relationship between price sensitivity (APT) and the experience of alcohol-related problems.

combination of price sensitivity and individuals' disposable income. The current study was the first to take disposable income into account to help isolate price insensitivity.

Even after controlling for gender and spending money, the APT accounted for significant variance in alcohol consumption and drinking-related problems. Furthermore, after controlling for alcohol expectancies and motives for drinking, the APT demonstrated incremental validity in accounting for consumption and problems. In sum, the results provide support for the value of the APT as a measure of an individual's willingness to consume alcohol at different price points.

Given that the APT, alcohol expectancies, and drinking motives each predicted alcohol consumption and the experience of alcohol-related problems, the current study examined whether and how these constructs may work in concert in predicting alcohol outcomes. Prior to the current study, this issue had not yet been addressed; previous studies had shown individuals who are more insensitive to the price of alcohol are more likely to be heavier drinkers, but steps have not yet been taken to explain this pattern. Given that there is a strong relationship between alcohol expectancies/motives for drinking and alcohol consumption, the current study formulated a model examining whether insensitivity to price contributes to drinking through these constructs. In other words, we hypothesized that the APT might affect consumption of alcohol and problems through expectancies and motives. The preliminary tests of mediation suggested this may be the case, but because the data are cross-sectional, future work is necessary before any kind of definitive answer can be reached.

The current study provided preliminary evidence for the convergent, predictive, and incremental validity of the APT. Furthermore, it made an important initial attempt at deconstructing the relationship between price insensitivity and alcohol use documented by earlier studies (Wagenaar et al., 2009). Despite these key strengths, the study has several limitations that should be considered. Previous studies either examined each of the APT subscales independently or created two factors representing price and consumption. Within the current study, the APT subscales evidenced significant overlap with one another, making it difficult to distinguish between them. In addition, the problems with multicollinearity among the subscales made it impossible to replicate the two factors used in previous papers. As such, the current study collapsed the APT subscales into a single factor representing a composite of willingness to purchase alcohol at increasing prices. Unfortunately, using the composite APT score made it impossible to assess the independent contributions of the APT subscales. Additionally, the study's inability to examine each subscale individually within the model may have contributed to the fact that many of the effect sizes had smaller magnitudes than expected, a second important limitation. It could be that the effect of a subscale that might have accounted for a moderate amount of variance when taken by itself was masked in the process of being combined with other, less important subscales. Alternatively, these small effects could reflect the true magnitude of the associations between the APT and other alcohol-related variables.

A third limitation of the current study was that all of the data came from self-report measures. Although the accuracy of

self-report data has been questioned, self-report measures have generally been shown to be reliable when it comes to assessing alcohol consumption and its consequences (Babor, Steinberg, Anton, & Del Boca, 2000).

Finally, it is important to acknowledge that all of the data within the current study were cross-sectional, representing only a snapshot of participants' experience. As such, it was impossible to assess relationships among variables across time, prohibiting conclusions about the directionality of effects (i.e., causality). While the current study suggested that insensitivity to price may translate to heavier drinking through positive expectancies and motives, it is also possible that a heavy drinker may be so motivated to consume alcohol as a result of developing positive expectancies over time that he or she is willing to incur higher costs for the reward of drinking. A third possibility is that price insensitivity leads to heavy drinking, which leads to more positive expectancies and stronger motives for drinking. Longitudinal data is needed to flesh out the temporal relationships among these constructs.

Despite the limitations, the current study suggested that price sensitivity is a construct worthy of further investigation. The preliminary mediation analyses indicated that price insensitivity may operate through expectancies and motives to influence consumption of alcohol and the experience of problems. In addition, this study demonstrated for the first time that spending money and gender must be taken into account when examining price sensitivity to alcohol consumption. Future research will benefit from including these important covariates.

On a broader level, the results of the current study may have implications for alcohol tax policies and for alcohol interventions. The study suggests that the relationship between price insensitivity and heavy drinking identified in previous research may be due, at least in part, to the amount of spending money individuals have and to the kinds of alcohol expectancies and drinking motives they hold. Replicating the results of this study would lend strength to the idea that large-scale interventions designed to temper alcohol consumption, like sin taxes, may only be effective for certain subgroups of the population. This is not to say that sin taxes are completely ineffective, or that they should be eradicated. Rather, when choosing an intervention designed to target those for whom drinking causes the most problems (i.e., heavy drinkers), attempting to control price may only be an effective intervention in concert with targeting expectations and motivations for alcohol use.

REFERENCES

Altman, A. (2009). A brief history of: Sin taxes. *Time Magazine*. Retrieved from http://www.time.com/time/magazine/article/0,9171,1889187,00.html.

Ashley, M. J., Ferrence, R., Room, R., Rankin, J., & Single, E. (1994). Moderate drinking and health: Report of an international symposium. *Canadian Medical Association Journal*, *151*, 809-824.

Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. (2000). Talk is cheap: Measuring drinking outcomes in clinical trials. *Journal of Studies on Alcohol, 61*, 55-63.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Blackwell, E., & Conrod, P. J. (2003). A five-dimensional measure of drinking motives. Unpublished manuscript, Department of Psychology, University of British Columbia.

Brown, S. A., Goldman, M. S., & Christiansen, B. A. (1985). Do alcohol expectancies mediate drinking patterns of adults? *Journal of Consulting and Clinical Psychology, 53*, 512-519.

Carey, K. B. (1995). Alcohol-related expectancies predict quantity and frequency of heavy drinking among college students. *Psychology of Addictive Behaviors, 9*, 236-241.

Centers for Disease Control. (2008). Alcohol and public health. Retrieved from http://www.cdc.gov/alcohol/index.htm. Last updated Aug. 6, 2008.

Centers for Disease Control. (2009). Community guide: Task force findings. Last updated Apr. 20, 2009.

Cox, M. W., & Klinger, E. (1988). A motivational model of alcohol use. *Journal of Abnormal Psychology, 92*, 168-180.

Cutter, H. S., & O'Farrell, T. J. (1984). Relationship between reasons for drinking and customary drinking behavior. *Journal of Studies on Alcohol, 45*, 321-325.

Grant, V. V., Stewart, S. H., & Mohr, C. D. (2009). Coping-anxiety and coping-depression motives predict different daily mood-drinking relationships. *Psychology of Addictive Behaviors, 23*, 226-237.

Grant, V. V., Stewart, S. H., O'Connor, R. M., Blackwell, E., & Conrod, P.J. (2007). Psychometric evaluation of the five-factor modified drinking motives questionnaire—Revised in undergraduates. *Addictive Behaviors, 32*, 2611-2632.

Grossman, M., Sindelar, J. L., Mullahy, J., & Anderson, R. (1993). Policy watch: Alcohol and cigarette taxes. *Journal of Economic Perspectives, 7*, 211-222.

Grube, J. W., & Agostinelli, G. E. (1999). Perceived consequences and adolescent drinking: Nonlinear and interactive models of alcohol expectancies. *Psychology of Addictive Behaviors, 13*, 303-312.

Guide to Community Preventive Services. (2006). Preventing excessive alcohol use: Increasing alcohol taxes. Retrieved from http://www.thecommunityguide.org/alcohol/increasing taxes.html. Last updated Feb. 10, 2009.

Ham, L. S., Stewart, S. H., Norton, P. J., & Hope, D. A. (2005). Psychometric assessment of the comprehensive effects of alcohol questionnaire: Comparing a brief version to the original full scale. *Journal of Psychopathology and Behavioral Assessment, 27*, 141-158.

Hansen, W. B., Raynor, A. E., & Wolkenstein, B. H. (1991). Perceived personal immunity to the consequences of drinking alcohol: The Relationship between behavior and perception. *Journal of Behavioral Medicine, 14*, 829-857.

Hursh, S. R. (2000). Behavioral economic concepts and methods for studying health behavior. In W. K. Bickel & Vuchinich, R. E. (Eds.), *Reframing Health Behavior Change with Behavioral Economics* (pp. 27-60). Mahwah, NJ: Erlbaum.

Hursh, S. R., Raslear, T. G., Shurtleff, D., Bauman, R., & Simmons, L. (1998). A cost-benefit analysis of demand for food. *Journal of the Experimental Analysis of Behavior, 50*, 419-440.

Leigh, B. C. (1999). Peril, chance, adventure: Concepts of risk, alcohol use and risky behavior in young adults. *Addiction, 94*, 371-383.

MacKillop, J., & Murphy, J. G. (2007). A behavioral economic measure of demand for alcohol predicts brief intervention outcomes. *Drug and Alcohol Dependence, 89*, 227-233.

MacKillop, J., Murphy, J. G., Ray, L. A., Eisenberg, D. T., Lisman, S. A., Lum, J. K., & Wilson, D. S. (2008). Further validation of a cigarette purchase task for assessing the relative reinforcing efficacy of nicotine in college smokers. *Experimental and Clinical Psychopharmacology, 16*, 57-65.

MacKillop, J., Murphy, J. G., Tidey, J. W., Kahler, C. W., Ray, L. R., & Bickel, W. K. (2009). Latent structure of facets of alcohol reinforcement from a behavioral economic demand curve. *Psychopharmacology, 203*, 33-40.

Murphy, J. G., & MacKillop, J. (2006). Relative reinforcing efficacy of alcohol among college student drinkers. *Experimental and Clinical Psychopharmacology, 14*, 219-227.

Newman, J. S. (2003). Slave tax as sin: 18th and 19th century perspectives. *Taxanalysts*. Retrieved from http://www.taxhistory.org/thp/readings.nsf/ArtWeb/4AF487C90CA 14FB985 256E000057B5EB?OpenDocument.

Newman, R. (2010). 8 sneaky ways to raise taxes. *US News & World Report*. Retrieved from http://www-origin.usnews.com/money/blogs/flowchart/2010/02/02/8-sneaky-ways-to-raise-taxes.

Wagenaar, A. C., Salois, M. J., & Komro, K. A. (2009). Effects of beverage alcohol price and tax levels on drinking: A meta-analysis of 1003 estimates from 112 studies. *Addiction, 104*, 179-190.

Wechsler, H., Dowdall, G. W., Maenner, G., Gledhill-Hoyt, J., & Lee, H. (1998). Changes in binge drinking and related problems among American college students between 1993 and 1997: Results of the Harvard School of Public Health and College Alcohol Study. *Journal of College Health, 47*, 57-68.

World Health Organization. (2007). Alcohol and injury in emergency departments: Summary of the Report from the WHO Collaborative Study on Alcohol and Injuries. Retrieved from http://www.who.int/substance_abuse/publications/alcohol_injury_summary.pdf.

Associations between Color and Music are Mediated by Emotion and Influenced by Tempo

Tawny Tsang & Karen B. Schloss

UNIVERSITY OF CALIFORNIA, BERKELEY

Historically, there have been a number of attempts to characterize the relationship between perceptions of color and perceptions of music (e.g., McClain, 1978; Cuddy, 1985; Shaw-Miller, 1959; Arnheim, 1986). One line of recent research suggests that associations between color and music are based on the shared emotions that they evoke (Schloss, Lawler, & Palmer, 2008). The current study explored this hypothesis further by examining particular features of music contributing to color-music associations. Corroborating previous research, we found consistent correlations between emotional ratings for musical selections and emotional ratings for colors. Specifically, we show that the same piece can be associated with different colors on the basis of manipulating tempo alone.

INTRODUCTION

Synesthetes are famed for their ability to "hear" colors or "see" sound. However, the widespread use of such colorful phrases as "white noise" and "the blues" suggests that ordinary people, too, perceive relationships between sound and color. Most individuals understand what it means for a color or a sound to be "bright" or "dark" or to have a "warm" tone versus a "cool" one. Novice and professional artists alike can agree that a trumpet is a "bright" instrument and that red is a "warm" color. In addition to these linguistic analogies, the historical concurrence of certain artistic and musical movements with overlapping elements, such as impressionism and minimalism, has led some to posit perceptual relations between the experiences of color and music (e.g., Simon, 1959; Arnheim, 1986).

Historically, the cognitive underpinnings of the relationship between color and music have been a heady source of investigation and debate for philosophers and psychologists. Plato, Aristotle, and Newton theorized associations between specific musical intervals and specific colors (for a historical overview, see, e.g., McClain, 1978; Cuddy, 1985). The findings of these thinkers dovetail with their metaphysical starting points. Plato, for example, drew on Pythagorean theory to link the interval of a perfect fourth to yellow. Newton was influenced by his optical experiments to link this same interval to orange due to the similar wavelength properties of light and sound (Westfall, 1962). More recent attempts to model the relationship between music and color have similarly been based on correspondences between their physical properties. For example, the psychophysical theory of music and color (Pridmore, 1992) explains this relationship as a function of amplitude and wavelength. Another proposal maps hue onto pitch and saturation onto timbre (Caivano, 1994).

