

UNDERGRADUATE JOURNAL OF PSYCHOLOGY

AT BERKELEY

Highlights of Insight and Excellence in Undergraduate Research

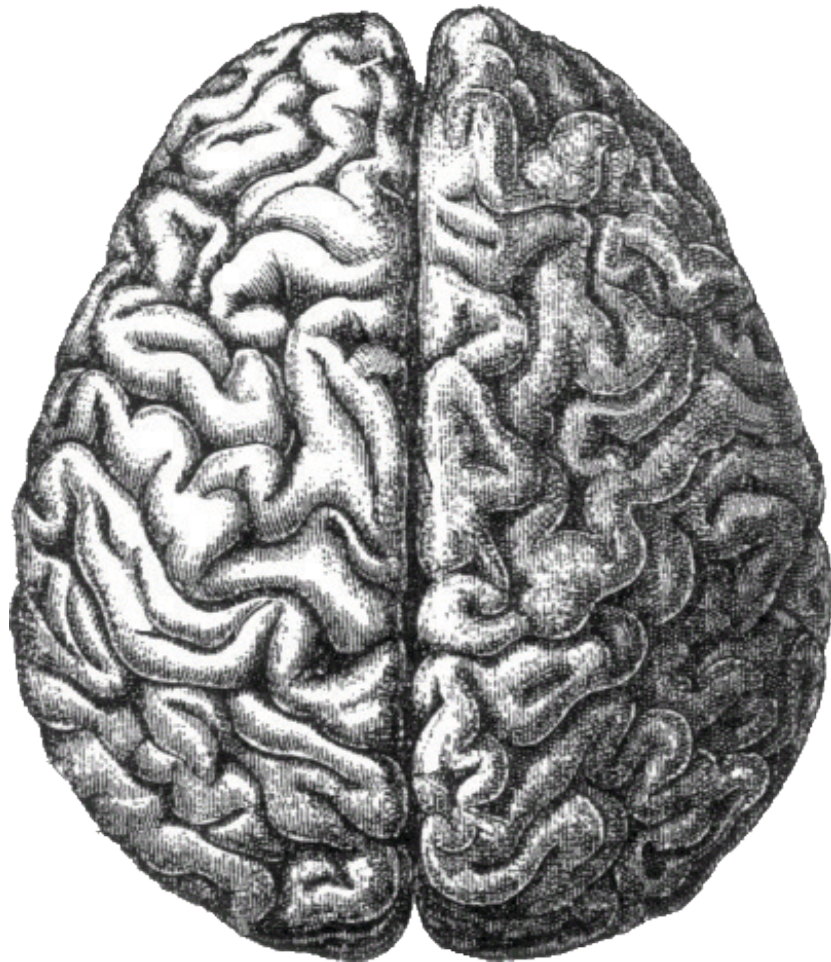


University of California,
Berkeley

Volume VII
2013 - 2014

UNDERGRADUATE JOURNAL OF PSYCHOLOGY AT BERKELEY

Highlights of Insight and Excellence in Undergraduate Research



University of California, Berkeley
Volume VII | Fall 2013 - Spring 2014

JOURNAL STAFF

EDITOR-IN-CHIEF

Katherine Wood

ASSOCIATE DIRECTOR

Aida Kohanteb

ASSOCIATE EDITORS

Anna Nguyen	Nada Rendradjaja
Armando Mota	Nicole Marsden
Dana Rosen	Sarah Alessi
Emma Farrow	Shanna Adler
Kalley Thompson	Sharon Chuang
Laura Belknap	Theresa Li
Louis Jacobus	William Pettus
Monica Breton	

LAYOUT DESIGN

Anna Nguyen

Special thanks to **Grayson Chao** for his work as web designer.

PSYCHOLOGY AT BERKELEY is an Associated Students of the University of California (ASUC) sponsored publication. The ideas and opinions contained in this publication solely reflect those of the authors and not the University of California

© 2013 – 2014 Undergraduate Journal of Psychology at Berkeley
<http://www.ujpb.org/>

TABLE OF CONTENTS

ii	Journal Staff
iv	Editor's Note Katherine Wood
v	Preface Richard Ivry

1	Jennifer Hatfield - Harvard University On the Horizon: Considering the Implications of Non-Western Cases of Anorexia Nervosa on Description and Diagnosis
9	Sydney Krueger - Princeton University Pre-Clinical Alzheimer's Disease: A Survey of the Observed Working Memory Deficits Among Non-Demented APOE e4 Carriers
18	Russell D. Martin - University of Minnesota Duluth Age Differences in Remembering Stereotypical Information by Targeting Implicit Stereotypes
27	Jessica E. Miner - Oklahoma State University Predictors of Toddler Behaviors from Infant Attention Measures
32	Shevantika Nanda - University of St. Andrews Are there Gender Differences in Empathy?
43	Thao B. Nguyen - University of California, San Diego Effects of Altering the Valence and Arousal of Music on Short Term Memory Task
51	Sasha Sommerfeldt - University of Minnesota Genetics of Behavioral Inhibition and Approach Systems: A Review of the Literature
59	Fushu Tan - University of Oregon Bullying, Victim, and Aggressor: Past Experience versus Current Behavior
69	Weston Uribe, Carole L. Yue, Robert A. Bjork - University of California, Los Angeles Toshiya Miyatsu - Washington University in St. Louis When To Take Notes: During Lecture or During Podcast?

EDITOR'S NOTE

Welcome to the 7th edition of the Undergraduate Journal of Psychology at Berkeley. This edition showcases careful experiments and thoughtful reviews conducted by undergraduate researchers from coast to coast.

This publication is the product of the very best that undergraduates have to offer. From the quality of the research to the dedication of the editing team, this journal highlights the potential and professionalism of the undergraduates who contributed to it. I hope that this inspires, educates, and helps to foster a thriving community of bright minds working in one of the most fertile and fascinating areas of science.

I extend my thanks to the authors for the privilege of publishing their work, to the editing staff for their tireless efforts, and to our webmaster, whose work allows us to remain open source and available to all. Finally, I thank our readers for their invaluable support.

Now, onto the science.

KATHERINE WOOD

Editor-in-Chief

PREFACE

I am pleased to introduce the 2014 edition of the Undergraduate Journal of Psychology. The discipline of psychology spans the social and life sciences. The papers contained in this volume reflect this breadth, with contributions spanning the many subfields of psychology. Here you will find review papers that explore important problems faced by our society such as bullying, stereotypes, and anorexia. The papers cover both basic and translational research, and even offer practical advice about how best to take notes in order to maximize retention. This is an engaging and sophisticated body of work.

At the University of California, Berkeley, our faculty are honored to have the opportunity to teach, and collaborate, with a very talented population of undergraduates. Psychology remains a very popular major, frequently resulting in large classes at the lower and upper division. Nonetheless, as shown by the work presented here, undergraduate students are able to create an intimate learning experience through their research projects. They are able to not only engage in the intensive study of a problem that builds on their idiosyncratic interests, but, as important, gain skills in the scientific method. An important part of this skill set is translating laboratory observations into a written work, one that makes clear the question at hand and then presents the results and conclusions in a concise and engaging manner. The reporting process is what makes science a cumulative, community endeavor. Our editors were pleased to receive submissions for this issue from universities and colleges across the country. The articles you will find here have been selected as representative examples of this excellent body of work.

I also want to congratulate our student editors for assembling the journal. They have refined a different set of skills, helping shape the ideas and writings of other individuals to ensure that the papers are maximally impactful.

Congratulations to all of the participants who have put together this year's edition of the Undergraduate Journal of Psychology.

RICHARD IVRY

Professor & Chair

Department of Psychology

University of California, Berkeley

ON THE HORIZON: CONSIDERING THE IMPLICATIONS OF NON-WESTERN CASES OF ANOREXIA NERVOSA ON DESCRIPTION AND DIAGNOSIS

JENNIFER HATFIELD

Harvard University

I would like to thank Professor Joshua Buckholtz and Teaching Fellow Elizabeth Mundy, the former instructors of Psychology 18: Abnormal Psychology, for their guidance and support in the research and writing process.

Eating disorders have often been characterized as byproducts of a flawed Western societal structure that places inordinate pressure on women to conform to media and peer standards of beauty and thinness. Although the DSM-5 can be said to reflect a Western description of eating disorders and their subtypes, its model lacks explanatory power for cases that deviate from its narrow diagnostic criteria and fails to attribute blame to specific components of said culture. The source of eating disorders in both specific cases and historical cases are still unclear due to variation. I outline these flaws, review the historical data, and assess recent changes in the incidence and course of eating disorders to illustrate the urgency of revising common conceptions of eating disorders. Devising more accurate diagnoses of eating disorders and more effective long-term treatments requires examining alternative hypotheses and rejecting an exclusively Western model of description.

In the Western world, the discrepancy between the size of an average woman and the size of an average model is increasing. Rates of eating disorders are increasing in Europe, Australia, and the United States. Media content about dieting has also increased, and women have reported more dieting behavior and increased dissatisfaction with their size (Miller & Pumariega, 2001, p. 94). In the United States alone, roughly 11 million people are affected by eating disorders, and the mortality rate for anorexia can be as high as 20% (Miller & Pumariega, 2001, p. 101; Eating Disorders Coalition, n.d.). All of these concerns seemingly implicate modern Western media, culture, and values for the relatively high incidence of eating disorders in the Western world. However, this explanation is inconsistent with research showing that eating disorders have existed throughout history, and are becoming prevalent in non-Western populations. I contend that eating disorders are not just manifestations of

Western values, but disorders with much historical precedent and with a large cultural component to their expression. This distinction is critical because the way that we conceptualize eating disorders has broad implications for the efficacy of treatment and the applicability of diagnostic criteria. In this paper, I explain and challenge the traditional description of eating disorders before providing a historical overview of anorexia nervosa. I conclude by considering hypotheses regarding the prevalence of eating disorders in non-Western cultures that could allow clinicians to reexamine the etiology of these disorders and to diagnose and better treat cases that do not fit the traditional Western model.

Traditional Concepts of Eating Disorders

The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) states the criteria necessary to determine a diagnosis of anorexia nervosa. First, the individual his or her consumption severely enough to cause “significantly

low body weight[, ... which] is defined as a weight that is less than minimally normal” (American Psychiatric Association & American Psychiatric Association Task Force on DSM-5, 2013). People with anorexia also exhibit extreme fear of gaining weight and a distorted perception of their body shape and size (American Psychiatric Association & American Psychiatric Association Task Force on DSM-5, 2013). Although the DSM definition does not implicate specific groups of people as being at higher risk, society usually considers eating disorders to be diseases primarily or exclusively affecting Westernized females.

Anorexia and other eating disorders are often considered derivatives of a society with misplaced cultural values and expectations for women. Media coverage bolsters the typical image of people with anorexia being white, upper class females by featuring young, white female models, which places pressure on this demographic to conform to perceived norms and might explain the lower prevalence of eating disorders in men (Miller & Pumariega, 2001, p. 96). This conceptualization of eating disorders reflects a belief in cultural determinism, which posits that ideas about beauty are culturally determined (Furnham & Baguma, 1994, p. 81). Accordingly, an important variable affecting the prevalence of eating disorders in minorities may be their acculturation into white Western society (Silber, 1986, as cited in Miller & Pumariega, 2001, p. 98). The traditional hypothesis thus implicates Western culture, with its competing pressures to conform and to succeed, as both a necessary and sufficient cause for the maladaptive cognitive mechanisms underlying anorexia and other eating disorders.

Problems with an Exclusively Western Framework

I argue that culture is merely a factor to the development of psychological disorders. Western culture in particular contributes to the development of eating disorders, including the specific form or diagnosis of a disorder, the expression of distress, and a sufferer’s concept of abnormality, but is not a necessary cause. Therefore, reliance on a narrow, modern, Western concept of anorexia nervosa is misleading for diagnosis and treatment.

Problem 1: Research Findings

Western culture cannot be the sole determinant of eating disorders because there have been several non-Western cases, particularly in Asia (Miller & Pumariega, 2001, p. 101). Miller and Pumariega (2001) attribute the increased international prevalence of eating disorders to familial and cultural conflicts that produce dissonance within individuals during a crucial period of their development. However, these conflicts surely vary given the specific cultural and societal elements to which individuals are exposed. For example, a study by Hill and Bhatti (1995) that examined the eating behaviors of British girls found that, while both Asians and Caucasian nine-year-olds were highly cognizant of body shape and thinness, the Asian girls showed much more “dietary restraint” than the Caucasian girls (Miller & Pumariega, 2001, p. 100). Furthermore, dietary restraint was positively correlated with the adoption of traditional cultural values, indicating that cultural values influenced Asian girls’ eating behavior, not just exposure to mainstream ideas and pressures. In the United States, a study by Snow and Harris (1989) found “a substantial incidence” (p. 334) of disordered eating in Southwestern Hispanics and Pueblo Indians. Snow and Harris report that 11% of participants met DSM criteria for bulimia, and over half reported some disordered eating behavior. The participants were minorities of low socioeconomic status and living in rural communities, making them markedly different from the stereotypical white, upper class, Westernized female with anorexia.

Another conflict between the existing research and the Western model of anorexia appears when one examines cross-cultural views of body image. A study by Furnham and Baguma (1994) measured Ugandan and British subjects’ evaluations of body shapes ranging from extremely underweight to obese (p. 83). If the traditional Western model is correct, one would expect that British participants would evaluate the relatively underweight body shapes as more attractive than would Ugandan participants because Britain is more Westernized. However, the authors report that the

main differences were in the groups' evaluations of the healthiness of each figure, not the attractiveness. Specifically, the Ugandan participants rated the four heaviest body shapes as relatively healthy, whereas the British participants considered those shapes to be relatively unhealthy. This challenges the common assumption that Western societies pressure adolescent women to conform through stringent definitions of attractiveness and therefore predispose that population to eating disorders. If Western culture is to be implicated at all, then, it must have another specific mechanism for predisposing its population to eating disorders.

Problem 2: Lack of Specificity

There are two ways in which the Western conceptualization of anorexia is too vague: it fails to specify which aspects of Western culture predispose adolescents to eating disorders,² and it fails to explain how eating disorders can develop in non-Western cultures. Regarding the former, some scholars believe that Western culture is responsible for the emergence of fat phobia, which is consistent with the remarkable absence of this phobia in most Asian countries (Miller & Pumariega, 2001, p. 101). Although the DSM-5 has made significant improvements over the DSM-IV in allowing flexibility in diagnosing cases that do not feature fat phobia, neither the Western model nor the DSM-5 allow for cases that are not aesthetically motivated, such as Katzman and Lee's (1997) example of a wife who starved herself to avoid having children (p. 388). The Western conceptualization of anorexia also omits an explanation of why eating disorders develop in non-Western cultures. Emphasizing Western aspects of anorexia leads one to fixate on fat phobia, when this symptom may be more broadly, and likely more accurately, described as a phobia of losing control (Katzman & Lee, 1997, p. 390). In this view, the societal balance of power, rather than gender or the media per se, may be the critical factor. Thus, this paper argues that the 'fat phobia' component of the second criterion should be exchanged for the 'fear of losing control' that is listed as a common sub-feature, because the latter is more applicable across cultures. This explanation

could apply to Western cases that are atypical (i.e., focused on control rather than appearance), as well as to non-Western cases, and even to cases in cultures who believe that beauty is not culturally determined at all. Although the Western model fails to explain these concerns, the DSM diagnostic criteria nevertheless reflect Western values.

The Historical Expression of Anorexia Nervosa

I now turn from a general overview of problems with the Western model of eating disorders to a historical example of anorexia nervosa, which highlights many of the aforementioned problems with the model. This case study will demonstrate the contribution of culture to the expression of symptoms – however, it seems that specific cultural experience affects symptoms more than Western culture in particular. This distinction broadens the Western model to be consistent with both historical and contemporary evidence regarding the prevalence of and variation in eating disorders.

Historically, anorexia nervosa was typically motivated by religious beliefs rather than body image or societal concepts of attractiveness. The first recorded death from self-starvation occurred in the fourth century A.D. in Rome, driven in part by the teachings of Saint Jerome that emphasized personal discipline (Bemporad, 1997, p. 404). There were also recorded cases in the fifth and eighth centuries in which girls believed that they were demonically possessed and had to refuse food to return to a state of purity or goodness. Additional cases of anorexia followed with the introduction of Eastern asceticism, which held that bodily needs are evil and must be resisted to achieve purity (Bemporad, 1996, p. 220). This cultural influence dominated over the previously esteemed Greek concept of “a sound mind in a sound body,” providing an early example of the power of acculturation in affecting the prevalence of eating disorders (Bemporad, 1997, p. 404). Some well-known figures, such as Vardhamana (the founder of Jainism), Buddha, and the goddess Uma, engaged in self-starvation for religious motivations (Bemporad, 1996, p. 218). This notion of 'holy anorexia' became especially

popular during the Renaissance, a period of relative affluence, and was often rewarded with sainthood (Bemporad, 1997, p. 407).

Another prevalent motivation for anorexia was a profound feeling of helplessness or a lack of control. For example, one explanation for the cases of anorexia that were reported in ancient Greece is that citizens were losing political control of a weakening city-state, prompting them to turn inward as a means of gaining control (Bemporad, 1997, p. 404). Another example involved Saint Wilgefortis, who wanted to be a nun and refused food to become less attractive and dissolve the plans for her arranged marriage. This example is antithetical to the Western model, as Saint Wilgefortis believed that becoming thinner would make her unattractive; furthermore, she only sought this aesthetic change as a means to an end – namely, gaining control over her future. As Bemporad (1997) remarks, “this behavior can be seen as a remedy for the physiological burdens associated with being female or as an assertion against the obedience expected of women” (p. 406). Furthermore, the phenomenon that eating disorders have historically been quite rare in poverty-stricken areas suggests that, when food is scarce, rejecting it does not provide an adequate feeling of control (Bemporad, 1996, p. 220). This fear of losing control is relevant to modern descriptions of anorexia, as it may apply to a larger proportion of current cases than the largely Western concept of fat phobia.

While there has been historical variation in individuals’ motivations underlying anorexia, there has also been a great deal of change in its definition. Originally considered a variant of hysteria, anorexia was not recognized as a psychological disorder until the mid-1800s (Bemporad, 1996, p. 225; Miller & Pumariega, 2001, p. 95). Some intellectuals in the 1700s were convinced that anorexia incorporated sitophobia, a dread of food stemming from either a fear of poisoning or from a divine command to abstain. A diagnosis of chlorosis was also common from the 1600s to the 1800s, which had very similar physical symptoms to modern-day anorexia, but was thought to be caused by a buildup of toxins in menstrual fluid (Bemporad, 1996, p. 226). In

the 1800s, secular motivations for eating disorders emerged, and the church began to interrogate people showing symptoms of anorexia, rather than rewarding them for adopting a lifestyle free of wants. Self-starvation was also seen as a novelty act. For instance, people paid money to see anorexic adolescents or professional hunger artists (men who resisted food as a show of strength) at circuses (Bemporad, 1996, p. 228; Bemporad, 1997, p. 409). A conflict developed between religion and medicine in explaining the ability of people with anorexia to subsist without food: religion proclaimed a divine role in the survival of individuals with anorexia, whereas doctors sought to explain it in biological terms for the first time (Bemporad, 1997, p. 409).

The first biological descriptions of anorexia emerged in the mid-1800s, with multiple modifications in the twentieth century. An early theory, designed during the era of hunger artists, suggested that the organs of people with anorexia created fuel by fermentation and eliminated the need for food; although patently false, the theory was an important step in the conception of anorexia as a disorder with biological rather than religious undertones (Bemporad, 1996, p. 224). In 1873, the first full medical description of anorexia nervosa was created, notably without the modern-day specification of fat phobia (Bemporad, 1997, p. 411; Miller & Pumariega, 2001, p. 95). This definition predated a clear change in body image in the Victorian period: well-fed women were no longer a symbol of wealth, and thinness was emphasized as an indicator of femininity (Bemporad, 1997, p. 410). Thus, the DSM’s emphasis on body weight and shape actually reflects relatively recent changes in anorexia nervosa. Weight phobia only became a criterion in 1970, and eating disorders were not subdivided into anorexic and bulimic types until 1979 (Russell, 1985, pp. 103-4). These changes in the description of anorexia nervosa reflect aspects of Western culture and society, but may not be universally applicable.

Recent Changes in the Expression and Description of Eating Disorders

By examining historical variations in the expression and prevalence of anorexia, I have challenged the assertion that anorexia is a modern

phenomenon of Western societies. Similarly, although the DSM-5 implicitly reflects a Westernized concept of anorexia, studying recent changes in eating disorders reveals how recently Western society has emerged as a contributory factor.

Incidence of Eating Disorders

In the twentieth century, and especially in the latter half, there has been a much higher incidence of eating disorders. This phenomenon is not exclusive to the United States: for example, Sweden experienced a fivefold increase in eating disorders between 1930 and 1950, which Russell (1985) attributes to increased societal pressures about body image (p. 102). Similarly, Hall, Cousins, and Power (1991) found that Mexican American daughters' ideal body shape was significantly thinner than that of their mothers, reflecting changes occurring even across one generation (as cited in Miller & Pumariega, 2001, p. 98).

An increasingly globalized, interconnected world exposes more people to the societal pressures common to many industrial countries, which partially explains the increased prevalence of eating disorders. Foreign exchange programs and immigration can also expose people to these pressures, especially adolescents and young adults, who are in a pivotal stage of development and often pursue these cross-cultural opportunities (Miller & Pumariega, 2001, p. 99). These experiences can then predispose an individual to an eating disorder that contains features of his or her native culture blended with this pressure to conform. For example, in the Middle East, vomiting rather than restriction tended to be a symptom of anorexia, likely reflecting cultural values or traditions that make the act of vomiting more feasible than restriction (Miller & Pumariega, 2001, p. 102). Furthermore, as irregular eating behavior becomes more prevalent, media or personal discussions of this behavior can contribute to its spread. Bruch posits that the increasingly high incidence of eating disorders can be attributed to copycat behavior, whereas as recently as the early twentieth century, people with anorexia had to come up with the idea of food refusal independently (as cited in

Bemporad, 1997, p. 413). Russell (1985) claims that these copycat cases are examples of “vulnerable individuals who, in earlier times, would have developed different forms of neurotic illness – possibly hysterical or anxiety states” (p. 106), which further implicates cultural pressures in determining the specific course of one's illness.

Weight Phobia in Anorexia Nervosa

Weight phobia as a necessary factor for diagnosis is a recent addition to the criteria for anorexia nervosa. This development may severely limit the diagnosis, detection, and treatment of non-Western or otherwise atypical cases. Weight phobia is common among Western people with anorexia, and it has scientific support: Russell (1985) demonstrates that people with anorexia tend to overestimate their own body width, reflecting the cognitive dysfunctions that spur weight loss even when they are severely underweight (p. 103). However, making weight phobia a necessary criterion for an anorexia nervosa diagnosis, as the DSM-IV did, ignores the role of culture in determining how one experiences distress and what is abnormal (Dolan, 1991, p. 76). As previously demonstrated, while weight phobia is common in Western cultures, this phenomenon may demonstrate as a fear of losing control in cultures with different body image ideals. This criterion predisposes healthcare professionals to adopt a Western model of the disease, which hampers their ability to recognize variations in symptoms across other populations. Given that the variability should increase as more cases emerge, this problem is even more important. Even though the DSM-5 has built in some flexibility surrounding this criterion, it is still not inclusive enough to diagnose the full range of variability properly because the criteria are still oriented toward appearance rather than power and control.