Other models focus on the emotional experiences that engender perceived similarities between color and music. Prior research has demonstrated that color and music independently carry emotional valence (e.g., Ballast, 2002; Kaya, 2004; Schubert, 2007; Hunter, 2010). For example, fast, major musical excerpts are reliably perceived as happy while minor excerpts are reliably perceived as sad (Hunter, 2010). Similarly, light colors are associated with more positive emotions while dark colors are associated with more negative emotions (Kaya, 2004). These effects extend to more complex emotions as well. One study found that Western-European university students consistently associated red with "excitement" and yellow with "playful" (D'Andrade & Egan, 1974). Moreover, associations between particular musical pieces and paintings may be drawn if both carry the same emotional effect. One study showed that participants use different color profiles to classify the same piece of

The authors would like to acknowledge Dan Levitin for his help developing experimental stimuli. The first author can be reached at ttsang@emory.edu.

music depending on instrumentation and the musician's emotional interpretation of the piece (e.g., happy, sad, shameful, or angry; Bresin, 2005). However, this study did not measure how participants judged the emotional content of the recordings or colors and did not explain how musicians' emotional intent was translated into performance (e.g., by change in articulation or dynamic range). Idiosyncrasies in both performance and personal preference may have influenced these color-music associations. A more recent study helped account for these ambiguities (Schloss et al., 2008). In addition to associating colors with musical excerpts, participants provided emotional ratings for the two types of stimuli. Tempo and mode were identified as features of music that influence these associations. For example, saturated and light colors and fast and major music were independently rated as happy and were also associated with each other.

Despite narrowing the factors that influence the emotional experience of music and its perceived associations with color, several other features besides tempo and mode may have accounted for the findings of Schloss et al. (2008). For example, stimuli were symphonic pieces that also varied in orchestral size and instrumentation. Considering that many of the terms used to describe music and color apply to timbre (e.g., brightness, warmth, saturation), it is conceivable that the differences that these authors found in color-music associations could be due to such factors.

The current study aimed to control for timbre, genre, and instrumentation in an attempt to clarify how much of the variation in documented music-color associations can be attributed to tempo and mode alone. Whereas previous studies used a variety of instruments to evaluate music-color associations, the current study only used piano pieces for musical stimuli. In addition, we electronically manipulated the speed of the excerpts such that stimuli representing fast and slow tempos were derived from a single piece. We also selected major and minor pieces from the same genre or compositional collection in order to narrow the amount of artistic and stylistic difference represented in our stimuli, in turn minimizing participants' unique responses to idiosyncratic features.

STUDY

Method

Participants

Ten undergraduates (3 male; all between 20-25 years old) enrolled in a course on color perception participated in this study on a voluntary basis. No participants had any known color deficiencies.

Musical stimuli

Thirty-second excerpts from six classical piano pieces were used to create 12 stimuli. The six original pieces were composed by J.S. Bach, Robert Schumann, and Frederic Chopin, each of whom contributed one major and one minor composition to our stimuli (see Appendix). Each major and minor piece reflected a single genre, which reduced the amount of stylistic variation between modes for composers. The composers were selected based on their prolific contributions to the standard piano repertoire and their representativeness of the Baroque, Classical, and Romantic periods. Where possible, pieces within the same genre (e.g., prelude to fugue, ballade, and sonata) were

recordings from the same pianist on the same piano. These six selections were also relatively slow in tempo (mean tempo: 77 +/- 16 beats per minute).

The original, full-length pieces were downloaded from http://www.pianosociety.com, a free classical music website with a large selection of piano works. The tempos of the pieces were sped up using Pitch 'n Time (Serato Audio Research, Auckland, New Zealand; http://serato.com/pitchntime), computer software from the Levitin Laboratory for Music, Perception, Cognition, and Expertise at McGill University. This manipulation did not leave any sound artifacts, such as distorted pitch. Mean tempo for the faster versions was 138 +/- 21 beats per minute.

Thirty seconds of music were excerpted from the original six pieces using a program called Audacity. All excerpts ended on a cadence, suggesting a natural stopping point in the music. For the sped-up versions, the excerpted material was looped until it reached 30 seconds in length.

Color Stimuli

The stimuli consisted of 37 colors from the Munsell color system (see Appendix). These included eight hues (red, yellow, green, blue, orange, chartreuse, cyan, and purple) with four saturation/lightness conditions: maximum saturation at medium lightness, medium saturation at medium lightness, high lightness, and low lightness. Five stimuli were achromatic, ranging from black to white.

Apparatus

Visual displays were presented using PowerPoint Mac 2004 on a MacBook Pro computer that was connected to a color projector. Musical stimuli were WAV audio files that were embedded into the PowerPoint slides, allowing for the simultaneous presentation of auditory and visual material. Speakers were also used to amplify the music so that it could be heard clearly.

Procedure

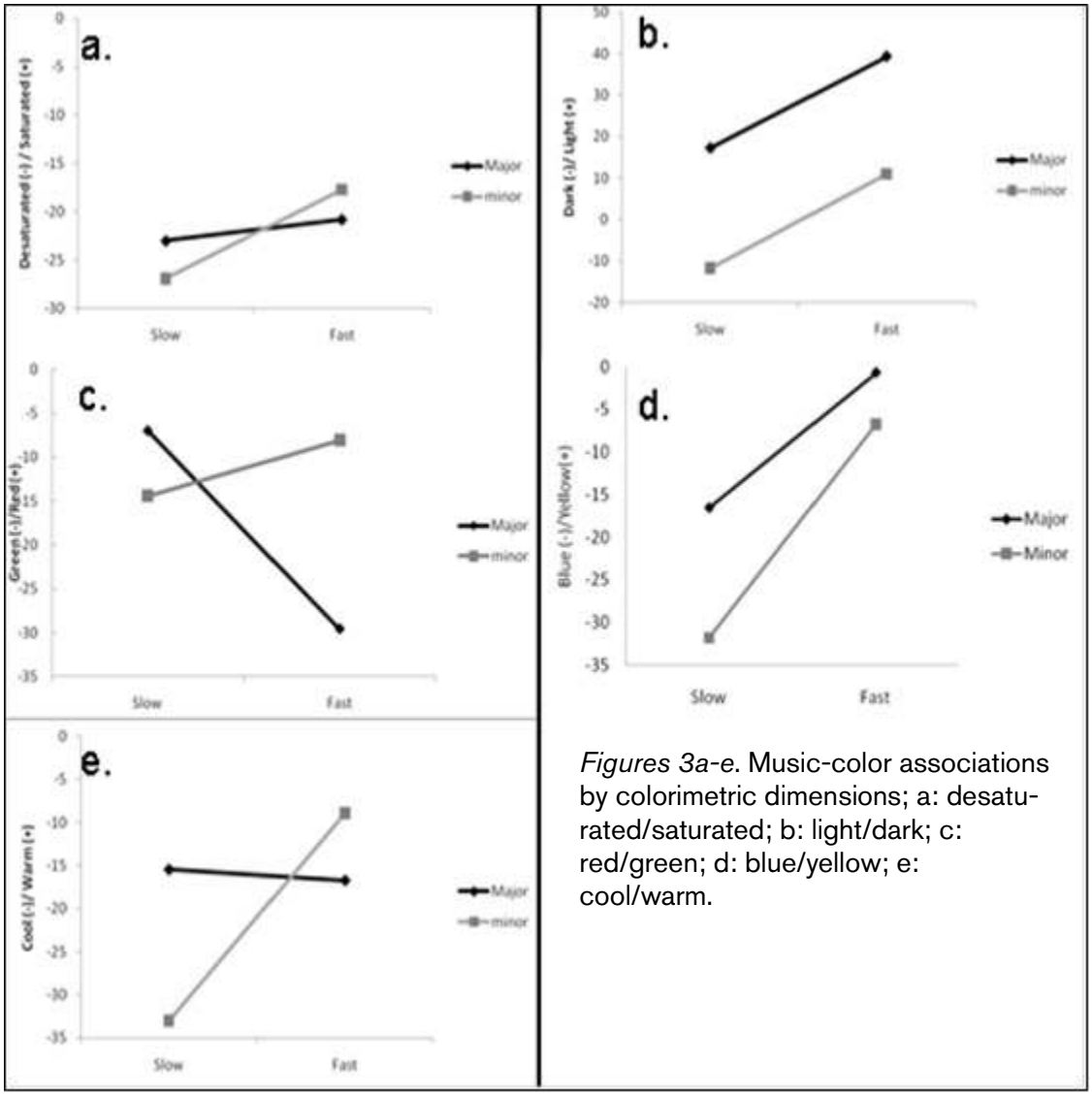
The experiment was separated into three blocks. In the first block, participants were asked to rate the emotional content of 37 color stimuli. The stimuli were displayed in a random order on a projector screen for 40 seconds each. During this time participants rated each color on five dimensions: sad/happy, calm/angry, weak/strong, passive/active, and dislike/like. Ratings were made on a scale from -5 to 5, with 0 indicating a neutral response.

In a second block five weeks later, participants were asked to associate these colors with the twelve aforementioned musical clips. Participants first previewed fifteen-second versions of each selection and were told that although some selections might sound the same, none of them were identical. Participants were then played the full thirty-second clips. After hearing each clip, participants were presented with the 37 colors encountered in the first block of the study and given 40 seconds to rank-order the five colors which they felt were most consistent and most inconsistent with the clip.* Consistency was explained as colors that participants most strongly associated with the music. Similarly, inconsistency was explained as colors that participants would

* Sound files were presented in pseudorandom order so that a slow or fast version of the same piece was not heard consecutively. Trials were only repeated for cases in which participants could not complete color associations in the allotted forty seconds.

not associate with the music. Participants were instructed to make their music-color associations based on the experience of the current music without relating it to other pieces heard either in the experiment or in real life.

The third block occurred a week after the second one. During this block, participants made music-emotion associations. The procedure was very similar to the color-emotion association task in the first block. For each of the 12 musical excerpts, participants rated their emotional response on the five dimensions previously mentioned (sad/happy, calm/angry, weak/strong, passive/active, and dislike/like). As in the color-emotion association task, ratings were made on a scale from -5 to 5. The clips were presented



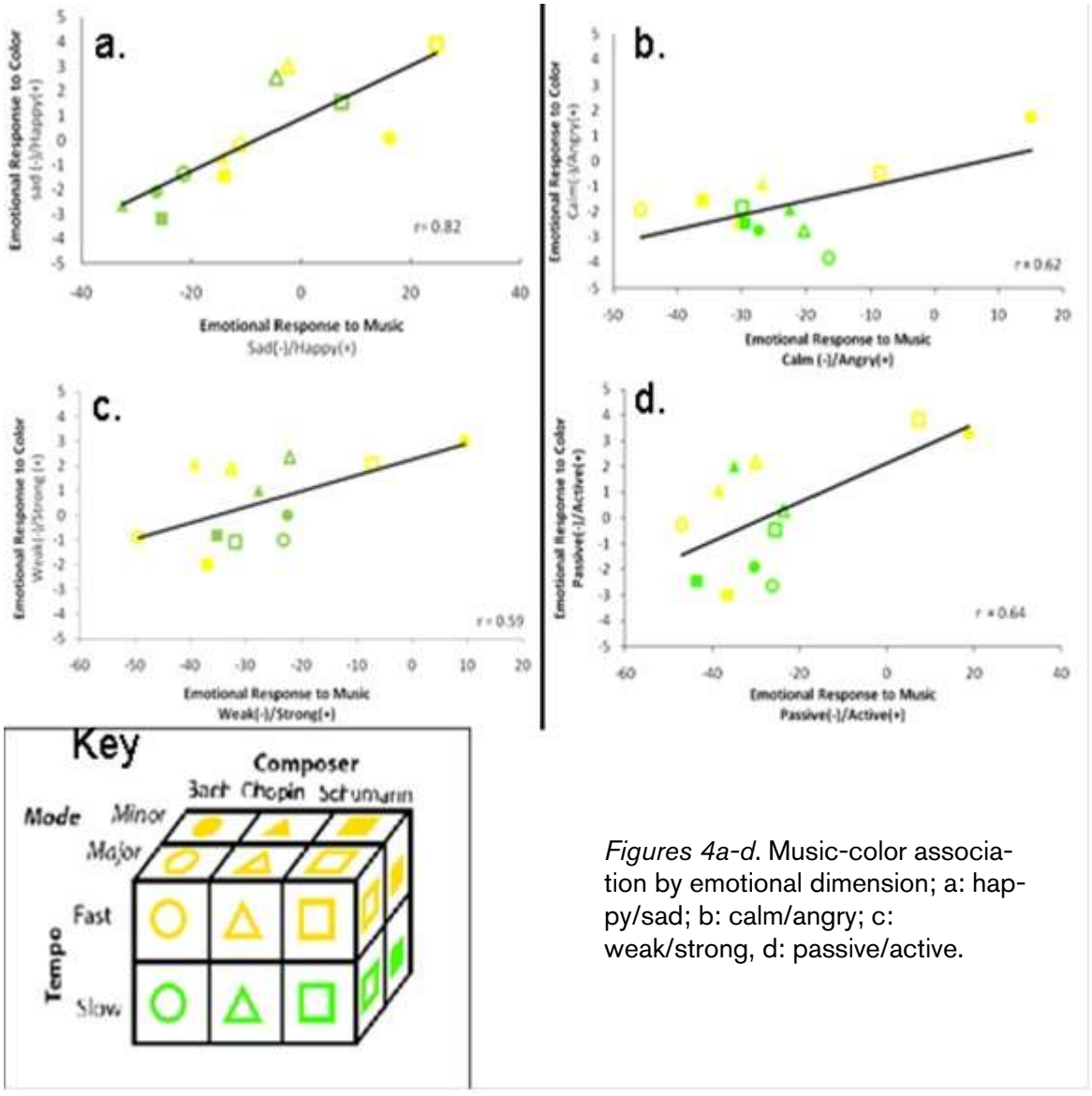
Figures 3a-e. Music-color associations by colorimetric dimensions; a: desaturated/saturated; b: light/dark; c: red/green; d: blue/yellow; e: cool/warm.

in the same order as in the previous week.

Results

Color-Music Association

Just as music can be analyzed in terms of timbre, tonality, and mode, every color can be broken down in terms of its colorimetric composition: saturation, lightness versus darkness, red versus green, blue versus yellow, and cool versus warm. These dimensions provide a standard basis for analyzing which component of the colors lends itself the most to color-music associations. In order to condense the selections into a single music-color association for a given piece,



Figures 4a-d. Music-color association by emotional dimension; a: happy/sad; b: calm/angry; c: weak/strong, d: passive/active.

colors that were ranked most consistent or most inconsistent were weighted more heavily according to the rank given to them by the participant. To quantify color-music associations, we conducted the following calculation for each music clip under the five colorimetric dimensions:

$C_D = 5c_{1,D} + 4c_{2,D} + 3c_{3,D} + 2c_{4,D}. + 1\ c_{5,D}$, where C = consistency score, D = dimension, and $c_{i,D}$ = the i^{th} color chosen

$I_D = 5i_{1,D} + 4i_{2,D} + 3i_{3,D} + 2i_{4,D}. + 1i_{5,D}$, where I = inconsistency score, D = dimension, and $i_{i,D}$ = the i^{th} color chosen

Saturation

For ratings of saturation, there was no main effect for composer, mode, or tempo: ($F(2, 18) < 1$; mode: $F(1, 9) < 1$; and tempo: $F(1, 9) < 1$). Participants generally associated all music selections with desaturated colors (Figure 3a). This contrasts with Schloss et al.'s (2008) findings that both faster music and music in the major mode are associated with more saturated colors.

Lightness/darkness

We did find a significant main effect on the light/dark colorimetric dimension (Figure 3b). Music that was in the major mode and at faster tempos was associated with lighter colors (Mode: $F(1, 9) = 9.51$, $p < 0.05$, Tempo: $F(1, 9) = 8.50$, $p < 0.05$).

Red/green and blue/yellow

All excerpts were associated with colors that contained more hues of green than red (Figure 3c). No main effects were observed for composer, mode, or tempo (Composer: $F(2,18) < 1$, $p > 0.05$; Mode : $F(1,9) = 1.48$, $p > 0.05$; Tempo: $F(1,9) = 4.27$, $p > 0.05$). However, there was an interaction between mode and tempo ($F(1, 9) = 10.31$, $p < 0.05$). Participants perceived faster major selections to be greener than their slower counterparts. In the blue/yellow colorimetric dimension, slower musical selections were associated with bluer colors than those associated with fast musical selections ($F(1, 9) = 8.13$, $p < 0.05$; Figure 2d). Minor musical selections were associated with colors that were more blue than major selections, although this effect was not quite statistically significant ($F(1,9) = 1.28$, $p > .05$).

Warmth/coolness

In general, participants associated the musical selections with colors that were cool (Figure 3e). A main effect was observed for tempo ($F(1, 9) = 8.06$, $p < 0.05$): the faster music selections were on average associated with less cool colors. This effect appeared especially strong for the minor pieces, but could not be shown to be statistically significant ($F(1, 9) = 4.55$, $p = 0.62$).

In sum, participants tended to associate the musical selections with desaturated, green, blue, and cool colors. Faster musical selections and those in a major key were associated with lighter colors. Music-color associations were significantly influenced by tempo alone. This accounts for how the same piece of music could be connected with two different colors.

Emotional Ratings

We examined whether these associations were mediated by participants' emotional reactions. The color-emotion ratings taken during the first block were averaged across participants to yield a single score for every color along the five emotion dimensions (sad/happy, calm/angry, weak/strong, passive/active, and dislike/like). These scores

were used to replace the 10 color selections made for each music stimuli. A music-emotion score was created for each musical stimulus by averaging the difference in color-emotion score between the five most and least consistent colors.

Happy/sad emotional dimension

We found consistent correlations between the emotional ratings of each musical selection and the emotional ratings of the colors that were chosen as consistent or inconsistent with the music. The correlation for emotional responses between color and music was highest for happy/sad ratings (r = 0.82, p < 0.001; Figure 4a). For any given piece, the faster version was happier than the slower version. Major selections were also rated happier than minor ones. For musical selections in which mode and tempo were incongruent (i.e., slow/major and fast/minor), happy/sad ratings were more neutral. This finding is consistent with the music-emotion ratings found in Hunter (2010). Moreover, musical selections that received a higher happiness rating were associated with colors that also received high happiness ratings.

Calm/angry, weak/strong, and passive/active emotional dimensions

Music and color were also correlated along other emotional dimensions, although to a lesser degree. The colors participants selected for musical stimuli had similar calm/angry (r = 0.62, p < 0.05), weak/strong (r = 0.59, p < 0.05), and passive/active (r = 0.64, p < 0.05) ratings as the corresponding musical stimuli (Figures 4b-d).

There was also significant overlap in how participants rated their emotional responses to musical selections in the calm/angry, passive/active, and weak/strong dimensions (correlation between weak/strong and calm/angry ratings: r = 0.98, p < 0.001; correlation for passive/active and weak/strong ratings: r = 0.95, p < 0.001; correlation between weak/strong and passive/active ratings: r = 0.97, p < 0.001). In other words, selections that were rated as calmer were generally rated as weaker and more passive, suggesting that these emotion dimensions may be variations of each other. Moreover, these selections were strongly correlated with colors that received similar emotional ratings. To give an example of how tempo modified music-color associations, the fast minor Bach selection was rated the most intense (i.e., most angry, active, and strong), and the fast major Bach selection was rated the least intense (i.e., most calm, passive, and weak). By contrast, the original, slow versions of these selections received similar emotional ratings along the three dimensions. This difference could be based on the accentuation of particular musical features under fast tempo.

DISCUSSION

The findings from this study underscore previous research demonstrating that color-music associations are mediated by the similar emotional responses that they evoke. In addition, by controlling for genre and composition in our musical stimuli, our findings reveal an important—if still open-ended—role for tempo and mode.

Importantly, considering the flexibility and variability of tempo from one performance of a musical piece to another, our findings suggest that music-color associations can vary significantly for different manifestations of musical stimuli even when

other features such as pitch, harmonics, and instrumentation are held constant. Considering that participants made independent color-emotion, music-emotion, and music-color associations at different points in time, the influence of tempo appears to be robust.

It is possible that tempo influences emotional response, in general, by accentuating particular features of musical pieces such as note density, chromaticism, and rhythm. This effect can take a variety of forms. For example, anger was more intensely perceived in the fast minor Bach excerpt than in the original slow version. Moreover, while slow minor stimuli were associated with greener colors than slow major stimuli, speeding up the music produced opposite associations; minor stimuli were associated with colors that were less green and major stimuli with colors that were greener. This was especially true for music-emotion ratings in the weak/strong and calm/angry dimensions.

It is worth noting that because excerpts from the fast conditions were looped in order to match the length of the slow selections, participants heard the same musical material more times in the fast conditions than in the slow ones. This repeated exposure could have influenced participants' responses. According to the basic assumptions of the mere exposure effect (e.g., Zajonc, 1968), such repetition could have produced more positive emotional ratings (i.e., happier or calmer). However, our data do not support this possibility; the musical selections from the fast condition did not necessarily receive more positive emotional ratings than those from the slow condition. For example, the fast major Schumann selection was rated as angrier than the slow version of it (Figure 4b).

A number of gaps in the present study present potentially fruitful avenues for further research. One shortcoming of the present study is that dynamic range was not held constant. Some music selections were much louder than others, which may have influenced both the emotional rating and corresponding color association. Moreover, musical selections were not uniformly sped up. For example, the fast version of the major Schumann piece was twice as fast as the slow version, whereas both of Chopin's *Ballades* were sped up to be 50% faster than the original versions. At the same time, our stimuli were limited to fast and slow tempos, with no "medium" condition. Further idiosyncrasies in our stimuli and results suggest an even broader range of future studies. For example, most of the music used here was associated with desaturated and cool colors. And despite intuitive connections between slow speed and lethargy, nearly all the fast stimuli were associated with passivity. These results may in part be due to the unique timbre of the piano. Overall, given the myriad dimensions upon which musical stimuli vary, as well as the documented influence of cultural and environmental conditions on color preferences and their associations with emotion (e.g., Terwogt & Hoeksma, 1995), we hope that our study is only one of many to come.

REFERENCES

Arnheim, R. (1986). *New essays on the psychology of art*. Berkeley, CA: University of California Press.

Bresin, R. (2005). What is the color of that music performance? *Proceedings of the International Computer Music Conference*, 367-370.

Caivano, J. L. (1994). Color and sound: Physical and psychophysical relations. *Color Research and Application*, *19*, 126-133.

Cuddy, L.L. (1985). The color of melody. *Music Perception: An Interdisciplinary Journal*, *2*, 345-360.

D'Andrade, R., & Egan, M. (1974). The colors of emotion. *American Ethnologist*, *1*, 49-63.

Hubbard, T.L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *American Journal of Psychology*, *109*, 219-238.

Hunter, P.G., Schellenberg, E., & Schimmack, U. (n.d.). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity and the Arts*, *4*, 47-56.

Kaya, N., & Epps, H. (2004). Relationship between color and emotion: A study of college students. *College Student Journal*, *38*, 396-405.

Marks, L. E. (1975). On colored hearing synaesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin*, *82*, 303-331.

McClain, E. G. (1978). *The Pythagorean Plato*. Stony Brook, N. Y.: Nicolas Hays Ltd.

Pridmore, R.W. (1992). Music and color: Relations in the psychophysical perspective. *Color Research & Application*, *17*, 57-61.

Schloss, K. B., Lawler, P., & Palmer, S. E. (2008). The color of music. Presented at the 8th Annual Meeting of the Vision Science Society, Naples, FL, May 2008.

Schubert, E. (2007). The influence of emotion, locus of emotion, and familiarity upon preference in music. *Psychology of Music*, *35*, 499-515.

Simon, P. (1959). *Visible deeds of music: Art and music from Wagner to Cage*. New Haven: Yale University Press.

Terwogt, M. M., & Hoeksma, J. B. (1995). Colors and emotions: Preferences and combinations, *Journal of General Psychology*, *122*, 5-17.

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, *123*, 394-409.

Westfall, R.S. (1962). The development of Newton's theory of color. *Isis*, *53*, 339-358.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*, 1-27.

APPENDIX

Musical selections

	Bach	Chopin	Schumann
Major	*Prelude No.7 in E-flat Major*	*Ballade in F Major*	*Album fur die Jugend*, *opus 68, No 1*
Minor	*Prelude No. 4 in c-sharp minor*	*Ballade in g minor*	*Album fur die Jugend*, *opus 68, No 9*

The 37 color stimuli

Testing Domain Specificity: Conceptual Knowledge of Living and Non-living Things

Eshin Jolly

UNIVERSITY OF ROCHESTER

Among current theories regarding the organization of conceptual knowledge, there is an accepted distinction between living and non-living things in ventral stream processing. Less clear, however, is what drives the organization of concepts and whether "living" is a unitary, coherent concept category. Using functional neuroimaging and three behavioral paradigms, this study aimed to test how verbal association differentially affects the categorization of plants, how animacy affects the categorization of animals, and how motor production affects the categorization of manipulable non-living things. The results reveal that linguistic labels may not strongly influence the organization of categories. Moreover, motor production only shows some differential effect between category domains (non-living more than living) and no effect within category domains (manipulable compared to non-manipulable objects). Animacy, however, does differentially affect the categorization of animals and no other category items, implicating a shared mechanism for the detection of biological motion and the categorization of animals. Finally, a meta-analysis of previous functional imaging literature reveals significant neural overlap between cortical areas responding to animals and those responding to "socialness." This finding supports the theory that social concepts might better explain the living/non-living dichotomy through shared mechanisms responsible for detecting animacy, thereby implying causality and intentionality within the domain of agency. Still unclear is how object categories such as plants, which are living but non-social, fit into the organization of conceptual knowledge.