Recent developments in the description of eating disorders reflect the higher incidence of (and variation in) eating disorders in Western society, but do not fully account for corresponding increases in other societies. The DSM-5 has taken steps toward correcting some of these diagnostic issues; however, the criterion of weight phobia, as well as the overall emphasis on appearance, may still mask

even greater increases in non-Western populations than are currently recognized. Accuracy is crucial for the prospects of treatment of and prevention in atypical cases, and for understanding the etiology and symptoms of specific eating disorders.

Alternate Cultural Hypotheses for the Description and Diagnosis of Eating Disorders

In consideration of the historical development and evolution of anorexia nervosa, I now introduce two hypotheses regarding how culture contributes to the onset of eating disorders. I thus address one of the aforementioned problems with the Western model: it declines to assert what facets of Western culture are at fault. Is the problem the pressure to conform to certain body image standards, either from the media or from peer groups? Or is it certain expectations for female behavior more generally that make women feel out of control? This ambiguity precludes the Western model from being a truly useful hypothesis for description, diagnosis, or treatment. In contrast, the following hypotheses use the previous discussion of description and diagnosis to motivate and support their specific etiological claims.

Hypothesis 1: Clash of Cultures

The clash of cultures hypothesis has the potential to add a more global dimension to the Western model, and it also suggests why a large percentage of Westernized females fail to develop eating disorders. Industrialization is one way that cultures rapidly evolve, but it can induce dissonance among individuals reluctant to experience such change (Miller & Pumariega, 2001, p. 99). Individuals who change communities often experience identity confusion and isolation, and, if coupled with oppression in the new community, these women may be especially likely to adopt disordered eating as a coping mechanism (Katzman & Lee, 1997, p. 391). As Hare (1981) notes, “diseases, like species, represent the balance of a process by which living organisms struggle to adjust to a continually changing environment” (as cited in Russell, 1985, p. 101). Historical evidence

for this hypothesis includes the aforementioned effect of East Asian asceticism on Greek concepts of the body, as well as Greek feelings of helplessness in politics. Based on this hypothesis, lack of exposure to Western ideas may be a preventive factor for eating disorders, but it does not preclude their development; furthermore, Western culture itself may not be especially dangerous if it does not produce cognitive conflict.

Hypothesis 2: Feminist Theories of Eating Disorders

Like the clash of cultures hypothesis, feminist theories are advantageous in their specificity and applicability to non-Western cultures. Feminist theories posit that restricting food intake (or absorption, via purging behaviors) is a method of self-definition, control, and/or anxiety expression for women in cultures with expressed or implied body image norms (Miller & Pumariega, 2001, p. 104). Furthermore, the potential for control is amplified when women’s freedoms increase, which is when eating disorders are most common. Historical examples show that wealthy societies do not increase risk for an eating disorder unless they also have relatively broad women’s rights. Bemporad (1997) cites examples from certain ancient Greek and modern Muslim societies in which there exist affluent communities with severely restricted freedoms for women, and virtually no eating disorders have been reported in such communities (p. 414). Thus, the specific mechanism implicated in feminist theories is the presence of body image norms, when coupled with broad women’s rights and relative affluence.

Reconciling the Alternative Hypotheses

Both sets of alternative hypotheses suggest specific cultural factors that are implicated in eating disorders; in this sense, they are more powerful than the common Western hypothesis. They are also consistent with previously reviewed literature on the description and diagnosis of eating disorders. Both hypotheses support the conclusion that a social cause of eating disorders stems from women’s roles within a larger cultural framework,

which suggests more variability in eating disorders than if culture itself was the primary determinant. Both hypotheses also explain why only certain people in a culture develop eating disorders. To be most effective, prevention and treatment initiatives should consider specific individual variables stressed in one or both theories and, as Katzman and Lee (1997) suggest, look comprehensively at patients as individuals rather than focusing solely on symptoms of anorexia (p. 391). It is also crucial to recognize that these new hypotheses may more accurately indicate risk for certain subtypes of eating disorders than for others. This would be particularly likely if the DSM-5 were revised in line with the changes I suggested earlier, motivated both by historical analysis and contemporary research.

Conclusions

In short, anorexia nervosa has strong historical and cultural components of its expression and conceptualization; it is not a recent outgrowth of a Western culture emphasizing female body image and thinness. A theory advocating Western society as a necessary cause of eating disorders cannot be reconciled with historical or modern evidence, and it is not particularly powerful without more specific risk factors for susceptibility to eating disorders. However, this theory is common among laypeople and has influenced the DSM's ill-advised use of weight phobia as a criterion for anorexia nervosa and bulimia nervosa. The conceptualization of anorexia nervosa and other disorders is crucial for proper diagnosis and treatment. Furthermore, it may be essential to properly define and categorize such disorders in order to produce accurate assessments of etiology. Future research should assess international rates of anorexia nervosa, especially in countries where little to no data is currently available, and reexamine the justifications for weight phobia and for the subdivisions of eating disorders.

References

American Psychiatric Association, & American Psychiatric Association Task Force on DSM-IV. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. American Psychiatric Publishing, Inc.

Bemporad, J. R. (1996). Self-starvation through the ages:

Reflections on the pre-history of anorexia nervosa. *International Journal of Eating Disorders*, 19(3), 217-237.

Bemporad, J. R. (1997). Cultural and historical aspects of eating disorders. *Theoretical Medicine and Bioethics*, 18(4), 401-420.

Crichton, P. (1996). Were the roman emperors Claudius and Vitellius bulimic? *International Journal of Eating Disorders*, 19(2), 203-207.

Dolan, B. (1991). Cross-cultural aspects of anorexia nervosa and bulimia: A review. *International Journal of Eating Disorders*, 15(1), 81-89.

Furnham, A., & Baguma, P. (1994). Cross-cultural differences in the evaluation of male and female body shapes. *International Journal of Eating Disorders*, 15(1), 81-89.

Katzman, M. A., & Lee, S. (1997). Beyond body image: The integration of feminist and transcultural theories in the understanding of self-starvation. *International Journal of Eating Disorders*, 22(4), 385-394.

Miller, M. N., & Pumariega, A. J. (2001). Culture and eating disorders: A historical and cross-cultural review. *Psychiatry: Interpersonal and Biological Processes*, 64(2), 93-110.

Parry-Jones, B., & Parry-Jones, W. L. (1991). Bulimia: An archival review of its history in psychosomatic medicine. *International Journal of Eating Disorders*, 10(2), 129-143.

Russell, G. F. (1985). The changing nature of anorexia nervosa: An introduction to the conference. *Journal of Psychiatric Research*, 23(2), 129-143.

Snow, J. T., & Harris, M. B. (1989). Disordered eating in southwestern Pueblo Indians and Hispanics. *Journal of Adolescence*, 12, 329-336.

Footnotes

¹ Note that the DSM-5 has made a few changes compared to the DSM-IV, the most notable being the changes to the second criterion and the changes to the amenorrhea feature. The second diagnostic criterion for a diagnosis of anorexia nervosa is "Intense fear of gaining weight or of becoming fat, or persistent behavior that interferes with weight gain, even though at a significantly low weight" (American Psychiatric Association & American Psychiatric Association Task Force on DSM-5, 2013). Previously, it was limited to the fear of gaining weight or becoming fat. The DSM-5 has also deemphasized the fact that females who are anorexic and who have reached pubescent maturity are usually amenorrheic, moving it into a discussion of potential additional features rather than listing it in the criteria.

² Put another way, why don't all Western adolescents develop eating disorders? What features of Western culture are especially risky, and to which adolescents in particular?

PRE-CLINICAL ALZHEIMER'S DISEASE: A SURVEY OF THE OBSERVED WORKING MEMORY DEFICITS AMONG NON-DEMENTED APOE ε4 CARRIERS

SYDNEY KRUEGER

Princeton University

Alzheimer's disease (AD) is a progressive neurodegenerative disease that affects people of all ages but is most prevalent in the aging population, a population that is growing as the average lifespan increases. Patients with AD suffer extreme memory loss and eventually die. The disease causes severe cognitive decline, including a gradual decline in working memory (the ability to store, manipulate and recall information in a short time period). Treatment for AD is very limited in its effectiveness and the disease itself is irreversible. There is, however, a known genotype, apolipoprotein epsilon 4 (APOε4), which puts people at a higher risk for developing the disease in old age. It is also known that APOε4 carriers, before the onset of AD, perform worse on certain working memory tasks than education and age-matched non-carriers do. Working memory deficits have been observed on some tasks and not on others. Exploring the types of working memory tasks that differentiate the APOε4 carriers from the non-carriers under different conditions helps to further understand the cognitive changes that occur in pre-clinical AD and the link between the genotype and the disease.

Alzheimer's disease is an irreversible neurodegenerative disease that mostly affects the elderly population. The average age of clinical diagnosis for AD is 65, and the average life span following diagnosis is 7 years (McKhann et al., 1984). Many elderly people suffer from dementia, which involves cognitive impairment and personality changes as a result of some form of disease or injury. AD is the most common cause of dementia among the elderly population. Estimates show that 5.1 million people in the United States are currently diagnosed with AD, over 1.5% of the United States population (McKhann et al., 1984). Approximately 26.6 million people have AD in the world today (World Population Prospects, 2006). With the baby boomer generation now in their 60s and the population at risk for AD growing, cases of AD are expected to nearly triple in the next forty years.

Late-onset AD is initially characterized by the inability to learn new skills, difficulty remembering how to navigate familiar places, difficulty multitasking, and mood swings (National Institute on Aging, 2012). As the disease progresses, patients experience more drastic memory loss and

eventually forget their own identity, family, and friends. As the body weakens, the patient ultimately becomes unable to perform vital functions and the patient dies as a result of the disease (McKhann et al., 1984).

The neurological pathologies and symptoms of normal aging and AD are very similar, so it is often difficult to distinguish between the two. Brain scans and evidence from autopsies have provided two hypotheses for the unique neural pathology of AD—neuritic plaques and neurofibrillary tangles. Researchers have discovered abnormally high amounts of beta-amyloid deposits on neurons in the brains of AD patients (Kang et al., 1987). These deposits form plaques throughout the brain. These plaques also happen to be common in aging, but they occur more frequently in AD patients than in non-AD patients of the same age (Kang et al., 1987). AD patients also have an abnormally high amount of tau protein in their brains, which causes neurofibrillary tangles to form in neurons (Perl, 2010). After time, the plaques halt new neuron growth and cause neuronal death (Lazarov & Marr, 2010). The progressive nature by which the plaques

and tangles cause neurons to rapidly malfunction and die offers an explanation for how AD patients progressively lose cognitive abilities like the ability to acquire new memories. Brain scans also indicate that as AD patients age and the disease progresses, they have less grey matter than age-matched controls without AD. This difference in mass indicates that more neurons have died out in the brains of AD patients (National Institute on Aging, 2012).

Transgenic models of animals that are genetically altered to develop neural plaques and tangles provide strong support for the two hypotheses. The animals mature with an abnormally high amount of beta-amyloid in the brain, causing plaques and tangles to form. When given simple cognitive tasks the transgenic animals perform worse than control wild type animals (Oddo, Caccamo, Kitazawa, & Tseng, 2003). These models provide strong evidence that the plaques and tangles found in AD patients are responsible for the observed cognitive deficits (Oddo et al., 2003).

The Apolipoprotein e4 Genotype and its Significance in AD

In 1993, Allen Roses discovered evidence that individuals who have the e4 allele on a gene responsible for encoding Apolipoprotein E (APOE) are more common among the AD population than the general population (Corder et al., 1993). Some studies indicate that up to 50% of people in the United States who suffer late-onset AD have the e4 allele (Raber, Huang, & Ashford, 2004). It should be noted that not all AD patients have the APOe4 genotype, and not all people with the APOe4 genotype develop AD. The likelihood of developing late-onset AD also increases with gene dosage. Homozygote APOe4 carriers, those with two alleles for the gene, are significantly more likely to develop AD than heterozygote carriers and non-carriers (Raber et al., 2004). There was evidence of a genetic link with AD prior to the APOe4 genotype discovery, when it was found that the average person's chances of developing AD increased with familial history (Raber et al., 2004). The discovery of the APOe4 genotype thus provided further evidence for a genetic determinant of AD.

APO is an Apolipoprotein involved in the catabolism of a specific group of lipoproteins. It is expressed throughout the body, however it is most heavily expressed in liver cells and cells in the central nervous system. An APO allele occurs in nearly 80% of the population, but the e4 allele, which is less common, is the only variation that has any known correlation with AD (Raber et al., 2004). The percentage of people with the APOe4 allele differs between different cultures and is generally higher for Caucasians than most other cultures. Estimates suggest that somewhere between 20% and 30% of the population have at least one allele for APOe4 (Raber et al., 2004). This isoform has a greater difficulty breaking up plaques and tangles than other isoforms, so APOe4 carriers are more likely to have more plaques and tangles than average age-matched non-carriers (Pericak-Vance, 1991). Researchers concluded that this genotype thus makes individuals more susceptible to the plaques and tangles that cause AD (Pericak-Vance, 1991).

Working Memory

Memory is divided into two broad categories: long-term memory and short-term memory. Long-term memory is where people store information about their past, such as faces, places, and experiences. Anything stored in long-term memory was at one point retained in a person's short-term memory, where recent information is stored. A part of short-term memory that is closely studied by psychologists is working memory (WM), where people retain, manipulate, and recall limited amounts of information over a short time period (Becker & Morris, 1999). WM includes the processing of both verbal and non-verbal information.

Psychologists Alan D. Baddeley and Graham Hitch proposed a model of WM in 1974 where a central executive is responsible for moderating the information in two slave systems, the phonological loop and the visuospatial sketchpad (1974). The phonological loop continually rehearses a list of information for later recall. The amount of information that a person can hold in a loop ranges between 5-9 objects, differing between numbers and

words (Miller, 1956). The visuospatial sketchpad is responsible for storing visual information, such as the color, shape, and location of objects. The central executive directs attention between the two systems and is responsible for blocking out irrelevant information (Baddeley & Hitch, 1974). There are various types of tasks used to measure WM capacity. Researchers may present participants with images or words on a screen and test them later on the accuracy of their memorization, while also taking into account reaction time. Different methods of measuring WM often lead to different results for the same participant, indicating that WM has several components that can be highlighted with these various tasks. Unlike age-matched controls, AD patients suffer a progressive decline in WM. The severity of the WM deficits differs across WM tasks and at different stages of the disease (Baddeley, Bressi, Della Sala, Logie, & Spinnler, 1991). The age of onset differs per person and the rate of decline also differs between patients, so measuring WM in AD patients is challenging and often very inconsistent (Baddeley et al., 1991).

APOe4 Carriers and the Primacy Effect

In 2002, Rosen, Bergeson, Putnam, Harwell and Sunderland published the first paper that discussed the relationship between the WM of healthy APOe4 carriers and non-carriers (2002). They worked with a sample of 21 APOe4 carriers and 21 non-carriers. They did not report how many people in the sample were heterozygote or homozygote carriers, however since homozygote carriers are more rare it is likely that a majority of the participants in the carrier group were heterozygote carriers (Raber et al., 2004; Rosen et al., 2002). The participants in both groups ranged in age from 50-79 years old. All participants had no prior history of any neurological, psychiatric, or learning disorders, nor had they been prescribed any medication for pain or depression at the time of the study. None of the participants met the criteria for mild cognitive impairment, a test for dementia. The two groups were matched on age and years of education, with average age being approximately 62 years old, and the average years of education approximately 17

years. The two groups had similar IQs, according to the WAIS-R full-scale IQ (M=123 for carriers and M=125 for non-carriers). The two groups also did not vary from one another on the WMS-r general memory index (Rosen et al., 2002).

All participants were tested on the operation span task. In the task the participants were told to answer a series of math operations by confirming if the stated expression was true or false. For example, they would be presented with “ $8 \times 2 = 16$ ” and they would need to respond with either “yes” or “no”. In between the math operations there would be short words that the participants were asked to remember for later recall. An experimenter monitored the task for every participant, so that the participant would not rehearse the words before recall (Rosen et al., 2002).

Rosen found that the math scores (percentage of equations answered correctly) and total span scores (total number of words recalled) did not differ significantly between the carrier and non-carrier groups. There was a significant difference ($p < .05$), however, between the carriers’ and the non-carriers’ primacy scores, a measure of how often the participant recalled the first word in the set. APOe4 carriers recalled the first word of the set on an average of 8.52 sets out of 15, compared to 10.43 for non-carriers (Rosen et al., 2002). The researchers then looked at the accuracy with which each group recalled the first word in the set, regardless of whether they remembered it as being the first word or not. These results were similar to the prior primacy results, implying that the WM deficits among APOe4 carriers were a result of poor word recall (Rosen et al., 2002). The operation span task forces participants to divide their attention between word recall and solving math operations. People with clinical AD show significant declines in divided attention, a central executive function (Baddeley et al., 1991). The results of the operation span task indicate that APOe4 carriers may experience similar types of WM decline as those seen among clinical AD patients (Baddeley et al., 1991).

Healthy APOe4 Carriers and Visuospatial Attention Tasks

In 2005, Greenwood, Lambert, Sunderland, and Parasuraman tested the second system of WM, the visuospatial sketchpad, within a sample of 113 non-carrier and 64 APOe4 carrier participants (12 of the APOe4 carriers were homozygous and 52 were heterozygous) (2005). The participants ranged in age from 41 to 85 years old, averaging 58 years old. Participants were excluded if they had reported having significant medical problems affecting their neuropsychology. The researchers also excluded all participants who did not fall within the normal age-related scores on the Mattis Dementia Rating and Wechsler Memory scales, as well as the Buschke Selective Reminding task. The researchers matched control non-carriers with two groups of carriers on age, intelligence and mental health (Greenwood et al., 2005).

Greenwood, Lambert, Sunderland and Parasuraman presented their participants with a task that measured spatial WM, the ability to recall locations within an image (2005). The participants were told to fixate on an picture on the computer screen at which point an “X” would appear for three seconds somewhere on the picture. After a delay, a red dot appeared on the picture, either in the same position as the “X” or in a novel position. The participants were asked to assess if the red dot location matched the target “X” location (Greenwood et al., 2005). The participants were asked to recall the location of one, two, or three target locations at a time. Under all conditions, the participants were asked to assess whether the location of the red dot was old or novel after viewing all of the target locations. Increasing memory load (i.e. asking participants to recall more target locations) increases task difficulty (Greenwood et al., 2005).

The results of the spatial WM task show a significant difference in scores between groups on the three locations condition, and little variation under the one and two locations conditions (Greenwood et al., 2005). Homozygote carriers, on average, performed worse than age-matched non-carriers and heterozygote carriers when asked to recall three locations. Homozygote carriers have a much higher risk of developing AD later in life

than heterozygote carriers, so it was predicted that they would perform worse on cognitive tasks than heterozygote carriers (Greenwood et al., 2005). Patients with AD often discuss having difficulty remembering where they parked their car or how to navigate familiar areas, which both indicate a decline in visuospatial attention (National Institute on Aging, 2012). There is also evidence that AD patients perform worse than age-matched non-demented controls on certain visuospatial attention tasks, implying that the observed differences between the homozygote group and the other groups may be indicative of early signs of AD (Greenwood et al., 2005; Parasuraman, Greenwood, Haxby, & Grady, 1992).

Healthy APOe4 Carriers and the A-X Continuous Performance Task

Reinvang, Winjevoll, Rootwelt, and Espeseth ran a controlled study with APOe4 carriers using the A-X Continuous Performance Task (2010). The study included 186 participants who ranged in age from 40-80 years old and all participants were recruited through a newspaper advertisement in Oslo, Norway. All participants with a reported history of psychological disorders that affect cognitive abilities were excluded from the study. The experimenters also excluded all participants who scored more than one standard deviation away from the age-specific baseline average on the Wechsler Abbreviated Scale of Intelligence (Reinvang et al., 2010). The experimenters tested DNA from blood samples to assess genotype, and they observed that 69 of the 187 participants had at least one allele for APOe4 and 14 people were homozygote carriers (Reinvang et al., 2010).

During the task, participants observed a computer screen where red capital letters were continuously presented on a black background. The participants were told to respond with the rightmost key when presented with the target stimuli, the letter “A” followed by the letter “X”, and the leftmost key when presented with one of two non-target stimuli, the letter “A” followed by any letter but “X” (the AY condition), and any letter but “A” followed by “X” (the BX condition) (Reinvang

et al., 2010). They were given 1500 ms to press the right or left key. The participants first did a practice block with 20 trials and then performed six blocks of 50 trials (Reinvang et al., 2010).

The researchers found that the male homozygote carriers (eight of the 14 homozygote carrier participants) performed significantly worse than all other participants in identifying the AY non-target condition. Homozygote male carriers achieved 63% accuracy, while all other groups achieved between 80% and 90% accuracy (Reinvang et al., 2010). Poor performance on the AY condition of the A-X CPT is indicative of poor response preparation. Changes that result from normal aging have been associated with poor performance on the A-X CPT (Emery, Myerson, & Hale, 2007). Since the neurological process of AD is a similar but much faster version of normal aging, these results serve to reinforce that association (Reinvang et al., 2010).

Healthy APOe4 Carriers Without Deficits in WM

The three previously discussed papers were all criticized for their sample sizes. The Greenwood et al. and Reinvang et al. studies included a sample size of nearly 200 participants, with around 35% of them carrying at least one allele for the genotype, and the Rosen et al. study included even fewer participants ($n = 42$) (Greenwood et al., 2005; Reinvang et al., 2010; Rosen et al., 2002). In 2007, Jorm and colleagues conducted a study comparing the cognitive performances of healthy APOe4 carriers to non-carriers with a much larger subject pool (Jorm et al., 2007). The Australian study recruited 6,500 Caucasian participants that were divided into three different age groups: 20-24, 40-44, and 60-64, with around 2,000 participants in each group (Jorm et al. 2007). The researchers recorded previous head traumas and surveyed their participants' levels of alcohol consumption, but did not eliminate any participants from the experiments based on the alcohol and head trauma data. They eliminated participants afterwards who were suspected of having dementia if they fell within a threshold on a dementia test, or if they had

scores that were outliers (more than three standard deviations below the mean) on cognitive tests (Jorm et al., 2007). The experiment was conducted entirely on the computer and the participants were given the option of completing them at home or at the University with an interviewer (Jorm et al. 2007).

The experimenters used blood samples to test the genotype for all participants. 27.4% of the participants had at least one APOe4 allele, 2.3% had two alleles, and the results were consistent among age groups (Jorm et al., 2007). Under the category of WM test, the researchers tested all participants on a subset of the Wechsler Memory Scale, the backward digit span task (Wechsler, 1945). A series of digits appeared on a screen at one-second intervals and the participants were then prompted to recall the digits in backwards order (Jorm et al., 2007). The paper does not provide how many digits they were asked to recall.

The researchers observed no significant differences between heterozygote, homozygote, and non-carrier group scores on this task (Jorm et al., 2007). The researchers did, however, find differences between age groups, with the oldest group performing worse than the other age groups on all tasks, regardless of genotype. The researchers later looked at the effect of alcohol consumption and head trauma and found a correlation between the genotype and instances of head trauma and alcohol consumption. Using an ANOVA they saw that both head trauma and alcohol consumption had an effect on all participants' performance on cognitive tests, but the effect was not related to genotype. The tests were thus sensitive to age effects, head trauma, and alcohol consumption, but they were not sensitive to differences in people with the APOe4 genotype (Jorm et al., 2007).