INTRODUCTION

The organization of conceptual knowledge is a much-debated area of study motivated by research from fields including neuropsychology, cognitive neuroscience, and developmental psychology. Neural processing of concept categories is localized in the ventral stream. While current researchers accept a dichotomy between living and non-living objects, significant deliberation surrounds the origins of these distinctions (e.g., Caramazza & Shelton, 1998; Martin, 2007; Martin, Ungerleider, Haxby, 1996; Moore & Price, 1999; Warrington & McCarthy, 1987; Warrington & Crutch, 2003; for a review, see Mahon & Caramazza, 2003). This dichotomy has been observed in both neuropsychological patient data (e.g., Caramazza & Shelton, 1998; Warrington & McCarthy, 1987) and functional neuroimaging data, in the form of a lateral-to-medial organization of categories in the fusiform gyrus (Mahon & Caramazza, 2009). Lesion data has indicated that lateral regions of the fusiform gyrus show more activation for living things such as animals and faces, whereas medial areas show more activation for non-living things such as tools and vehicles (e.g., Chao, Haxby, & Martin, 1999; Mahon, Millevill, Negri, Runiati, & Caramazza, 2007; Noppeney, Price, Penny, & Firston, 2006; for a review, see Mahon & Caramazza, 2009). The debate within this field of study regards *why* patient deficits and functional neuroimaging data reveal this particular pattern of conceptual knowledge organization. Since the first case of a category-specific deficit was reported, over 100 such cases have been observed; the most common pattern of category-specific deficits is an impairment in either biological living categories or non-biological inanimate objects (Martin & Caramazza, 2003). Two major theories that attempt to provide an explanation include the Sensory/Functional Theory (S/FT; Warrington & McCarthy, 1987; Warrington & Crutch, 2003) and the Domain Specific Hypothesis (DSH; Caramazza & Shelton, 1998). Originally developed around the observation of neuropsychological patient testing, each theory attempts to explain the origin of category-specific deficits exhibited by these patients.

Sensory/Functional Theory attempts to explain category specificity as a product of

The author would like to thank Dr. Jessica Cantlon for all of her guidance, sincere encouragement, and support; Dr. Brad Mahon for his support, particularly with paradigm implementation, MATLAB scripting, Brain Voyager QX analysis, and sharing his fMRI data set; Dr. Elissa Newport for her support as a thesis committee advisor; Robert Emerson for his guidance with E-Prime and experiment programming; and Rosa Li for her guidance with IRB protocols and logistics. This research was performed to satisfy the requirement for a degree with honors in the department of Brain and Cognitive Science at the University of Rochester. Research was completed with funding from Dr. Jessica Cantlon and the Concepts, Actions, and Objects Laboratory at the Rochester Center for Brain Imaging through the University of Rochester. The author can be reached at eshin.jolly@gmail.com or at Harvard University Center for Brain Science, Northwest 290, 52 Oxford St., Cambridge, MA 02138.

experience-based semantic knowledge of object properties. More specifically, S/FT posits that conceptual knowledge is organized into clusters of common sensory features for living things (e.g., shape, form, color) and common functional properties for non-living things (e.g., behaviors with regard to use or purpose of use; Warrington & McCarthy, 1987). In a foundational study, Warrington and McCarthy (1987) observed a patient, Y.O.T., who suffered from severe global aphasia and noted selective impairments not only for broad categories of semantic knowledge, but also for very specific subsets of these categories. Y.O.T. demonstrated a significant impairment for categories of non-living objects as compared to foods or living things, and small manipulable objects as compared to large man-made objects. Research in support of S/FT specifically sought to identify categories that were impaired together to find similarities between objects within these categories. For example, earlier work by Warrington and Shallice (1984) showed that categories of animals, plants and foods ("living" things) tended to be impaired or spared together. These types of observations led researchers to hypothesize a system of multiple processing channels based on an interactive network of relative weighted inputs, which now forms the basis of the S/FT (Warrington & Crutch, 2003). The weightings are thought to be based on modality-specific experience such as visual features, which are assumed to be a more critical component for identifying and differentiating a living animal, versus functional use, which is assumed to be more important in the identification of a non-living object such as a tool (Warrington & McCarthy, 1987; Warrington & Crutch, 2003). Further support for this theory has come from more recent studies, such as Borgo and Shallice (2001; 2003), who reported an association between impairment for living things and impairments for sensory-quality categories. Thus, for S/FT theorists, category-specific deficits occur because information is organized according to its sensory inputs, which results in a categorical organization of objects in the ventral stream.

As an alternative explanation for the observed organization of conceptual knowledge, Caramazza and Shelton (1998) outlined the Domain Specific Hypothesis (DSH), which draws upon the principle of modularity (Fodor, 1983). Specifically, the DSH suggests that the information is constrained by evolutionary processes in the human brain (Caramazza & Shelton, 1998; Mahon & Caramazza, 2009). In this manner, category-specific deficits exhibited by patients are examples of broad category domains formed through pressures of natural selection and evolutionary relevance and importance (Martin & Caramazza, 2003; Mahon & Caramazza, 2009; Caramazza & Shelton, 1998). The basis for the DSH comes from the observation of a patient, E.W., who exhibited modality independent, category-specific deficits for animals, vis-à-vis other livings things and all non-living things (Caramazza & Shelton, 1998). These observations lead to the idea that conceptual knowledge is domain-specific. Rather than being organized by sensory features, the DSH argues that domains are defined by evolutionary processes that have acted on the semantic content of information. Thus, category-specific deficits in patients should reflect impairments with all types of knowledge about the domain (Mahon & Caramazza, 2009). A case study by Farah

and Rabinowitz (2003) supports this idea: Adam, who suffered a stroke at 1 year of age, failed to acquire both visual and non-visual knowledge about living things, but displayed normal levels of understanding about non-living things 15 years later. Further support for the DSH comes largely comes from work in developmental psychology, which demonstrates innate "knowledge" in human infants that is not mediated by experience with modality-specific associations (e.g., Mandler & McDonough, 1993; Carey & Spelke, 1994; Gelman, 1990).

Despite the differences between the theories, S/FT and the DSH agree that there is substantial evidence of a neural dissociation between the processing of living and non-living things in both neuropsychological and functional neuroimaging data (e.g., Chao, Weisberg, & Martin, 2002; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009; Martin & Chao, 2001; Warrington & Crutch, 2003). Whereas S/FT and the DSH attempt to explain *why* category-specific deficits emerge, the focus of the study at hand is *what drives* the organization of concepts. That is, what types and forms of inputs are relevant to categorizing objects as living or non-living?

Early work by Piaget (1929/1960) pointed to the importance of "animism" (animacy) or biological motion in the emergence of a concept of "living." He described how children often attribute life to things that move or seem to move. For example, a child thinks that a bicycle is alive just as a cat is alive because it "goes," but a table is not alive because it cannot move (Piaget, 1929/1960). In his work, Piaget also showed that children have particular difficulty considering plants as alive. He argued that since plants are stationary and acted upon (rather than being actors), they are typically considered non-living by young children (Piaget, 1929/1960). In addition to animacy, Carey (2009) describes the importance of "agency"; the cognitive representation of action intentions, goals, causality and self-directed motion is important in the categorization of something as alive. For example, infants show surprise when non-living objects, such as a beanbag, seem to be capable of self-motion, and seek an external causal explanation, such as a human hand (Saxe, Tzelnic, & Carey, 2006). In another study, Wagner and Carey (2005) reported that after viewing various scenes of object motion, infants show an understanding of action intentions (that one ball is trying to "chase" another) through habituation to the appropriate end scenes.

Language, propositions, and verbal labeling have been proposed as additional means for children to group objects with feature-based similarities (e.g., Waxman & Markow, 1995; Woodward & Markman, 1998). For example, studies have shown the importance of words in the "formation of sortal concepts" (Bloom & Keil, 2001; Xu, 1999; Xu & Carey 1991), and several researchers have argued that "language is used to partition the world" into meaningful units (Berlin, 1992; Goldberg & Thompson-Schill, 2009). Waxman and Markow (1995) have shown that children as young as 12-months of age, during early stages of lexical acquisition, use words and object names to guide the formation of categories.

More recent studies provide further evidence that motion processing and verbal learning influence the categorization of living things. For example, several studies have recently shown that children can correct the kinds of errors that they exhibit in Piaget's

living/non-living studies once they understand the difference between types of motion and animacy; once children understand the cause of an object's movement and the different kinds of motion exhibited by living and non-living things, they reshape their living object category to include non-moving, living objects such as plants (Goldberg & Thompson-Schill, 2009; Opfer & Siegler, 2004). This developmental change in the organization of the "living" concept (i.e., to include plants) is thought to reflect a qualitative difference between mature systems of knowledge organization employed by adults and undifferentiated child based systems, with the assumption that the child system no longer "plays any role in the adult conceptual system" (Carey, 1988, 1999). Surprisingly, however, adults and even domain-specific experts exhibit remnants of child-like errors, especially in their response times, when performing a category decision task involving animals and plants (Goldberg & Thompson-Schill, 2009). Using a word-based speeded category decision task (living or non-living), Goldberg and Thompson-Schill reported significantly lower accuracy and longer response times for the categorization of plants as living things as compared to animals in a population of college undergraduates and biology professors (domain-specific experts). Although the difference between plants and animals was reduced by nearly two thirds for professors, the authors suggested that experience and expertise may only provide a "subtle reconstruction of conceptual knowledge." Additionally, they argued that conceptual categorization involves perceptual similarities based on "coarsely coded associations" that are largely established in early development and mediated through the use of language. These studies are important in highlighting the role of biological motion in the categorization of animals as living things and the role of verbal or propositional processing in the categorization of plants as living things.

A final domain in which different input variables appear to influence the categorization of living and non-living things is the domain of "tools." Recent studies have shown that information relevant for action (motor preparation and planning) plays a differential role in the categorization of tools compared to other non-living objects (e.g., Mahon et al., 2007). This evidence is based primarily on fMRI data wherein people exhibit greater functional connectivity between parietal action regions and "non-living" object processing regions of the ventral stream when the stimulus is a tool than when it is a non-manipulable object (Mahon et al., 2007). The central argument here is that motor-relevant processing contributes more toward the categorization of tools than other objects within the "non-living" domain.

If there is a pure dissociation between living and non-living things in the organization of conceptual knowledge, then why do plants cognitively dissociate from animals within the domain of "living"? And, why do tools engage neural structures that are not engaged by other non-living objects? Are different input variables important for the categorization of animals, plants, tools, and non-living things? To answer these questions, we aimed to test whether verbal associations, motor information, and animacy each play a unique role in the categorization of animals, plants, tools, and other objects. We used a generic behavioral categorization paradigm based on Goldberg and Thompson-Schill's (2009) study, coupled with a

series of cognitive distracter tasks (verbal, motor, and biological motion judgment tasks), to test whether these tasks would differentially impact category judgments of living and non-living things. In light of previous research, we hypothesized that verbal distraction would differentially impair the categorization of plants, motor distraction would differentially impair the categorization of tools, and animacy distraction would uniquely impair the categorization of animals.

STUDY 1: NEURAL CORRELATES OF CATEGORY DIFFERENCES

A novel re-analysis of fMRI data collected by Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza (2009) was performed to compare neural activation related to processing animals and plants within the domain "living," in the ventral stream. It was hypothesized that if category domains are organized as living and non-living, then animals and plants should not differ significantly in the cortical activation within lateral areas of the fusiform gyrus (which have been associated with processing living things in prior neuropsychological and neuroimaging studies as described above).

Methods

Participants

Twenty-one participants (9 males; 12 females; ages 20-51 years old, *M* = 31.2) were recruited from the Center for Mind/Brain

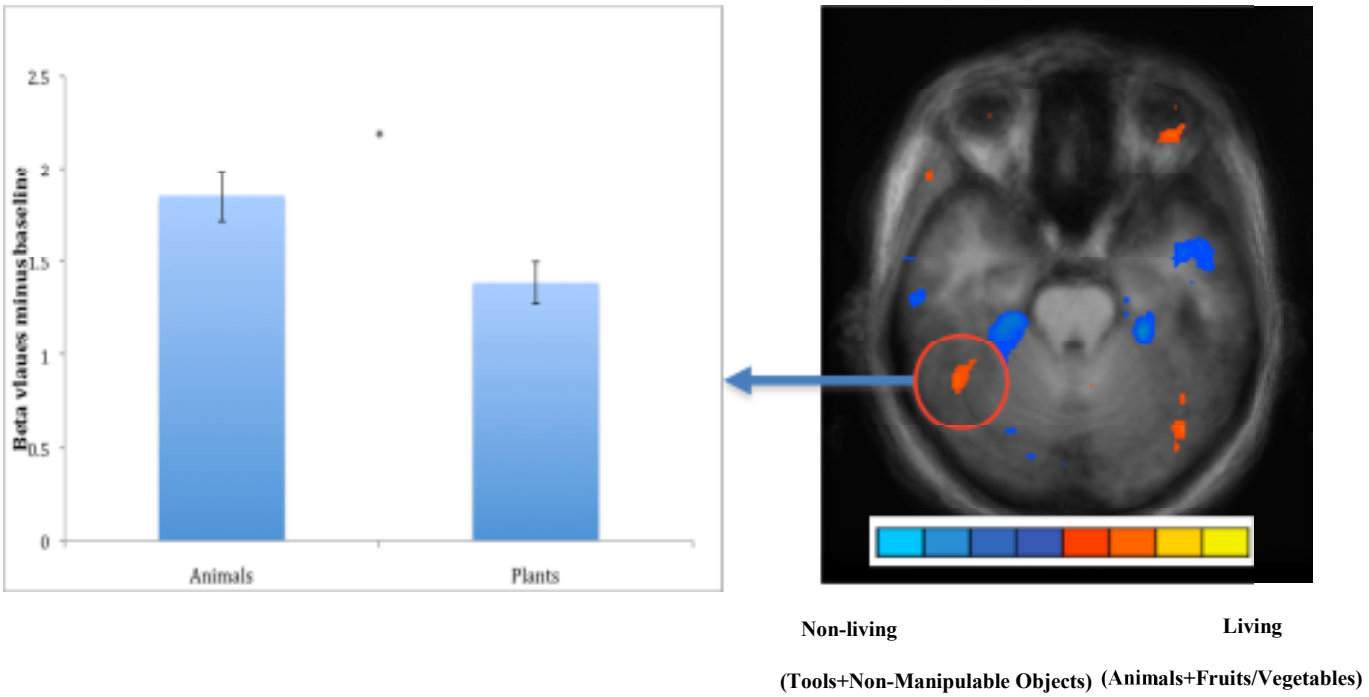


Figure 1. Shown on the right is the activation contrast for lateral (living) and medial (non-living) areas in the ventral occipital-temporal cortex. On the left is the comparison between activation for animals and activation for plants within the lateral ROI (peak activation 42, -46, -20; $t(19) = 3.13$, $p < .001$), which are significantly different, $t(19) = 4.7$, $p < .001$, and indicate more activation for animals than plants.