Discussion and Further Directions

The Jorm et al. study presents an interesting counter-argument to the work of Rosen et al., Greenwood et al., and Reinvang et al.. Their results indicate that neither homozygote nor heterozygote APOe4 carriers exhibit abnormal cognitive decline prior to the onset of AD. Jorm et al. argue that based on the sample size, differences

in WM scores between carriers and non-carriers in the other papers could easily be attributed to noise. The results from the Jorm et al. study are the most convincing, because their sample size was considerably larger than the others' ($n = 6,500$). With a much larger sample size, all of the studies that observed WM differences between healthy APOe4 carriers and non-carriers could guarantee with more confidence that their results were less influenced by noise. It is possible that the observed deficits in WM capacity among APOe4 carriers were too subtle to appear on Jorm et al.'s study. In their paper, Jorm et al. argue that, if their tests were not sensitive to APOe4 genotype effects, they would not have picked up age-related differences or differences associated with head trauma or alcohol consumption. Since there is no relation between AD symptoms and performance on the backward digit span task, it is not surprising that APOe4 carriers perform similarly to non-carrier controls. The task in question most likely does not measure a process of WM affected by AD or any of the neurological processes associated with AD. While no differences were observed between the carrier and non-carrier groups on any of the other cognitive tasks used by Jorm et al., such as an immediate and delayed recall, they are too broad to be considered WM tests (Jorm et al., 2007).

Rosen et al. provided strong evidence for a decline in primacy scores on operation span due to carrying the APOe4 gene. These results cannot be directly compared to Jorm et al.'s results, because number recall is different than word recall. Additionally, Jorm et al. did not report primacy scores for any of the tests, so a future study with a sample size as large as the one in question should examine primacy scores on an operation span task. Also, it is unknown how many people in the carrier groups were later diagnosed with AD. This knowledge would distinguish the effects of the genotype from the effects of early symptoms of AD (Rosen et al., 2002).

Greenwood explored age and genotype effects on spatial memory, and observed that homozygote carriers had a greater difficulty recalling a high load of target positions than

heterozygote carriers and non-carriers. The Jorm et al. paper divided the heterozygotes from the homozygotes, but contrary to previous findings there were no observed differences between gene dosage on any of the cognitive tasks within each age group (Jorm et al., 2007). Reinvang et al.'s A-X CPT experiment also showed differences in gene dosage, but the A-X CPT measures a different resources of WM that cannot be easily related to any of Jorm et al.'s results, further limiting the relevance of Jorm et al.'s experiment.

The direct cause of AD remains unknown, but neuroscience research shows that plaques and tangles form in the brains of AD patients much more rapidly than in the adult aging brain. The discovery that healthy APOe4 carriers display deficits in WM prior to any symptoms of AD further emphasizes the link between the genotype and the disorder. As is the case with many diseases, understanding the cause can take us one step closer to finding a cure. Perhaps by using information learned about APOe4 carriers before the onset of AD, researchers can better understand the progression of the disease and potentially trace the decline in cognitive functioning before the onset of clinical AD.

Future studies should aim to replicate the previously discussed results using the WM tasks that have distinguished healthy APOe4 carriers from education and age-matched non-carriers. These experiments need to have a larger sample size, much closer to the Jorm et al. 2007 sample size ($n > 6,000$) (Jorm et al., 2007). Researchers should also consider combining WM tasks and other measures of fluid intelligence to create a composite score that serves as a latent variable of WM. Additionally, researchers should return to the data several years later and compare the performance of the APOe4 carriers who later develop AD to those who do not. The scores of the non-carriers who later develop AD should be compared to the non-carriers who do not develop AD. Being able to control for later development of AD will help distinguish whether the observed differences in WM are solely related to the genotype or are simply early symptoms of AD, too subtle to show up on tests that measure dementia.

References

- Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R., & Spinnler, H. (1991). The decline of working memory in Alzheimer's disease a longitudinal study. *Brain*, *114*(6), 2521-2542
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, *8*, 47-89.
- Becker, J. T., & Morris, R. G. (1999). Working memory(s). *Brain and Cognition*, *41*(1), 1-8
- Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., ... & Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, *261*(5123), 921-923.
- Emery, L., Myerson, J., & Hale, S. (2007). Age differences in item-manipulation span: the case of letter-number sequencing. *Psychology & Aging*, *22*(1), 75.
- Greenwood, P. M., Lambert, C., Sunderland, T., & Parasuraman, R. (2005). Effects of Apolipoprotein E genotype on spatial attention, working memory, and their interaction in healthy, middle-aged adults: results from the National Institute of Mental Health's BIOCARD study. *Neuropsychology*, *19*(2), 199.
- Jorm, A. F., Mather, K. A., Butterworth, P., Anstey, K. J., Christensen, H., & Easteal, S. (2007). APOE genotype and cognitive functioning in a large age-stratified population sample. *Neuropsychology*, *21*(1), 1.
- Kang, J., Lemaire, H. G., Unterbeck, A., Salbaum, J. M., Masters, C. L., Grzeschik, K. H., ... & Müller-Hill, B. (1987). The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature*, *325*(6106), 733-6.
- Lazarov, O., & Marr, R. A. (2010). Neurogenesis and Alzheimer's disease: at the crossroads. *Experimental neurology*, *223*(2), 267-281.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*(7), 939-939.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.
- National Institute on Aging. (2012). About Alzheimer's Disease. <http://www.nia.nih.gov/alzheimers/>
- Oddo, S., Caccamo, A., Kitazawa, M., Tseng, B. P., & LaFerla, F. M. (2003). Amyloid deposition precedes tangle formation in a triple transgenic model of Alzheimer's disease. *Neurobiology of aging*, *24*(8), 1063-1070.
- Parasuraman, R., Greenwood, P. M., Haxby, J. V., & Grady, C. L. (1992). Visuospatial attention in dementia of the Alzheimer type. *Brain*, *115*(3), 711-733.
- Pericak-Vance, M.A., Bebout, J.L., Gaskell, P.C. Jr., Yamaoka, L.H., Hung, W.Y. et al. (1991). Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet*, *48*, 1034-1050.
- Perl, D. P. (2010). Neuropathology of Alzheimer's disease. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, *77*(1), 32-42.
- Raber, J., Huang, Y., & Ashford, J. W. (2004). ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiology of Aging*, *25*(5), 641-650.
- Reinvang, I., Winjevoll, I. L., Rootwelt, H., & Espeseth, T. (2010). Working memory deficits in healthy APOE epsilon 4 carriers. *Neuropsychologia*, *48*(2), 566-573.
- Rosen, V. M., Bergeson, J. L., Putnam, K., Harwell, A., & Sunderland, T. (2002). Working memory and apolipoprotein E: what's the connection? *Neuropsychologia*, *40*(13), 2226-2233.
- Wechsler, D. (1945) Wechsler Memory Scale®--Third edition-(WMS---III). San Antonio, TX: Harcourt Assessment.

AGE DIFFERENCES IN REMEMBERING STEREOTYPICAL INFORMATION BY TARGETING IMPLICIT STEREOTYPES

RUSSELL D. MARTIN

University of Minnesota Duluth

I'd like to thank Dr. Aydin Durgunoglu (University of Minnesota – Duluth) for her expert advising on this research project. This paper was also possible thanks to early collaboration with Della Dustin (University of Minnesota – Duluth), who helped set the foundation for the hypotheses tested.

An experiment was conducted to investigate how false memories are created as a function of the stereotypical information in a text. Participants were 230 male and female US-residents recruited via Amazon Mechanical Turk and assessed using Qualtrics. In this web-based study, participants were given a news story about a homicide which either contained stereotypical or no stereotypical descriptions of the suspect. In each story, the ethnicity of the suspect was never explicitly stated. The participants had three minutes to read the story and then were taken to questionnaire pages that contained factual questions from the story along with the key question that asked for the suspect's ethnicity. Participants also provided confidence ratings for their responses. Results indicated that participants who read the story that included stereotypical descriptions believed the suspect was African-American, Asian, or Hispanic more often than the participants who read the non-stereotyped version. Additionally, age-group comparison analysis indicated that older participants falsely remembered ethnicity more often than younger participants in the story with stereotypical descriptions but not within the story without stereotypical descriptions.

Research indicates that racial bias has shifted from overt to subtle (Bristor, Lee & Hunt, 1995; Entman & Rojecki, 2000; Kern-Foxworth, 1994; Sweeney, 2005). The subtleness of racial bias can be measured by examining implicit stereotypes.

Undetected biases are considered to be indirect influences of today's stereotypical thoughts. Research has demonstrated that indirect influences alone can produce false remembering of information (Lenton, Blair & Hastie, 2001; Roediger, 1996). Additionally, our level of engagement with the media significantly contributes to shaping our stereotypical thoughts (Laney & Loftus, 2008).

Implicit cognition involves traces of a past experience that affect performance and decision making. Greenwald and Banaji (1995) define implicit stereotypes as introspectively unidentified traces of past experiences that mediate attributions of qualities to members of a social category. These experiences are not explicitly remembered as they are unavailable via self-report (Graf & Schacter,

1985). This idea of implicit memory is a well-established phenomenon with multiple methods of investigation. These investigation techniques include implicit association tests, go/no-go association tasks, and lexical decision tasks (among others). We examined an alternate approach by using stereotypical descriptors within an everyday context: a criminal news story.

Priming is an implicit memory effect in which exposure to a stimulus influences a response to a later stimulus. The effects of priming are very salient and long lasting (Tulving, Schacter & Start, 1982). Older individuals will typically show decrements on explicit tests (due to the biological nature of the aging process and the brain) relative to healthy young adults despite comparative priming levels (Fleischman, 2007). Furthermore, Ward, Berry, and Shanks (2013) outlined research findings that show explicit memory systems are affected by age, whereas implicit memory systems are not. However, priming in comparative age group

research is mixed. Theoretically, older individuals should be more susceptible to priming effects since topics will be more meaningful compared to younger individuals (Valsiner & Connolly, 2003). When priming is observed to have a similar effect across all age-groups in a particular study, it is considered to be age-invariant priming. This has been observed with different age-groups in word-stem completion tasks and word-identification tasks (Light & Singh, 1987; Park and Shaw, 1992). However, priming deficits were achieved during similar tasks in other studies (Chiarello & Hoyer, 1998; Hulstsch, Masson, & Small, 1991). Fleischman (2007) suggests that these discrepancies between findings are due to methodological differences.

The present research experimented with priming via stereotypical descriptions in an attempt to elicit individuals to falsely remember the ethnicity of a suspect. Also, we sought to find age differences to determine if a certain age-group is more susceptible to false remembering ethnicity compared to another. In accordance with the aforementioned research, we believe older individuals will have proportionally more incidents of false remembering ethnicity compared to younger generations in the stereotypical story but not within the non-stereotypical story. Additionally, we believe the stereotypical descriptors within our experiment will elicit more false remembering of African-American criminals compared to the story without stereotypical descriptors. The story that includes priming will attempt to indirectly coerce the participant to take implicit stereotypes into consideration when later asked about the suspect.

Methods

Participants

Two hundred thirty individuals (128 males and 102 females) above the age of 18 participated in the experiment. Participants were enlisted from Amazon Mechanical Turk and were paid fifty cents upon completion of the experiment. The experiment was advertised for domestic participants only (within the United States). Participants were divided into two groups, the experiment condition and the control group. The experiment condition

included one hundred fifteen individuals (65 males and 50 females). The control condition included the remaining one hundred fifteen (63 males and 52 females). Each person was randomly assigned to their treatment level.

Participants after Exclusions

After exclusions, one hundred seventy-two individuals were included in the statistical analysis. The experiment condition included eighty-six participants (44 males and 42 females). The control condition also included eighty-six participants (51 males and 35 females). Demographic information indicated a wide-range of residential states, education levels, and income. 77% of the participants were White or Caucasian, 8% were Asian, 7% were Black or African-American, 5% Hispanic or Latino, and 3% indicated other.