Science volunteer pool at the University of Trento, Italy, and were paid for their participation.

Materials and procedure
Stimuli consisted of four object categories of 24 pictures each: animals, tools, non-manipulable objects, and fruits/vegetables. While BOLD responses were recorded, participants passively viewed pictures in 20 second blocks, followed by 20 seconds of blank screen (fixation). Each block contained 24 pictures of the same category, each stimulus presented for 50 refreshes of the monitor, and each category block (24 pictures) was repeated three times per run. The order of blocks was randomized; each run lasted approximately 10 minutes.

Imaging procedure
Cerebral images were acquired by Mahon et al. (2009) on a Bruker BioSpin MedSpec 4T MRI machine. They collected functional data using an echo planar 2D imaging sequence with phase over-sampling (Image matrix: 70 × 64, TR: 2250 ms TE: 33 ms, Flip angle: 76°, Slice thickness = 3mm, gap = .45mm, with 3 × 3 in plane resolution). Volumes were acquired in the axial plane in 37 slices and the acquisition order was ascending interleaved odd-even.

Data analysis
Images were pre-processed (slice time correction, realignment, normalization) and transformed into Talairach space by Mahon et al. (2009) using Brain Voyager QX. The authors applied BOLD activity to the individual subject data, which included parameters for stimulus onsets and motion parameters. This individual-subject regression model provided an estimate of the strength of activation related to each category for each subject. We used these estimates to determine regions that responded to Living Things > Non-living Things at the group level. That analysis defined our regions of interest (ROIs). Then, we looked at the individual-subject responses within these ROIs to test the relative activation to plants versus animals at the group level, thresholded at $p < .05$.

Results and Discussion

Cortical activation was mapped for areas representing living things (animals and fruits/vegetables) and non-living things (tools and non-manipulable objects). As shown in the right panel of Figure 1, data were representative of previous functional imaging findings with significant lateral-to-medial dissociation of living and non-living things in ventral occipital-temporal cortex, $p = .05$, uncorrected (e.g., Chao et al., 1999, 2002; Mahon & Caramazza, 2009; Martin, 2007; Martin, et al., 1996). Additionally, a significant difference was observed between animals and plants within the lateral ROI in right hemisphere ($t(19) = 4.7$, $p < .001$). However, as shown in the left panel of Figure 1, animals exhibited significantly greater activation in this area than plants, suggesting that not all living objects contribute equally to cortical activation of the lateral regions of the ventral stream. Thus, cortical organization may not be indicative of a "living versus non-living" dissociation because not all living objects carry an equal weight within this domain. To further investigate why plants are not represented like animals in the ventral stream and to test what drives the organization of concepts of living and non-living, we performed a series of behavioral interference experiments.

STUDY 2a: VERBAL ASSOCIATIONS AND PLANTS

The first experiment was designed to test whether verbal learning accounts for the difference in performance when categorizing plants and animals as living things as reported by Goldberg and Thompson-Schill (2009). This was performed by first replicating the previous study and then by adding a secondary verbal distracter condition. Distraction was created by asking participants to sing the children's nursery rhyme "Twinkle, Twinkle, Little Star" while simultaneously completing the living-non-living categorization task. It was hypothesized that if verbal learning and linguistic labels are employed as a method of associating plants with the "living" domain, then performance for judging them as such should be significantly worse when a verbal distracter is employed.

Methods

Participants

Twenty-one participants (10 males; 11 females; ages 18-21 years old, *M* = 19.8) were recruited from the undergraduate subject pool of the departments of Psychology and Brain and Cognitive Sciences (BCS) at the University of Rochester. Subjects were asked to participate in a 1-hour behavioral BCS study, which was held at the Rochester Center from Brain Imaging, and were paid $8 for their participation.

Materials

Word items were taken from Goldberg and Thompson-Schill (2009) and contained 140 items in six different categories. The set consisted of 30 animals, 30 plants, 20 non-moving artifacts (e.g., towel), 20 non-moving naturals (e.g., rock), 20 moving artifacts (e.g., bus), and 20 moving naturals (e.g., river). All words within each domain (living and non-living) were matched and

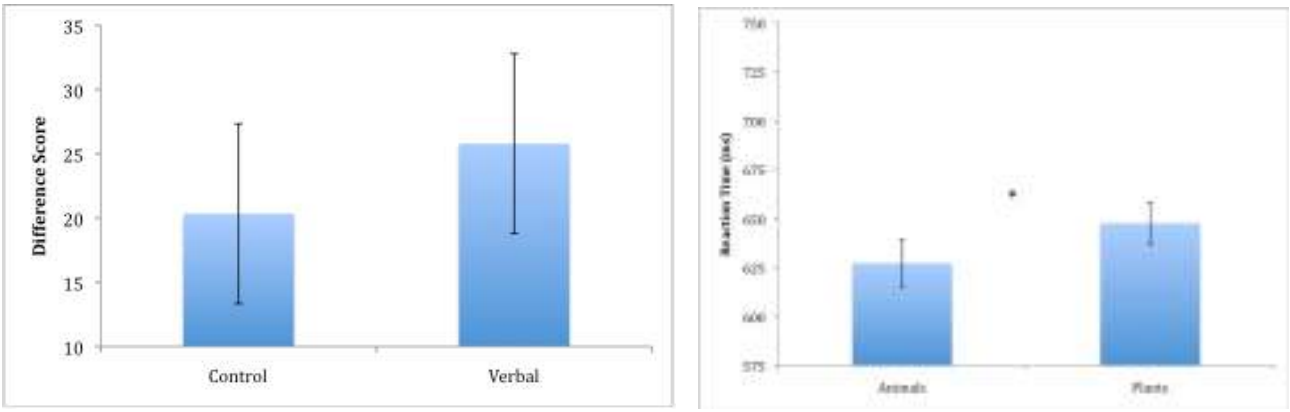


Figure 2. Results from Study 2a: (a) Participants' mean reaction times for identifying plants (*M* = 647.6 +/- 48.25) and animals (*M* = 627.2 +/- 54.45) as living (correct trials only). A significant difference was observed between animals and plants, $t(20) = 2.62$, $p < .05$. (b) Difference scores for the control condition (*M* = 20.34 +/- 7.76) and the verbal distracter task (*M* = 25.79 +/- 6.47). No significant difference was observed. Error bars represent standard error of the mean.

balanced for number of letters, syllables and frequency (for details refer to Goldberg & Thompson-Schill, 2009).

Procedure

Participants were told to follow onscreen instructions, which appeared in two varieties. In the control condition participants were instructed to indicate, as quickly and accurately as possible, whether the word that appeared onscreen represented a living or non-living object, by pressing either of two keys using their dominant hand. In the verbal distracter condition, participants were instructed to perform the same judgment task while at the same time holding a small microphone with their non-dominant hand and singing "Twinkle, Twinkle, Little Star" aloud. Each trial began with the presentation of a visual fixation-cross for a random duration between 500-1500 milliseconds followed by a random word from the stimuli set for a maximum of 1000 ms. If subjects did not make a decision within the allotted time, "no response detected" appeared on the screen for 500 ms. Each trial ended with a visual word mask (e.g., #####) for 500 ms.

Statistical testing

Data analysis was performed on reaction times for responses to correct trials only in both conditions. In the control condition (replication), data were analyzed using a paired sample t-test between reaction times to animals and plants. In the experimental condition (verbal distracter), difference scores were calculated for each category (plant RT-animal RT) and conditions were compared using a paired sample t-test.

Results

As shown in Figure 2, the results of Goldberg and Thompson-Schill (2009) were replicated successfully. Participants exhibited a significantly slower reaction time of 20.4 ms for plants ($M = 647.6$ +/- 10.53) as compared to animals ($M = 627.2$ +/- 11.88), ($t(20) = 2.62$, $p < .05$). However, when control difference scores ($M = 20.34$ +/- 7.76) were compared to the difference scores in the verbal distracter condition ($M = 25.79$ +/- 6.47), no significant difference was observed, ($t(20) = .62$, $p = .55$), suggesting no further impairment of either category judgment through the use of a verbal distracter task.

Discussion

Despite previous research indicating the importance of language use in forming certain categories during development (e.g., Bloom & Keil, 2001; Waxman & Markow, 1995; Xu, 1999; Xu & Carey 1991), our results do not support this theory for the categorization of animals and plants. Results were expected to indicate a greater difference between reaction times for animals and plants in the verbal distracter condition based on the idea that language is used as an association tool to link plants to the domain of living things along with animals (Goldberg & Thompson-Schill, 2009; Woodward & Markman, 1998). However, it seems that there is no significant change in reaction time for the categorization of living and non-living things during verbal distraction. Thus, an inability to access verbal processes does not disrupt the categorization of plants as living things, and verbal learning is not necessarily something driving the organization of conceptual knowledge in this case.

STUDY 2b: BIOLOGICAL MOTION AND ANIMALS

The aim of the second experiment was to test the importance of animate processing on the categorization of conceptual knowledge. Here, we used a dual-task paradigm to test the interference of biological motion processing on the categorization of living and non-living things. Subjects were asked to judge the gender of a point-light-walker (PLW) immediately preceding a categorical judgment of a word as living or non-living. We manipulated the distance between the point-light-walker judgment and the categorical judgment in a stimulus onset asynchrony (SOA) design to assess interference effects during the psychological refractory period (PRP). To help clarify the methods, predictions, and results of this experiment, it is important to first explain the SOA paradigm, the PRP, and the bottleneck model of cognitive processing.

Dual-Task Interference and the SOA paradigm

Starting with the early work of Telford (1931), the PRP effect was shown to be the delaying of response to a second stimulus in an experimental design where the delay between two stimuli is asynchronous. This delayed response becomes greater as the distance between the stimuli is reduced and occurs even when the response types are of differing modalities (Pashler, 1994). Although many theories have been proposed as to the reason for the effect, one of the most prominent and relevant to the study at hand is the bottleneck or task-switching model (Sternberg, 1969). This model describes that if two stimuli require the same type of mechanism to be processed, then when both stimuli require the use of that mechanism simultaneously, a processing bottleneck occurs; because the first stimulus occupies the mechanism first, delays occur in the processing of the second stimulus until the first has been finished processing due to the necessary mechanism switching (Pashler, 1994; see Figure 3).

Many studies have replicated this effect and have shown that the bottleneck is not a voluntary aspect of stimulus processing and therefore cannot be explained by such things as temporal uncertainty or physical limitations (e.g., Bertleson, 1967; Broadbant & Gregory, 1967; Pashler & Christian, 1994; Pashler & Johnston, 1989; Sternberg, 1969; Telford, 1931; for a review, see Pashler, 1994). Additionally, key outcomes of an SOA paradigm include a high correlation between reaction time to the first and second stimulus at short SOAs (Gottsdanker & Way,

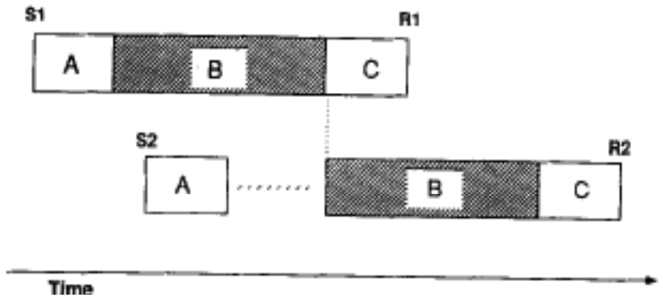


Figure 3. A central bottleneck model schematic: the shaded part of stimulus 2 (mechanism-based processing) cannot occur until the shaded part of stimulus 1 is complete (Pashler, 1994).

1966; Pashler, 1994; Pashler & Johnston, 1989; Welford, 1967), as well as a change in reaction time to the second stimulus, which approaches a slope of -1 (Pashler, 1994; Telford, 1931). Both of these characteristics are predicted by the bottleneck model. They occur because the necessary processing mechanism is occupied by the first stimulus, rendering this mechanism temporarily unavailable for processing the second stimulus and thereby introducing a delay (Pashler, 1994).