The first exclusion occurred if the participant read the story too fast (indicated by how long it took a participant to move to the next portion of the experiment). If a person spent less than 78.7s and 74.2s on the story portion of the control and experiment condition, respectively, or 1 SD below the average, their score was excluded (35 participants were excluded; SD = 44.3s, SD = 50.6s). The second exclusion occurred if the participant took more than 15 minutes to complete the entire experiment (four participants were excluded).

The third criteria addressed the potential for users to copy and paste the story in order to answer the questions in the later part of the experiment. To combat this, mouse clicks were measured during the story. If a person clicked more than 3 times on this section of the experiment, their data was excluded (10 participants were excluded). Lastly, the experiment required U.S. residents and each IP address was traced (nine participants were excluded). It is important to note that many violators of the exclusion criteria met two or more of the criteria outlined above. This overlap indicated they were indeed not engaged in the experiment properly.

Materials

The experiment was advertised on Mechanical Turk as a reading memory task that would take under 15 minutes to complete. If the task

was accepted, the person was forwarded to a survey created using the Qualtrics research survey platform. After completion, each participant was given a unique generated code which they entered back into the task completion screen on Amazon Mechanical Turk. This code was then submitted which indicated successful completion and the person was paid fifty cents after confirmation by the researcher.

The experiment was in the form of a closely controlled survey that was completely automated (hidden timer on each page, number of clicks were measured, IP address was saved). Participants were given a fictitious criminal news story involving the murder of a 29 year old woman from Los Angeles by a 25 year old suspect. Half of the participants (the control group) received a story with no stereotypical descriptors when referring to the suspect (see Appendix B). The remaining half of the participants (the experimental group) read the same criminal news story that included stereotypical descriptors (see Appendix C). After the participants read the story, they were taken to the last section of the experiment – a questionnaire that tested their knowledge about the story (see Appendix D).

Stereotypical Descriptors

A stereotypical descriptor is an adjective or a set of adjectives that motivate implicit stereotypes. The control group received neutral descriptions when referring to the suspect. For example, when discussing what a witness saw after the crime, the control condition story used a neutral descriptor, “witnesses say the man was wearing a sweatshirt and jeans.” In contrast, the experimental condition used a stereotypical descriptor, “witnesses say the man was wearing an over-sized hooded sweatshirt and baggy jeans.” The other six descriptors used were, “South Central Los Angeles,” “gang member’s house,” “gang violence,” “cocaine addiction,” and “cocaine,” which appeared twice. These stereotypical descriptors were used in the experimental group only (see Appendix C).

Procedure

The first page of the experimental survey was informed consent. Before agreeing to participate, it was clearly indicated that the

participant voluntarily agreed to participate, was at least 18 years old, and lived in the United States. Clicking “I agree” allowed the participant to see the demographic questionnaire pre-experiment (see Appendix A). Clicking “I disagree” sent the participant to the end of survey message (without the code they needed to receive payment in Amazon Mechanical Turk).

After answering the demographic questions (forced response on each) and after the participant clicked “next,” Qualtrics randomly (evenly) showed either the control, non-stereotypical story (see Appendix B), or the experimental story that included stereotypical descriptions (see Appendix C). The directions located at the top of each story were identical. Within each version of the story, there was a hidden timer that would automatically forward the participant to the questionnaire page. There was also a “next” button that would allow the participant to move on if they read the story before the automatic transfer. Mouse clicks and timing data were observed for the purposes of exclusions post-data collection. After the story was read, the participant was brought to the questionnaire page.

The final step of the experiment included 11 questions relating to the story presented in the previous portion of the survey (see Appendix D). Each question required a response and they would not be allowed to finish without answering each question and the corresponding confidence value for each. Nine questions were factual questions explicitly stated in the story (6 of which were multiple choice questions and 3 required fill-in answers). The remaining two questions were the questions of interest. The first asked whether or not a deadly weapon was present during the crime. Each version of the story had no mention of a weapon that was used. The second question asked for the ethnicity of the suspect, which was also never explicitly stated in either version of the story.

Results

Preliminary Analysis

The study included only 7 individuals who were over the age of 60. Age differences

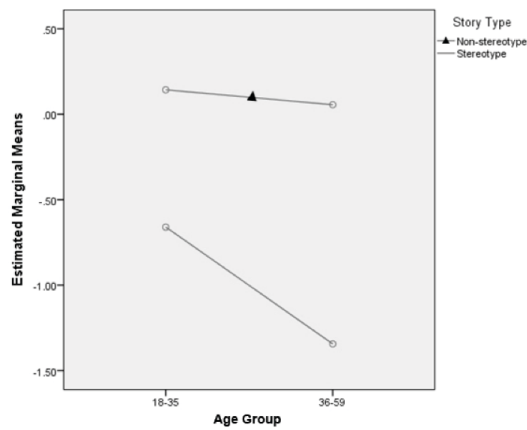


Figure 1. Estimated marginal means of ethnicity question. Note. The lower the estimated marginal mean, the more that group confidently remembered “African-American” as the criminal.

between this age group were not viable due to the low sample size. Consequently, these participants were not used for statistical analysis in age group comparisons. They were however used in the factual question and false remembering ethnicity analysis. Additionally, the presence of a deadly weapon question was never statistically analyzed due to concerns of how it was mentioned in the story. After exclusions, the average time it took participants to read the non-stereotype story was 133.8s (SD = 35.9s) and 142.3s (SD = 33.5s) to read the story with stereotypical descriptors. Raw-data coding information is available in the appendix section (see Appendix G).

Factual Questions

The nine factual questions included on the questionnaire were analyzed first. This is to verify the integrity of the story and questions and to make sure no significant difference occurred between both story types (manipulation check). An independent-samples t-test was conducted to compare the nine factual questions in the story without stereotypical descriptors and the story that did include stereotypical descriptors. There was no significant difference in the amount of correct answers on the non-stereotype story ($M = 1.98$, $SD = .70$) and the stereotype story ($M = 2.10$, $SD = .67$); $t(170) = 1.12$, $p = .266$. This enables us to assume that the questions adequately tested the knowledge of the participants in both stories equally.

False remembering a non-white suspect

A chi-square test was performed to examine the differences of story type in false remembering a non-white suspect on the ethnicity question of interest. The relationship between these variables was significant, $(1, N = 172) = 17.485$, $p < .001$. This result indicates that a significant difference occurred in false remembering a non-white ethnicity within the story that included stereotypical descriptions compared to the non-stereotype story. Therefore, the use of stereotypical descriptors played a role in eliciting more stereotype responses.

Age-group comparisons in false remembering non-white suspect

Age group comparisons were subjected to a univariate analysis of variance having two levels of story type (non-stereotypical, stereotypical) and two age groups (18-34, 35-59) (see Figure 1; Appendix E). No main effect was present for age groups, $F(1,172) = 2.11$, $p > .05$. A significant main effect was found for story type, $F(1,172) = 17.21$, $p < .001$. These data agrees with the chi-square analysis during the non-white suspect analysis. Participants had falsely remembered more non-white suspects in the stereotypical story across both age groups. No interaction was observed with age groups and story type, $F < 1.27$.

Tests of two a priori hypotheses were conducted using Bonferroni adjustments (see Table 1; Appendix F). Results indicated that the number of incidents of falsely remembering a non-white suspect was not statistically significant within the non-stereotype story of the 18-35 age-group ($M = .14$) and 36-59 age group ($M = .06$), $F(1,162) = .046$, $p = .830$. However, results indicate a significant difference in the number of incidents of falsely remembering a non-white suspect within the stereotypical story of the 18-35 age-group ($M = -.66$) and 36-59 age group ($M = -1.34$), $F(1,162) = 4.02$, $p = .047$. These data suggest that participants within the older age group falsely remembered more non-white suspects compared to the younger age group within the stereotypical story.

What is your age?	Story Type	Mean	Std. Deviation	N
18-35	Non-stereotype	.1429	1.55395	63
	Stereotype	-.6604	1.47991	53
	Total	-.2241	1.56642	116
35-59	Non-stereotype	.0566	1.69679	18
	Stereotype	-1.3438	1.42805	32
	Total	-.8400	1.65813	50
Total	Non-stereotype	.1235	1.57625	81
	Stereotype	-.9176	1.48974	85
	Total	-.4096	1.61460	166

Table 1. Age-group comparisons of ethnicity answers.

Note. The lower the mean, the more that group confidently remembered “African-American” as the criminal.

Discussion

The present findings suggest that stereotypical descriptors are an accurate measure of implicit stereotypes. Additionally, age group comparisons with the data show that older generations have more internal/implicit stereotypical motivations compared to younger generations. Moreover, older generations have higher confidence values when answering “African American” within the study. This could be explained by a number of factors. First, it is possible that the more experiences an individual has, the more matured and reinforced the implicit stereotypes are. Or secondly, our society could be becoming more ethnic neutral, with help from the new “subtle” nature of racial bias which exists within today’s media (Bristol, Lee & Hunt, 1995; Entman & Rojecki, 2000; Kern-Foxworth, 1994; Sweeney, 2005). The latter of the two is good news when it comes to ethnic equality. It appears that younger generations are perhaps more open-minded and don’t have internal biases as strong as older generations.

One must be wary of the impact that internet based studies have on the external validity of raw data, especially when the study offers cash incentives for participants. Consequently, the possibility of frivolous participants becomes a reality. Our study did not account for the possibility of copy/pasting the story using keyboard shortcuts (for reference during the questionnaire). To combat this, using a jpeg/gif instead of plain-text for the content of the story would make this concern trivial. Mechanical

Turk was chosen due to the increasing validity tests performed in recent studies. Shapiro, Chandler, and Mueller (2013) found that Mechanical Turk offers several advantages to research given their findings that participants are mostly genuine and adequately represent the population – as long as careful consideration and controls are implemented to prevent less than honest participants.

Other limitations of the present study include the use of specific stereotypical descriptors and the different internal perceptions of the confidence value verbiage. The stereotypical descriptors used in the present study were based off the opinion of the experimenter and the primary investigator’s discussion with colleagues. Because stereotypes are not ubiquitous, and are different from person to person, this is a clear limitation. Lastly, the definition of “not confident,” “somewhat confident,” and “very confident” could vary across participants. Carefully constructing this verbiage and clearly defining each in the directions of the questionnaire may have limited any differences that may exist.

The specific age groups the present study used were determined by historical events that may have explained age differences in internal biases. Specifically, the 36-59 group were late baby boomers and/or children raised during the after-effects of segregation – they may have seen and/or heard more stereotypical rhetoric compared to the younger age-group of 18-35. This may explain the difference in incorrectly identifying African-American as the criminal in the present study

(which was intended and which agrees with our initial hypothesis). Further studies may use different parameters but we believe the motif found in this study will be intact – older individuals will have more implicit racial-stereotypes than younger generations.

Conclusion

Future research would need to replicate these results. Many other factors could be looked at, including: educational differences, geographic differences within the United States, implicit stereotype differences across races, and the use of different stereotypical descriptors. Upon discovering that certain geographic locations and/or age groups are behind others, in terms of minimizing stereotypical thoughts, they could then be targeted by educators, media, and other relevant organizations.

Descriptors could easily target other ethnic stereotypes in an attempt to measure the strength of them. It would seem that measuring the implicit stereotypes within a person is an accurate measure of reviewing trends in the possible continued improvement of hidden biases. Although our data show an improving trend of open-mindedness, or perhaps an awareness of internal biases within individuals, continuing to monitor the implicit stereotypes within ourselves is an important step our society must undergo in order to make light of any unwanted or disadvantageous opinions.

References

Bristor, J. M., Lee, R. E. G., & Hunt, M. R. (1995). Race and ideology: African-American images in television advertising. *Journal of Public Policy & Marketing*, 14(1), 48–59.

Chiarello, C., & Hoyer, W. J. (1988). Adult age differences in implicit and explicit memory: Time course and encoding effects. *Psychology and Aging*, 3, 358–366.

Chiricos, T. & Eschholz, S. (2002). The racial and ethnic typification of crime and the criminal typification of race and ethnicity in local television news. *Journal of Research in Crime and Delinquency*, 39(400).

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1), 20-33.

Entman, R., & Rojecki, A. (2000). *The Black image in the White mind: Race and media in America*. Chicago, IL: University of Chicago Press.

Fleischman, D. A. (2007). Repetition priming in aging and alzheimer's disease: An integrative review and future

directions. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 43(7), 889-897.

Graf, P. & Schacter, D. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 501-518.

Greenwald, A. G., & Banaji, M. R. (1995). *Implicit social cognition: Attitudes, self-esteem, and stereotypes*. *Psychological Review*, 102(1), 4-27.

Hultsch, D. F., Masson, M. E. J., & Small, B. J. (1991). Adult age differences in direct and indirect tests of memory. *Journal of Gerontology: Psychological Sciences*, 46, 22–30.

Kern-Foxworth, M. (1994). Aunt Jemima, Uncle Ben, and Rastus: Blacks in advertising, yesterday, today, and tomorrow. Westport, CN: Greenwood Publishing Group, Inc.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100, 109-128

Laney, C. & Loftus, E. (2008). Emotional content of true and false memories. *Psychology Press*, 15(5).

Lenton, A., Blair, I. & Hastie, R. (2001). Illusions of gender: Stereotypes evoke false memories. *Journal of Experimental Social Psychology*, 37(3-14).

Light, L. L., & Singh, A. (1987). Implicit and explicit memory in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 531–541.

Likert, R. (1932). *A technique for the measurement of attitudes*. *Archives of Psychology*, 140, 1–55.

Park, D. C., & Shaw, R. J. (1992). Effect of environmental support on implicit and explicit memory in younger and older adults. *Psychology and Aging*, 7, 632–642.

Ramasubramanian, S. (2009). Pride, prejudice, and policy preferences: Exploring the relationships between TV stereotypes, racial attitudes, and support for affirmative action. Paper presented at the annual meeting of the International Communication Association, Marriott, Chicago, IL.

Roediger, H. (1996). Memory Illusions. *Journal of Memory and Language*. 35, 76-100.

Shapiro, D., Chandler, J., & Mueller, P. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science*. 1(2), 213-220.

Sweeney, C. (2005). 'One of them, but white': The disappearance of Negro: An anthology (1934). *Women: A Cultural Review*, 16(1), 93–107.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(4), 336-342.

Valsiner, J. & Connolly, K. (Eds.). (2003). *Handbook of Developmental Psychology*. London, SAGE Publications Ltd.

Walker, S., & Spohn, C. M. DeLone. 1996. *The Color of Justice: Race, Ethnicity and Crime in America*.

Ward, E. V., Berry, C. J., & Shanks, D. R. (2013). An effect of age on implicit memory that is not due to explicit contamination: Implications for single and multiple-systems theories. *Psychology and Aging*

Appendix A

**Referred to as Appendix B in original article.*



Please read the story below and then continue to the next section. You will automatically be transferred to the next part of the survey after 3 minutes.

A suspect was found and charged Wednesday with the murder of a 29-year-old woman in the Los Angeles area late last Saturday night. A 25-year-old male is accused of murdering the woman whose name has not been released. Police investigators believe the woman had gotten home from work Saturday evening and the perpetrator was waiting for her. The victim left work at 7 PM and the crime wasn't reported until later in the night. The perpetrator said he saw the woman with another man and wanted to confront her about it. The discussion escalated out of control. Prosecutors said he attempted to clean the crime scene and fled by car to a friend's house. He managed to hit a few parked cars at the end of the block. He was under the influence of drugs at the time of the arrest which may have contributed to his behavior. Neighbors said the victim and perpetrator knew each other for a while and dated in the past. They were unsure if they were dating at the time of the attack but they had seen the perpetrator around more often than usual. According to Los Angeles County district attorney's office, the suspect claimed the victim was his girlfriend and he had only been living with her for a few weeks prior to the murder.

News personnel were on scene to ask witnesses what they saw. Witnesses say the man was wearing a sweatshirt and jeans, and had his hands around his waist in a suspicious manner. One young, male witness said, "I saw a guy run through the park there and get into a car and speed off. He hit a whole bunch of cars down at the end of the block and that's when everyone came outside to see what was going on." One woman commented that, "This is all a tragedy. I don't know why stuff like this has to happen and it's a shame it happens so often."

Those who knew the perpetrator said his life had been plagued by violence since he was young. He had dropped out of school at 16 and had been in and out of jail his whole childhood. His mother had abandoned him when he was a baby after she succumbed to addiction and his father was never in his life.

The victim was desperately trying to make a better life for herself. She tirelessly worked two jobs to afford the monthly bills. She worked as a waitress during the day and by night she worked for a shipping company in the distribution center. Her co-workers said she always had a smile on her face, was always working hard and often had to be told her shift was over and needed to leave. Friends of the victim said she "had the warmest heart and put others before herself."

Los Angeles County prosecutors say the suspect has two prior convictions for assault with a deadly weapon. The suspect has been previously arrested for possession of drugs and drug paraphernalia, and possession of two unregistered handguns. With his current charge of murder he faces a minimum sentence of life in prison without the possibility of parole. The brutality of the case also makes him eligible for the death penalty.

next

Appendix B

**Referred to as Appendix C in original article.*



You will automatically be transferred to the next part of the survey after 3 minutes. Please read the following:

A suspect was found and charged Wednesday with the murder of a 29-year-old woman in the South Central Los Angeles housing projects late last Saturday night. A 25-year-old male is accused of murdering the woman whose name has not been released. Police investigators believe the woman had gotten home from work Saturday evening and the perpetrator was waiting for her. The victim left work at 7 PM and the crime wasn't reported until later in the night. The perpetrator said he saw the woman with another man and wanted to confront her about it. The discussion escalated out of control. Prosecutors said he attempted to clean the crime scene and fled by car to a gang member's house. He managed to hit a few parked cars at the end of the block. He was under the influence of cocaine at the time of the arrest which may have contributed to his behavior. Neighbors said the victim and perpetrator knew each other for a while and dated in the past. They were unsure if they were dating at the time of the attack but they had seen the perpetrator around more often than usual. According to the Los Angeles County district attorney's office, the suspect claimed the victim was his girlfriend and he had only been living with her for a few weeks prior to the murder.

News personnel were on scene to ask witnesses what they saw. Witnesses say the man was wearing an over-sized hooded sweatshirt and baggy jeans, and had his hands around his waist in a suspicious manner. One young, male witness said, "I saw a guy run through the park there and get into a car and speed off. He hit a whole bunch of cars down at the end of the block and that's when everyone came outside to see what was going on." One woman commented that, "This is all a tragedy. I don't know why stuff like this has to happen and it's a shame it happens so often."

Those who knew the perpetrator said his life had been plagued by gang violence since he was young. He had dropped out of school at 16 and had been in and out of jail his whole childhood. His mother had abandoned him when he was a baby after she succumbed to cocaine addiction and his father was never in his life.

The victim was desperately trying to make a better life for herself. She tirelessly worked two jobs to afford the monthly bills. She worked as a waitress during the day and by night she worked for a shipping company in the distribution center. Her co-workers said she always had a smile on her face, was always working hard and often had to be told her shift was over and needed to leave. Friends of the victim said she "had the warmest heart and put others before herself."

Los Angeles County prosecutors say the suspect has two prior convictions for assault with a deadly weapon. The suspect has been previously arrested for possession of cocaine and drug paraphernalia, and possession of two unregistered handguns. With his current charge of murder he faces a minimum sentence of life in prison without the possibility of parole. The brutality of the case also makes him eligible for the death penalty.

[next](#)

PREDICTORS OF TODDLER BEHAVIORS FROM INFANT ATTENTION MEASURES

JESSICA E. MINER

Oklahoma State University

The present study focuses on infant attention and the later predictions of Attention Deficit Hyperactivity Disorder (ADHD) in toddlers. ADHD can lead to many behavioral, social, and academic problems as these children age (American Psychiatric Association, 2013), causing a gap in their development. We hypothesize that the percentage of sustained attention (SA) and/or attention termination (AT) found in infant heart rate (HR) measures will be predictive of higher scores of inattentiveness on the Child-Behavioral Checklist (CBCL). Thirty-three infants/toddlers and their mothers participated in this longitudinal study. The data was collected over a four-year span, in which HR data from the infants and CBCL data from the mothers were collected. The results did not support our hypothesis of percentage time spent in SA and/or AT being predictive of higher scores of inattentiveness. However, there were significant results in percentage of time spent in AT and higher scores of aggressive behaviors in the toddlers. This finding supports past research that ADHD in younger populations is manifested as aggressive behaviors instead of inattentiveness, which is the manifestation in older populations.

Attention Deficit Hyperactivity Disorder (ADHD) has become an extensively studied childhood disorder, but its causes and earliest symptoms are still unknown (Alessandri, 1992; Kasper, Alderson, & Hudec, 2012; Kofler, Rapport, & Alderson, 2008). ADHD is a “persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with development, has symptoms presenting in two or more settings (e.g. at home, school, or work), and negatively impacts directly on social, academic or occupational functioning” (American Psychiatric Association, 2013). There is little research that looks at ADHD onset prior to school age, with most not looking until second grade (Deutscher & Fewell, 2001). The lack of research on ADHD prior to school age can be attributed to the fact that studies show that ADHD cannot be clinically diagnosed before school years (Alessandri, 1992; Rapport, Kofler, Alderson, & Raiker, 2007; Deutscher & Fewell, 2001). The authors also state that there could be a discrepancy between the diagnostic criteria (attention or aggression) when observing these two populations (toddlers and school age children) for ADHD. How symptoms of ADHD are expressed at

different ages could be attributed to environmental differences, (Deutscher & Fewell, 2001) such as being at school and needing to pay attention to the teacher to succeed compared to being at home or in a daycare where directed attention is not necessarily expected or required to succeed. Furthermore, there is evidence that in younger samples ADHD symptoms are more closely related to aggressive behaviors than strictly lack of attention (Campbell, 1990; Campbell & Cluss, 1982; Rubin & Clark, 1983; Schleifer et al., 1975).

Although literature has indicated that attention measures in infancy are associated with later intellectual functioning (Lawson & Ruff, 2004), surprisingly few attempts have been made to use these infant measures of attention to predict later attention problems in childhood. Studying these early signs of attention problems in infants and toddlers would hopefully lead to earlier detection of ADHD. Earlier detection is needed because ADHD has become one of the most prevalent neurobehavioral disorders affecting children. Studies suggest that 3% to 5% of school-aged children in the United States have significant educational problems because of ADHD (Deutscher & Fewell, 2001). Numerous studies

reveal many adverse outcomes associated with ADHD, ranging from scholastic underachievement and school failure to dysfunctional interpersonal and employment-related relationships (Rappport et al., 2007; Barkley, DuPaul, & McMurray, 1990; Goodyear & Hynd, 1992).

Findings by Richards and Casey (1992) on infant attention using HR defined phases led to our hypothesis that variations in infant attention may be related to attention deficits in childhood. Their data demonstrates that infants show a sustained parasympathetic response (heart-rate deceleration) while processing visual information. The sustained attention (SA) phase is defined as a deceleration in the heart rate. This phase represents the period that infants are engaged and encoding information. Richards and Casey found that visual information is encoded more efficiently during SA compared to the orienting (OR) and attention termination (AT) phases. Both the OR and AT phases are represented by an increase in heart rate. According to Richards and Casey (1992) AT involves reduced information processing and can be used as a measure of the infant's ability to disengage their attention. Further, Ruff and her colleagues (Ruff, Capozzoli, & Weissberg, 1998; Ruff, Lawson, Parinello, & Weissberg, 1990.; Ruff, Capozzoli, & Saltarelli, 1996) found that lower percentages of sustained attention at age 1-2 years correlated with inattention at age 3.5. This information leads us to believe that by comparing infant attention measures (SA and/or AT) with later toddler attention measures from the Child-Behavioral Checklist (CBCL), there may be evidence to identify specific infant attention measures that suggest later attention problems.