For these reasons, the SOA paradigm was chosen for Study 2b to test the hypothesis that a system or mechanism that is used to process biological motion (animacy) is also used in the categorization of animals as living things. If this hypothesis is true, then the results should show an effect of SOA on the categorization of animals only, and the difference in performance between categorizing animals and plants as living things (observed in the previous experiment) should be eliminated.

Methods

Participants

Twenty-two participants (12 males; 10 females; ages 19-22 years old, *M* = 20.1) were recruited from the undergraduate subject pool of the departments of Psychology and Brain and Cognitive Sciences (BCS) at the University of Rochester. Subjects were asked to participate in a 1-hour behavioral BCS study, which was held at the Rochester Center from Brain Imaging and were paid $8 for their participation.

Materials

Subjects performed two tasks separated by a brief SOA: a biological motion judgment on PLWs and a living/non-living categorization on words. The same word stimuli that were used in the previous experiment for presentation were used for the second task, and 3second video clips of left- and right-facing "male" and "female" PLWs were used for the first task (Biomotion Lab at Queen's University, Kingston, Ontario). PLWs were chosen because of their accepted use as representations of biological motion in experimental testing paradigms (e.g., Chang & Troje, 2008; Troje, Westhoff, & Lavrov, 2005; Zhang, & Troje, 2005). Based on pilot testing performed prior to the experiment, the SOAs used included: 96 ms, 208 ms, 400 ms, 800 ms, 1600 ms, 2000 ms. In order to produce the strongest statistical PRP effect, the distribution of SOAs was normalized within each stimulus category, such that the greatest number of SOAs included those in the range of 200-800 ms.

Procedure

Participants were told to follow a set of on-screen instructions, which came in two varieties: a control condition, which was the same as the control condition in the previous experiment, and an experimental condition in which participants engaged in the dual-task SOA paradigm. Each experimental trial began with a fixation-cross presented at a random interval from 500-1500 ms followed by the first stimulus, the PLW, presented for 3000 ms in total. Participants were required to judge whether the PLW on screen was a male or female and indicated their response by pressing one of two buttons with their right hand. During this time, a random word stimuli appeared after a specific SOA duration and participants had the remainder of the time (800 ms – 3000 ms based on the length of the delay) to indicate whether the word represented a living or

non-living thing, by pressing one of two buttons with their left hand.

Statistical testing

Data from non-living categories were collapsed into only two categories: moving and non-moving objects. This was done to observe a possible interaction between animacy and the property of movement among non-living objects. Data analysis was performed first for the purpose of checking the effectiveness of the dual-task interference using a correlation between the reaction times to the first and second stimuli. Secondly, changes in slope between categories for the SOA range of 200-800 ms were observed. This range was chosen because of physical limitations on reaction time given the complexity of the task and because pilot testing had shown that for the stimuli used, this range best exhibited dual-task interference. Paired sample t-tests were performed comparing the changes in slope between all categories. In addition, a two-way repeated measures ANOVA was performed between the collapsed reaction times of the 200-800 ms SOA range and the control condition.

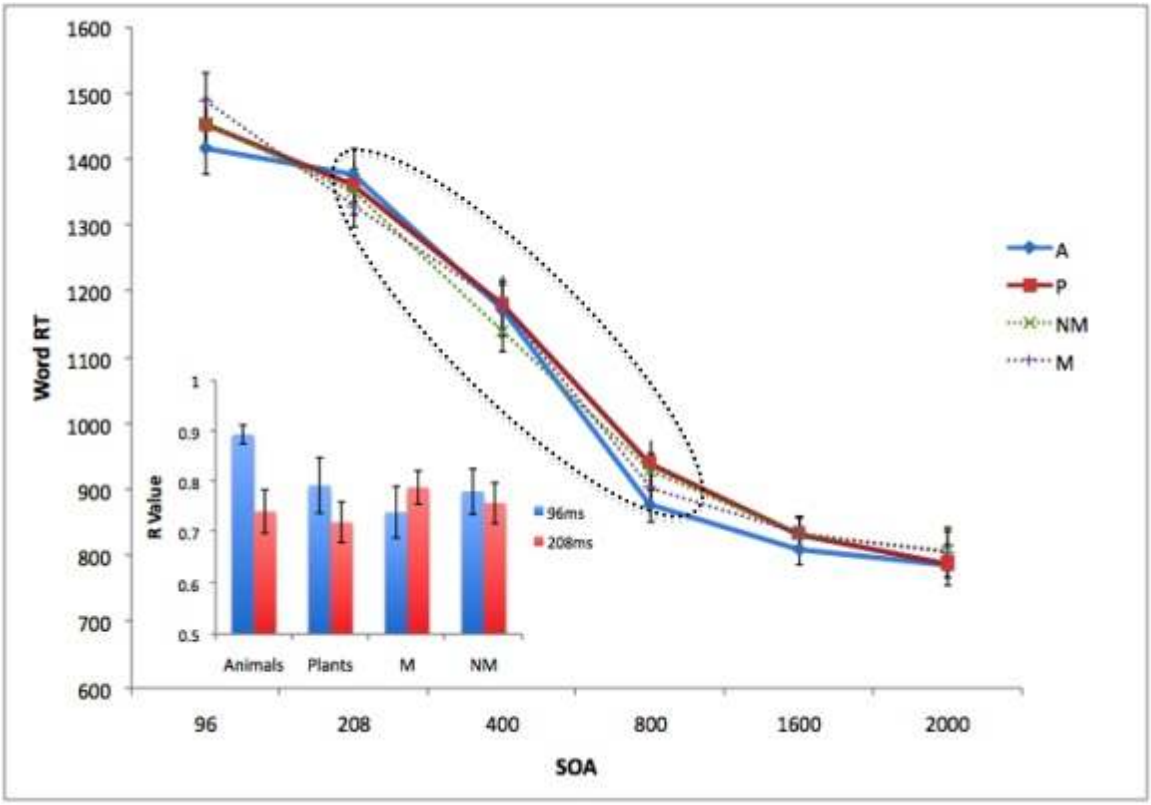


Figure 4. Results from the second experiment: correlations between reaction times for both stimuli were strong and positive $r(21) > .7$, $p < .05$, for all categories at short SOAs. Additionally, the change in reaction time (slope) for animals (M = -.84 +/- .04) was exclusively affected by dual-task interference compared to all other categories and showed the strongest PRP effect, approaching a slope of -1. A: animals; P: plants; NM: non-moving non-living things; M: moving non-living things.

Results

As indicated by Figure 4, correlations between reaction times for the biological motion and living/non-living stimuli were strong and positive ($r(21) > .7$, $p < .05$) for all categories at short SOAs, indicating a significant effect of dual-task interference. Upon comparison of slopes (within the 200-800 ms SOAs), significant differences were observed between animals ($M = -.84$ +/- .04) and plants ($M = -.71$ +/- .04, $t(21) = 2.85$, $p < .01$); animals and non-moving objects ($M = -.72$ +/- .04, $t(21) = 2.58$, $p < .05$); and animals and moving objects ($M = -.72$ +/- .03, $t(21) = 3.57$, $p < .01$). No significant differences were observed between plants and moving objects, ($t(21) = .16$, $p = .87$); plants and non-moving objects ($t(21) = .11$, $p = .91$); or moving and non-moving objects, ($t(21) = .08$, $p = .94$). These results suggest a specific effect of dual-task interference as a function of SOA on the categorization of animals only and not on the categorization of plants or non-living objects (irrespective of their motion properties).

A two-way analysis of variance for plants and animals (category × stimulus onset) yielded a main effect for stimulus onset ($F(1, 42) = 604.1$, $p < .001$), indicating a significant difference between the comparison of animals and plants in the control condition and in the SOA condition (200-800 ms). No main effect was observed for category ($F(1, 42) = .40$, $p = .53$) and there was no observable interaction ($F(1, 42) = .45$, $p = .50$). As

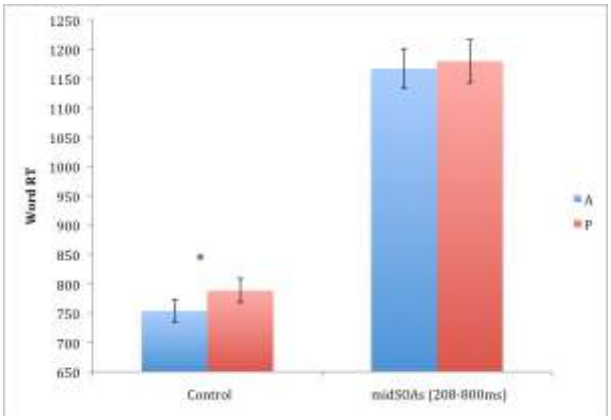


Figure 5. Comparison of animal and plant reaction time differences between control and SOA conditions. No significantly slower reaction time to categorizing plants as living, $t(21) = 4.25$, $p < .001$, was observed in the SOA condition, $t(21) = .75$, $p = .46$. A: animals; P: plants.

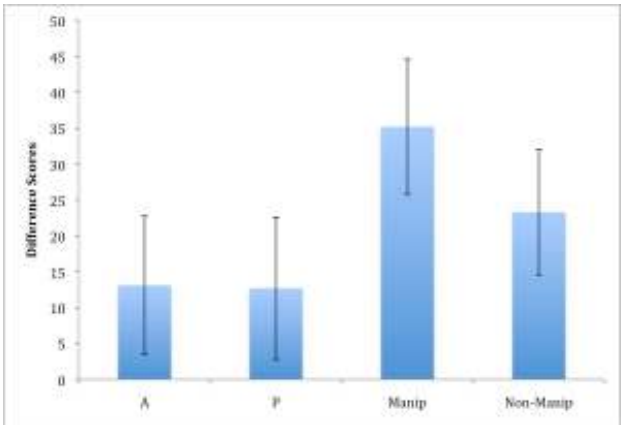


Figure 6. Difference scores between control and motor distracter task conditions for all category items. Significant differences were observed between manipulable objects and animals, $t(20) = 3.09$, $p < .01$, and plants, $t(20) = 2.41$, $p < .05$. However, no differences were observed between category items within the non-living domain (manipulable compared to non-manipulable words). A: animals; P: plants; Manip: manipulable non-living things; Non-Manip: non-manipulable non-living things.

seen in Figure 5, comparisons of animal (M = 1167.27 +/- 33.91) and plant (M = 1179.74 +/- 36.69) categorization reaction times for the SOA condition showed no significant difference ($t(21) = .75$, $p = .46$). This finding from the experimental condition contrasts with the comparison of animals (M = 753.9 +/- 18.98) and plants (M = 788.4 +/- 20.58) in the control condition, which, as expected, exhibited a significant difference of 34.5 ms, with plants being slower than animals, ($t(21) = 4.25$, $p < .001$). These results suggest that the dual-task biological motion SOA paradigm eliminated the previously observed categorization difference between animals and plants by slowing down judgments of animals to the level of plants.

Discussion

These findings implicate the role of a single mechanism used to detect biological motion in the categorization of living things. Dual-task interference was observed only for animals and not for plants or non-living categories, suggesting the occurrence of a processing bottleneck as two different stimuli called upon a single mechanism in succession (Pashler, 1994). If a processing bottleneck was not occurring, then results should have shown a more distributed and general dual-task interference across all categories. However, the exclusivity of the PRP effect observed for animals suggests that the mechanism used to detect animacy is also used to categorize animals as living things. Access to this mechanism creates a bottleneck of cognitive resources resulting in a performance deficit in categorization. These results, along with the lack of difference between performance in categorizing animals and plants during the biological motion dual-task paradigm, suggest that that the decrease in the performance gap occurs because of the impairment of animal categorization, not an improvement of plant categorization, during biological motion processing. This specific effect on animal categorization alone strongly suggests the importance of animacy detection as a mechanism for the formation of concept categories.

STUDY 2c: MOTOR ACTIONS AND MANIPULABLE OBJECTS

This experiment tested the relationship between motor action processing and the categorization of manipulable non-living things. Using the same general experimental design as Study 2a, a secondary motor distraction task was employed by asking participants to simultaneously draw concentric circles as they made living/non-living judgments on word stimuli. It was hypothesized that if the categorization of concepts pertaining to manipulable objects is influenced by systems responsible for carrying out motor actions, then interference would be observed exclusively for manipulable non-living objects compared to living objects in the motor distractor paradigm.

Method

Participants

The same twenty-one participants from Study 2a were tested in this experiment (10 males; 11 females; ages 18-21 years old, M = 19.8).

Materials

The same word items from the first two experiments, taken from Goldberg and Thompson-Schill (2009), were used in this experiment.

Procedure

Participants completed the motor action and living/non-living categorization task as a block of 140 (word) trials either before or after the sessions from Study 2a (order was counterbalanced). The control condition was the same control session as that of Study 2a. In the motor distracter condition, participants were instructed to perform the living-non-living judgment task with their dominant hand, while repeatedly drawing spirals on a sheet of paper with their non-dominant hand. As in Study 2a, each trial began with the presentation of a visual fixation-cross for a random duration between 500-1500 ms followed by a random word from the stimuli set for a maximum of 1000 ms. If subjects did not make a decision within the allotted time, "no response detected" appeared on the screen for 500 ms; each trial ended with a visual word mask (e.g., #####) for 500 ms.

Statistical testing

Data analysis was performed on reaction times for responses to correct trials only in both conditions and words were binned into four object categories: animals, plants, manipulable objects, and non-manipulable objects. Difference scores were calculated for each category RT (e.g., manip (motor distract) – manip (control)), and category scores were compared using paired sample t-tests.

Results and Discussion

As seen in Figure 6, there was a significant difference in performance between: manipulable objects (M = 35.20 +/- 9.37) and animals (M = 12.64 +/- 9.61; $t(20) = 3.09$, $p < .01$) and manipulable objects and plants (M = 12.72 +/- 9.90; $t(20) = 2.41$, $p < .05$). No significant difference was observed between performance for non-manipulable objects (M = 23.29 +/- 2.72) and animals, ($t(20) = 1.02$, $p = .32$); non-manipulable objects and plants ($t(20) = 1.04$, $p = .31$); or manipulable and non-manipulable non-living things ($t(20) = 1.41$, $p = .17$). These results suggest that motor information may be more important for the categorization of non-living things than living things as indicated by differential impairments in performance for non-living category words. However, these results are less transparent for interpreting the role of motor action on the categorization of manipulable versus non-manipulable objects. There was a measureable difference between living things and manipulable objects but not living things and non-manipulable objects. This lack of difference between manipulable and non-manipulable objects raises the question of whether motor actions specifically influence the categorization of manipulable objects within the non-living domain.

GENERAL DISCUSSION, STUDIES 2a-c

We performed three behavioral studies to test the hypothesis that verbal associations, animacy, and motor information are important in the development of the living/non-living dissociation. Taken together, our results indicate important influences of motor and biological motion information on the conceptual distinction between "living" and "non-living."

The results from Study 2a indicate that verbal distraction has no differential impact on the processing of semantic categories. Previous studies that suggest language, as a mechanism to create semantic associations

and group objects with similarities, still provides a viable explanation for the formation of concept categories (e.g., Berlin, 1992; Bloom & Keil, 2001; Waxman & Markow, 1995; Xu, 1999; Xu & Carey, 1991; Woodward & Markman, 1998). As suggested by Goldberg and Thompson-Schill (2009), the "groups" produced through linguistic association may reflect developmental "roots" that are only partially modifiable by further experience. If this is case, then the findings observed in Study 2a are to be expected, as verbal distraction would have little impact on categories that are already well established. If this explanation is correct then we should expect to see a greater effect of verbal distraction on living/non-living categorization in children. Alternatively, it is also possible that subjects were not affected by verbal distraction in our task as the nursery rhyme could have been over-learned or subjects may have used the prosody of the melody ("Twinkle, Twinkle, Little Star"), instead of language in order to recite the rhyme.

Interestingly, the results from the motor interference paradigm of Study 2c indicate a disproportionate impact of motor input on the categorization of non-living things compared to living things. However, the within-category distinction between manipulable and non-manipulable objects is less clear. Although there is a greater effect of motor interference on manipulable objects than on non-manipulable objects, this difference in the interference effect only approaches significance. It is unclear as to why domain dissociations are observable for this experiment (living and non-living) but category distinctions are not (manipulable and non-manipulable). These results could be explained by two different theories: S/FT and the "grounding by interaction" hypothesis (Warrington & McCarthy, 1987; Mahon & Caramazza, 2008). S/FT provides an explanation for these results based on the hypothesis that functional information is more important than sensory features for the categorization of non-living things; objects such as tools, for example, will be categorized based on *perceptions* of characteristics about function and purpose, rather than shape or color (Warrington & McCarthy, 1987; Warrington & Crutch, 2003). In contrast, the "grounding by interaction" hypothesis suggests that conceptual knowledge, particularly for tools, is organized based on both an "abstract" and "symbolic" representation of the concept (knowledge about what it is) and interactions between this knowledge and sensory and motor systems (Mahon & Caramazza, 2008). Under this hypothesis, the instantiation of a concept may require the interaction of specific sensory and motor systems responsible for producing *actions*. Therefore, activation of a particular motor system may complement the "abstract" or "symbolic" knowledge about an object concept. In our study, the motor distracter task apparently interfered with a generic action mechanism that is necessary to categorize many non-living objects. In order to determine whether specific types of motor inference disproportionately impact manipulable objects relative to non-manipulable objects (as in the "grounding by interaction" hypothesis), further research would have to be conducted to specifically create motor interference with a task that is specific to tool manipulation (e.g., grasping, pounding).

The results from Study 2b suggest an alternative explanation as to how concept categories are organized. In light of Piaget's

(1929/1960) early work with children and animacy, as well as Carey's (1999, 2009) more recent work on agency, it is possible that the formation of concept categories is not dissociated along a dimension of living versus non-living things. The exclusivity of the PRP observed for animals and not for other category specific word stimuli in Study 2b suggests a common mechanism (a biological motion system) that is important for both the detection of animacy and the organization of conceptual knowledge for animals. Carey (1988, 1999) suggests that there exists a "qualitative difference" between the organization of a conceptual knowledge system of children and adults. In particular, she posits that understanding the difference between "alive" on the one hand, and "movement" or "naturalness" on the other, is a key component of this difference. The concept "alive" permits animals to be coalesced with plants into a common conceptual category (Carey, 1988, 1999). However, if animals and plants are combined as such, then children, adults, and especially domain-specific experts should not exhibit the category errors and performance differences for plants versus animals that we observed in Study 2a (see also Goldberg & Thompson-Schill, 2009). Our data indicate that differences between animals and plants are due to differences in the inputs that bear on their categorization: biological motion is more important for the categorization of animals as living things than that of plants. The implication of these findings is that adults are not so different from children in this context, in that animacy remains an important dimension for the categorization of living objects throughout the lifespan. The results from Study 2b and prior research suggest that animacy may be a more likely dimension along which conceptual knowledge is organized. Since animacy has been implicated as important to the property of "socialness," (i.e., social versus non-social; Martin & Weisberg, 2003; Ross & Olson, 2010; Zahn et al., 2007), this possibility is further explored in the following section.

STUDY 3: META-ANALYSIS OF CONCEPT CATEGORIES

Several researchers have hinted at the hypothesis that social relevance plays a role in the organization of conceptual knowledge with respect to the living/non-living distinction (e.g., Martin & Weisberg, 2003; Zahn et al., 2007). For example, in a neuroimaging study with social and mechanical vignettes, Martin and Weisberg (2003) implicated a "core system" that is particularly important for understanding social interaction. Their findings indicated the same lateral-to-medial activation for living and non-living objects in the ventral stream; areas associated with living things (i.e., lateral fusiform gyrus) also showed activation for social vignettes, and areas associated with non-living and manipulable objects (i.e., medial fusiform gyrus) also showed activation for mechanical vignettes controlled for visual form and properties (Martin, 2007; Martin & Weisberg, 2003). Similarly, Zahn et al. (2007) suggested that there could exist a specific neural network important for the representation of social concepts, which is mediated in part by the anterior temporal lobes. This network includes overlap with lateral areas of the fusiform gyrus, normally activated for living things. One possibility that emerges from these prior conclusions is that domain-specific processing in the ventral stream is organized according to the

distinction of "socialness" rather than that of being "alive." If this were the case, then the results of Study 2b would be expected, in that animacy, a key component of "socialness" (e.g., Carey 1998; 1999; Martin & Weisberg, 2003; Ross & Olson, 2010) would be exclusively linked to categories within the domain of social (i.e., animals) rather than non-social (i.e., plants). To further investigate the possibility of "socialness" as a dimension along which concepts are organized, a meta-analysis of previous fMRI studies was conducted to examine neural overlap between living things, and social things (Chao et al., 1999, 2002; Martin & Weisberg, 2003; Noppeney et al., 2006; Ross & Olson, 2010). "Socialness," in these studies, was a dimension tested using social vignette stimuli modeled after stimuli created by Heider and Simmel (1944). These vignettes consisted of movie clips of shapes exhibiting biological motion as rated by behavioral testing data and previous literature (Martin & Weisberg, 2003; Ross & Olson, 2010).

Method

Activation maps

Stereotaxic foci representing Talairach coordinates from five different studies (fusiform gyrus and superior temporal sulcus activations) were projected onto a population-averaged brain using CARET (computerized anatomical reconstruction and editing toolkit; Van Essen et al., 2001; Van Essen, 2005; Van Essen & Dierker, 2007; see Figure 7). Foci were grouped by color-spectrum and shape according to category-based activation: circular represents "living" and square represents "socialness."

Statistical tests

Data analysis was performed using a 3 × 2 (Coordinates x Category) ANOVA for each hemisphere separately (analysis was modeled after Kaddosh, Lammertyn, & Izard, 2007). Means and standard deviations were computed for each category (within a hemisphere) separately. To keep the analysis balanced, only coordinate activations represented bilaterally were used in the ANOVA

Study	Category Activation	Task	Left			Right		
			X	Y	Z	X	Y	Z
Chao, Haxby & Martin, 1999	Living	picture-viewing	-38	-58	-10	38	-56	-12
Chao, Haxby & Martin, 1999	Living	picture-viewing	-49	-59	-10	39	-59	11
Chao, Haxby & Martin, 1999	Living	picture-viewing	-42	-59	19	36	-57	-10
Chao, Haxby & Martin, 1999	Living	picture-viewing	-39	-55	-8	53	-54	16
Chao, Weisberg & Martin, 2002	Living	picture-naming	-42	-59	19	53	-54	16
Noppeney et al., 2006	Living	picture/word viewing		n.s		39	-60	-21
Mean*			-42	-58	2	44	-56	4
S.D.*			4	2	16	9	2	14
Martin & Weisberg, 2003	Social	social vignette viewing	-44	-57	-23	41	-52	-15
Martin & Weisberg, 2003	Social	social vignette viewing	-49	-57	17	56	-58	19
Martin & Weisberg, 2003	Social	social vignette viewing	-40	-38	-20	40	-33	-21
Ross & Olson 2010	Social	Heider & Simmel animations	-40	-4	-13	54	-12	-12
Ross & Olson 2010	Social	Heider & Simmel animations	-40	-44	-14	36	-43	-14
Ross & Olson 2010	Social	social word semanticity	-31	-72	-17		n.s	
Mean*			-43	-40	-11	45	-40	-9
S.D.*			4	22	16	9	18	16

Figure 7. Activation foci for the meta-analysis and corresponding cortical maps (Figure 9). Noppeney et al. (2006) and Ross & Olson (2010; social word semanticity) coordinates not included in mean, S.D., or ANOVA analyses. Mean and S.D. rounded to nearest whole integer.

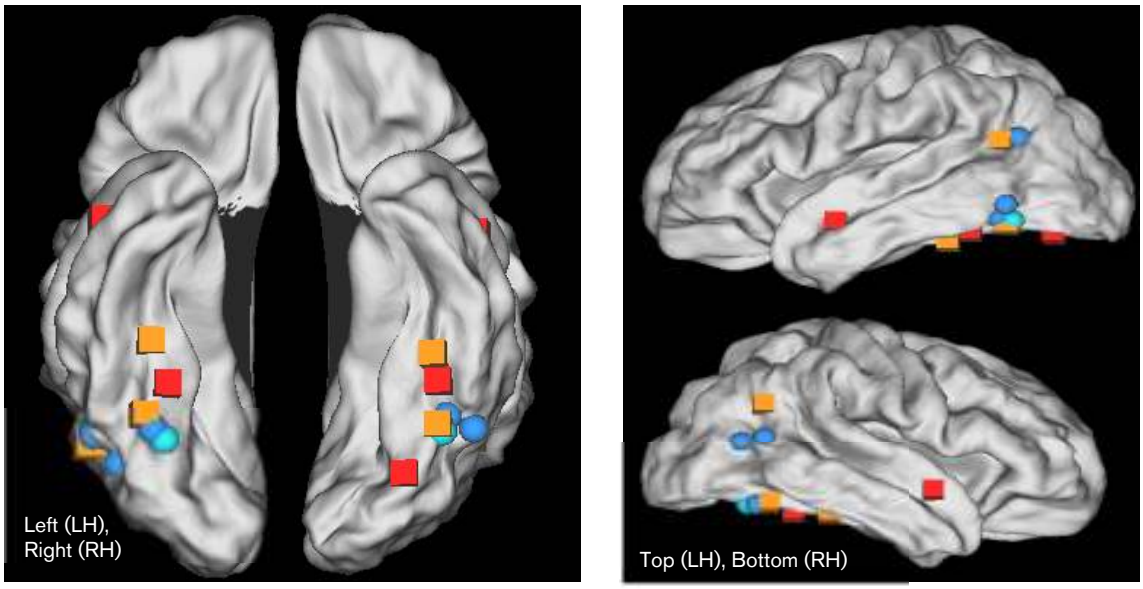
(that is, coordinates representing category specific activation for both hemispheres), and, as such, one set of coordinates per category was not used in testing (living: Noppeney et al., 2006; socialness (word stimuli): Ross & Olson, 2010).

Results

A two-way analysis for left hemisphere acti-

	Living			Social		
	X	Y	Z	X	Y	Z
LH						
Mean	-40	-58	2	-43	-44	-11
S.D.	2	2	16	4	23	16
RH						
Mean	44	-56	4	45	-40	-9
S.D.	9	2	14	9	18	16

Figure 8. Means and standard deviations rounded to the nearest integer for living and social foci coordinates. The main interaction observed in the RH seems to be driven by the difference in Y coordinates between living and social categories.



(a) (b)

Figure 9. A cortical mapping of studies reported in Figure 7. Squares represent social category activation; circles represent living category activations. (a) Ventral view foci. (b) Lateral view foci. No significant difference was observed for activation between categories, LH ((*F*(1, 24) = .01, *p* = .92), RH (*F*(1, 24) = .15, *p* = .71), and a significant interaction was observed in RH (*F*(2, 24) = 3.44, *p* < .05). Results suggest significant neural overlap between areas responsible for categories of living objects and categories of "socialness." Color code: *light blue*, Chao et al. (1999); *turquoise*, Chao et al. (2002); *light green*: Noppeney et al. (2006); *orange*: Martin & Weisberg (2003); *red*: Ross & Olson (2010).

vation yielded a main effect for coordinates ($F(2, 24) = 34.63$, $p < .001$), indicating a significant difference between all three dimensions (x,y,z) of the Talairach Coordinate System (Talairach & Tournoux, 1988; see Figure 8). No main effect was observed for category (($F(1, 24) = .01$, $p = .92$), and there was no interaction ($F(2, 24) = 2.57$, $p < .10$) suggesting that neural representations for both living and social categories are not significantly different in the left hemisphere. A two-way analysis for right hemisphere activation yielded a main effect for coordinates ($F(2, 24) = 137.6$, $p < .001$), again indicating a significant difference between all three dimensions. No main effect was observed for category ($F(1, 24) = .15$, $p = .71$), but a significant interaction was observed ($F(2, 24) = 3.44$, $p < .05$), which seems to be driven by the Y coordinate differences due to the more anterior location of social foci. These results suggest that neural representations of social and living categories in the right hemisphere are not significantly different.

Discussion

This meta-analysis suggests a significant neural overlap between areas that show activation for "socialness" and activation for living things. Social concepts are traditionally hard to define as they represent abstract semantic knowledge that enables us to describe our own as well as others' behaviors (Zahn et al., 2007). Neural regions representing "socialness" have most often been reported as cortical responses to social vignettes (e.g., Martin & Weisberg, 2003; Ross & Olson, 2010). These vignettes are based on the work of Heider and Simmel (1944), who first developed a movie clip of simple shapes that seemed to be capable of causal motion. When observers viewed this clip, they consistently attributed personality traits and emotions to the shapes regardless of the instructions that they were given (see also Scholl & Tremoulet, 2000). Follow-up studies on the phenomenon have repeated these findings and have demonstrated that these observations of animacy are consistent across cultures (Scholl & Tremoulet, 2000). Research in developmental psychology (e.g., Csibra, Gergely, Biro, Koos, & Brockbank,, 1999; Dasser, Ulbaek, & Premack, 1989; Gergely, Nadasdy, Csibra, & Biro, 1995; Premack, 1990) has shown that perceptual animacy is a phenomenon directly related to the elements of intentionality, goal-directed behavior, and causality, all of which make up the domain of agency, as described by Carey (e.g., 1988, 1999). Additionally, recent research has suggested that motion kinematics, and not feature properties of objects, are responsible for perceptual animacy and understanding agency (Scholl & Tremoulet, 2000).

Based on the behavioral evidence from studies such as these and functional imaging (e.g., Chao et al., 1999, 2002; Martin & Weisberg, 2003; Noppeney et al. 2006; Ross & Olson, 2010), "socialness," as observed through animacy and causal motion, seems to be an important factor in assessing the life state of an object. Castelli, Happé, Frith, and Frith (2000) reported findings in which goal-directed, and intentional movement, compared to random motion, elicited strong responses in fusiform gyrus and were accompanied by anthropomorphic language descriptions of stimuli. These findings, along with those discussed previously, indicate a strong relationship between areas linked to conceptual knowledge about living things and things that exhibit "socialness."

The Domain Specific Hypothesis can support animacy as an important factor for driving the organization of conceptual knowledge for living things, as well. Given that the DSH implicates evolutionary pressures as constraints on the development of concept categories (Caramazza & Shelton, 1998), animacy detection and social relevance are factors that would have been very important from an evolutionary standpoint. Therefore, understanding and parsing the environment into living and non-living things based on "socialness" would have been an effective means of survival and would have been promoted as a domain-specific mechanism. Plants, although living things, do not exhibit animacy or any other properties of the domain "social," and would therefore have been grouped as non-social objects, sharing properties similar to those of non-living objects.

CONCLUSION

Theories of conceptual knowledge organization agree on the dissociation between living and non-living things and are supported by both functional imaging and behavioral evidence. However, it is unclear whether the domain "living" is a coherent category representation. Results from this study show a dissociation between animals and plants in lateral, "living" regions of the ventral stream and a differential impact of animacy (biological motion) on the categorization of animals. If living things were a unitary concept category, then these findings would not be expected. On the other hand, intentional movement, causality, animacy, and agency seem to be important contributing factors in determining whether an object is alive (e.g., Piaget, 1929/1960; Carey, 1998, 1999, 2009; Scholl & Tremoulet, 2000; Martin & Weisberg, 2003; Ross & Olson, 2010). For these reasons, it may be the case that "socialness," a domain that represents elements of agency, animacy, causality and intentionality, better explains the organization of living and non-living things. The domain "socialness" fits well within the model of the Domain Specific Hypothesis as a cognitive mechanism that has been shaped by pressures of evolution. However, future studies are still required to understand the finer distinctions both between and within domains of "living" and "non-living." Specifically, studying an object category such plants, which is representative of both living and non-social things, might provide a more comprehensive understanding of the organization of conceptual knowledge.

REFERENCES

Anggoro, F.K., Waxman, S.R., & Medin, D. L. (2008). Naming practices and the acquisition of key biological concepts: Evidence from English and Indonesia. *Psychological Science, 19*, 314-319.

Berlin, E. (1992). *Ethnobiological classification.* Princeton, NJ: Princeton University.

Berteleson, P. (1967). The refractory period of choice reactions with regular and irregular interstimuli intervals. *Acta Psychologica, 27*, 45-56.

Borgo F., & Shallice, T. (2001). When living things and other "sensory-quality" categories behave in the same fashion: a novel category-specific effect. *Neurocase, 7*, 201-20.

Borgo F, Shallice T. (2003). Category specificity and feature knowledge: Evidence from new sensory-quality categories. *Cognitive Neuropsychology, 20*, 327-53.

Bloom, P., & Keil, F.C. (2001). Thinking through language. *Mind and Language, 16*, 351-367.

Broadbant, D.E., & Gregory, M. (1967). Psychological refractory period and the length of time required to make a decision. *Proceedings of the Royal Society Series B, 168*, 181-193.

Castelli, F., Happé, F. G., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12*, 314-325.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10*, 1-34.

Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language, 3*, 167-181.

Carey, S. (1999). Sources of conceptual change. *Conceptual development: Piaget's legacy* (pp. 293-326). Mahwah, NJ: Erlbaum.

Carey, S. (2009). *Origin of Concepts*. New York: Oxford University Press.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169-200). New York: Cambridge University Press.

Chao, L.L., Haxby, J.V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2*, 913-919.

Chao, L.L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex, 12*, 545-551.

Chang, D., & Troje, N. F. (2008). Perception of animacy and direction from local biological motion signals. *Journal of Vision, 8*, 1-10.

Csibra, G., Gergely, G., Biro, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of "pure reason" in infancy. *Cognition, 72*, 237-267.

Dasser, V., Ulbaek, I., & Premack, D. (1989). The perception of intention. *Science, 243*, 365-367.

Downing, P.E., Chan, A.W.-Y., Peelen, M.V., Dodds., C.M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 10*, 1453-1461.

Farah, M.J., & Rabinowitz C. (2003). Genetic and environmental influences on the organization of semantic memory in the brain: Is "living things" an innate category? *Cognitive Neuropsychology, 20*, 401-408.

Fodor, J.A. (1983). *The modularity of mind: An essay in faculty psychology*. Cambridge: MIT Press.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and animate-inanimate distinction examples. *Cognitive Science, 14*, 79-106.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165-193.

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science, 20*, 480-487.

Gottsdankcer, R., & Way, T.C. (1966). Varied and constant intersignal intervals in psychological refractoriness. *Journal of Experimental Psychology, 72*, 792-804.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243–249

Kadon, R.C., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology, 84*, 132-147.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88*, 197-227.

Mahon, B.Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron, 63*, 397-405.

Mahon, B.Z., & Carmazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris, 102*, 59-70.

Mahon, B.Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review Psychology, 60*, 27-51.

Mahon, B.Z., Milleville, S.C., Negri, G.A.L., Rumiati, R.I., & Caramazza, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron, 55*, 507-520.

Mandler, J.M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development, 8*, 291-318.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review Psychology, 58*, 25-45.

Martin, A., & Caramazza, A. (2003). Neuropsychological and neuroimaging perspectives on conceptual knowledge: An introduction. *Cognitive Neuropsychology, 20*, 195-212.

Martin, A., & Chao, L.L. (2001). Semantic memory and the brain: Structure and process. *Current Opinion in Neurobiology, 11*, 194-201.

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology, 20*, 575-587.

Martin, A., Wiggs, C.L., Ungerleider, L.G., & Haxby, J.V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*, 649-652.

Moore, C.J., & Price, C.J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain, 122*, 943-962.

Noppeney, U., Price, C.J., Penny, W.D., & Friston, K.J. (2006). Two distinct mechanisms for category-specific responses. *Cerebral Cortex, 16*, 437-445.

Opfer, J.E., & Siegler, R.S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology, 49*, 301-332.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220-244.

Pashler, H., & Christian, C. (1994). Bottlenecks in planning and producing vocal, manual and foot responses. *Center for Human Information Processing: Technical Report.*

Pashler, H., & Johnston, J.C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *Quarterly Journal of Experimental Psychology, 41A,* 19-45.

Piaget, J. (1929/1960). *The child's conception of the world.* Paterson, NJ: Littlefield, Adams & Co.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36,* 1-16.

Ross, L.A., & Olson, I.R. (2010). Social cognition and the anterior temporal lobes. *Neuroimage, 49,* 3452-3462.

Saxe, R., Tzelnic, T., & Carey, S. (2006). Five-month-old infants know humans are solid, like inanimate objects. *Cognition, 101,* B1-B8.

Scholl, B.J., & Tremoulet, P.D. (2000). Perceptual causality and animacy. *Trends in Cognitive Science, 4,* 299-309.

Spelke, E.S., & Kinzler K.D. (2007). Core knowledge. *Developmental Science, 10,* 89-96.

Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. In W.G. Koster (Ed.), *Attention and performance II* (pp. 276-315). Amsterdam: North-Holland.

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Telford, C.W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14,* 1-36.

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics, 67,* 667-675.

Van Essen, D.C. (2005). A population-average, landmark-and-surface based (PALS) atlas of human cerebral cortex. *Neuroimage, 28,* 635-662.

Van Essen, D.C., Dickson, J., Harwell, J., Hanlon, D., Anderson, C.H., & Drury, H.A. (2001). An integrated software system for surface-based analyses of cerebral cortex. *Journal of American Medical Informatics Association, 8,* 443-459.

Van Essen, D.C., & Dierker, D. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron, 56,* 209-222.

Wagner, L., & Carey, S. (2005). 12-month-old infants represent probable endings of motion events. *Infancy, 7,* 73-83.

Warrington, E. K., & Crutch, S. J. (2003). The selective impairment of fruit and vegetable knowledge: A multiple processing channels account of fine-grain category specificity. *Cognitive Neuropsychology, 20,* 355-372.

Warrington, E. K., & McCarthy, R. (1987). Categories of knowledge: Further fractionations and an attempted integration. *Brain, 110,* 1273-1296.

Warrington, E.K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107,* 829-854.

Waxman, S., & Markow, D. (1995). Words as invitations to form categories: Evidence from 12-to-13-month-old infants. *Cognitive Psychology, 29,* 257-302.

Welford, A.T. (1967). Single-channel operation in the brain. *Acta Psychologica, 27,* 5-22.

Woodward, A.L., & Markman, E.M. (1998). Early word learning. W. Damon & Lerner, R.M. (Eds.), *Handbook of child psychology* (5th ed., pp. 371-420). New York: Wiley.

Xu, F. (1999). Object individuation and object identity in infancy: The role of spatio-temporal information, object property information, and language. *Acta Psychologica, 102,* 113-136.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30,* 111-153.

Zahn, R., Moll, J., Krueger, F., Huey, E.D., Garrido, G., & Grafman, J. (2007) Social concepts are represented in the superior anterior temporal cortex. *PNAS, 104*, 6430-6435.

Zhang, Z., & Troje, N. F. (2005). View-independent person identification from human gait. *Neurocomputing, 69*, 250-256.

The journal is set in Quadraat and uses Akzidenz Grotesk for display.

Papers submitted to the Yale Review of Undergraduate Research in Psychology are reviewed using a blind review procedure.

The lead author of each paper submitting to this journal must have conducted her or his research while an undergraduate (or equivalent) student at an accredited college or university anywhere in the world.

For more information, visit www.yale.edu/yrurp or email psych.review@yale.edu.