Methods

Participants

We recruited thirty-three mothers and their children from an original experimental infant group from study based on nutrition and development. All participants were recruited from the Stillwater, Oklahoma area through local breastfeeding groups, publicly posted flyers, and via word-of-mouth. The experimenters tested the infants at 3, 6, and 9 months of age and then again as toddlers. The toddlers were tested at a mean age of 3.2

years of age ($M=1174.64$ days, $SD=170.469$), and there was an even distribution of sex with a total of 15 males and 18 females. The majority of these mothers were white (86%) and well-educated (64% were college graduated). In addition, all infants had single, full-term, births without complications, and were healthy and primarily breastfed at 3 months. The Institutional Review Board of Oklahoma State University approved the study, and each mother provided informed consent for herself and for her infant/toddler.

Infant Measures

The present study used the visual information processing (Vishab) task which is an infant controlled habituation procedure to a single face stimulus (Colombo et al., 2009). The infant sat in a car seat centered 60 cm from a 43 cm X 27 cm computer screen. The experimenter dimmed the lights. The computer screen presented a randomly selected adult Caucasian expressionless human face (set against a white rectangular background of 18.5 X 14 cm) for each infant. The computer presented the face for as long as the infant was fixated on it, but removed the face when the infant looked away for one second or longer. The face was again presented after a two second interval during which the screen remained blank. The experimenter recorded the duration of each fixation in another room by observing the infant on a television screen through a video camera set above the infant's computer screen. Habituation criterion was two consecutive looks < 50% of the mean of the two longest looks. The computer controlling stimulus presentation also sent signals to a second computer that controlled acquisition of the electrocardiogram (EKG).

During the Vishab procedure, the experimenter measured electrocardiogram (EKG) for each infant with disposable Ag-AgCl electrodes placed in a triangular configuration on the infant's chest and abdomen. The computer digitized the EKG data at 250 Hz with a Biopac™ (BioPac Inc., Santa Barbara, CA) system that controlled data acquisition and synchronized it with the stimuli and the coding of the infant's fixations; the computer data file thus recorded these events along with the EKG.

Each infant's data file was parsed into the categories of baseline (BL), orienting (OR), sustained attention (SA), and attention termination (AT). Baseline was calculated for each fixation as the median heart rate in beats per minute for the two seconds prior to each fixation. Orienting was defined as two or more consecutive heartbeats above BL, sustained attention as five or more beats below baseline, and attention termination as two or more heartbeats above BL following SA. The durations of OR, SA, and AT were recorded as were the percentages of each of these heart-rate-defined attentional phase.

Toddler Measures

The present study used the Child-Behavioral Checklist (CBCL/ 1 ½ -5) to measure the attention of the toddlers (Achenbach & Rescorla, 2000). The CBCL is a parent-report measure of 99 child problem behaviors. It assesses the parents' descriptions of overall child problems and disabilities, concerns about the child, and opinions of the best things about the child. The experimenter scored the problem items using a three-point Likert-type scale (0 = Not True, 1 = Somewhat True or Sometimes True, 2 = Very True or Often True). Two broadband scores of externalizing and internalizing behaviors were obtained, and T-scores above the 98th percentile were considered clinically significant. The CBCL online data log calculated individual subscale scores (e.g., Withdrawn, Sleep Problems, Attention Problems, Aggressive Behavior), and a Total Problems score. We used the total scores from attention problems (0-10) and aggressive behavior (0-38) for the analysis (higher score signifying higher attention problems and/or aggressive behavior). Internal-consistency coefficients range from .66 to .92 for the subscale scores. Achenbach and Rescorla demonstrated the validity of this instrument by using it to distinguish between referred and non-referred children.

Predictor variables were the percentage of time infants spent in each of the three HR-defined attentional phases, namely OR, SA, and AT, at 3, 6, and 9 months of age. Outcome variables

were the CBCL scores for Attention Problems and Aggressive Behavior. We calculated Pearson correlation coefficients for each predictor variable-outcome variable pair.

Results

This study used data from when the infants were tested at 3, 6, and 9 months of age and again from when they were between two and a half and four years of age. The hypothesis was that the percentage of time spent in SA and/or AT would correlate with higher scores of attention problems on the CBCL. Our data did not support our hypothesis, since the results yielded no significant correlation between the infancy measures of attention and attention problems in toddlers. However, infant AT at 3 months of age came close to predicting attention problems in toddlers ($r(31) = .274$, $p = .068$). The switch from CBCL attention measures to aggressive behavior showed a significant predictor at 9 months of age in infant AT ($r(31) = .319$, $p = .035$). When we looked over the data, only 14 of our 33 participants had any time spent in AT. We ran a post-hoc correlation analysis using the 14 participants who had spent time in AT. The results showed a very significant predictor as expected at 9 months of age in infant AT for aggressive behavior in toddlers ($r(12) = .653$, $p = .006$). When looking at attention problems in the 14 subject sample there was an unexplained correlation at 9 months in infant OR phase ($r(12) = -.566$, $p = .018$). Lastly, we examined the number of looks during the visual habituation procedure at 3, 6, and 9 months of age and their possible predicting value in toddler attention. The results yielded a significant association between number of looks at 6 months of age and reported aggressive behavior in toddlers ($r(31) = -.393$, $p = .024$).

Discussion

This study suggests that when examining infant measures as possible predictors of ADHD in toddlers there may need to be a focus on aggressive behavior reported in toddlers in addition to that of the actual attention measures. This finding agrees with the findings from prior studies on

ADHD in samples of children younger than 5 years (Campbell, 1990; Campbell & Cluss, 1982; Rubin & Clark, 1983; Schleifer et al., 1975). A proportion of young children who show aggression at very high rates is also at an elevated risk for associated symptoms of ADHD (Hay, Hudson, & Liang, 2010). Because of this, Hay et al. (2010) hypothesized that the negative associations between prosocial behavior and externalizing problems could actually be accounted for by the symptoms of ADHD, not aggression. This suggests that the presence of aggressive behavior in children is actually a symptom of ADHD. Hay et al. (2010) also found that no association between aggression and pro social behavior existed when controlling for symptoms of ADHD. In particular, they found a negative correlation with the children's symptoms of inattention/over activity and the teachers' reports of prosocial behavior (Hay et al., 2010). Therefore, since low rates of prosocial behavior are associated with lower rates of peer acceptance (Denham, McKinley, Couchoud, & Holt, 1990; Hay et al., 2010), children with activity and attention problems may be at a higher risk for peer rejection (Hay et al., 2010). This peer rejection could be the manifestation of aggression that is seen in children with ADHD, especially in the younger age groups. As mentioned in the introduction, the significance of early detection is the ability to start intervention at the earliest sign of attention problems. By identifying at risk infants/ toddlers and introducing early interventions this could aid in the gap that develops during the early school years that can lead to social and academic problems.

Further, when we examined the number of looks and the toddler's aggressive behavior, it seemed that fewer looks predicted more aggressive behavior. We suggest that since these infants were more easily distracted by outside stimuli and lacked engaged attention at the presented stimulus, this could be the manifestation of distractibility. Distractibility is often seen as a cue symptom in ADHD. Something else to be explored is that roughly half of the sample had no AT at 9 months and the aggressive measures were only significant in the AT phase. This finding could be consistent

with the Richards and Casey article explaining the significance of AT. They describe it as being involved in reduced information processing and state it could be used as a measure of the infant's ability to disengage his/her attention. This could explain why AT at 9 months of age was predictive of aggressive behaviors in the toddlers. Since AT seems to measure the ability to disengage attention, and deficits in being able to regulate attention are cue symptoms in ADHD, more attention should be paid to AT measures and ADHD.

The results are to be interpreted cautiously because of the very small sample size. An increase in sample size would give this line of study a higher probability of recruiting children who will actually be clinically diagnosed with ADHD thus obtaining a better representation of the population of interest. In addition, we need to be able to test the toddlers again after 7 years of age since that is the suggested age to make a clinical diagnosis of ADHD. Nevertheless, this study suggests that early signs of ADHD may appear during the first 9 months of life. If other studies can replicate these findings then there could be evidence to support early intervention for these at risk infants and toddlers. Beginning intervention at the earliest signs of attention problems could be a major step in preventing the behavioral and social problems seen in many ADHD cases as they advance into the school years.

References

- Alessandri, S. M. (1992). Attention, play, and social behavior in ADHD preschoolers. *Journal of Abnormal Child Psychology*, 20(3), 289-302. doi:10.1007/BF00916693.
- Achenbach, T., & Rescorla, L. (2000). *Manual for the ASEBA Preschool Forms and Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual for Mental Disorders, 5th Edition (DSM5)*. Washington, D.C.: American Psychiatric Association.
- Aubuchon-Endsley, N. L., Grant, S. L., Berhanu, G., Thomas, D. G., Schrader, S. E., Eldridge, D., & Hambidge, M. (2011). Hemoglobin, growth, and attention of infants in southern Ethiopia. *Child Development*, 82(4), 1238-1251. doi:10.1111/j.1467-8624.2011.01596.x
- Barkley, R. A., DuPaul, G. J., & McMurray, M. B. (1990). Comprehensive evaluation of attention deficit disorder with and without hyperactivity as defined by research criteria.

- Journal of Consulting And Clinical Psychology*, 58(6), 775-789. doi:10.1037/0022-006X.58.6.775
- Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry*, 36(1), 113-149.
- Campbell, S. B., & Cluss, P. (1982). Peer relationships of young children with behavior problems. In *Peer relationships and social skills in childhood* (pp. 323-351). New York, NY: Springer.
- Colombo, J., Shaddy, D. J., Blaga, O. M., Anderson, C. J., Kannass, K. N., & Richman, W. A. (2009). Early attentional predictors of vocabulary in childhood. In *Infant pathways to language: Methods, models, and research directions* (pp. 143-167). New York, NY: Erlbaum.
- Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of preschool peer ratings. *Child Development*, 61(4), 1145-1152. doi:10.1111/j.1467-8624.1990.tb02848.x
- Deutscher, B., & Fewell, R. R. (2001). The development and use of the Attention Deficit Hyperactivity Disorder-Observational Rating Scale: Factor analysis and a preliminary investigation of predictive validity. *Journal of Psychoeducational Assessment*, 19(4), 317-333. doi:10.1177/073428290101900402
- Goodyear, P., & Hynd, G. W. (1992). Attention-deficit disorder with (ADD/H) and without (ADD/WO) hyperactivity: Behavioral and neuropsychological differentiation. *Journal of Clinical Child Psychology*, 21(3), 273-305. doi:10.1207/s15374424jccp2103_9
- Hay, D. F., Hudson, K., & Liang, W. (2010). Links between preschool children's prosocial skills and aggressive conduct problems: The contribution of ADHD symptoms. *Early Childhood Research Quarterly*, 25(4), 493-501. doi:10.1016/j.ecresq.2010.01.003
- Jansiewicz, E. M., Newschaffer, C. J., Denckla, M. B., & Mostofsky, S. H. (2004). Impaired habituation in children with attention deficit hyperactivity disorder. *Cognitive and Behavioral Neurology*, 17(1), 1-8. doi:10.1097/00146965-200403000-00001
- Kasper, L. J., Alderson, R. M., & Hudec, K. L. (2012). Moderators of working memory deficits in children with attention-deficit/hyperactivity disorder (ADHD): A meta-analytic review. *Clinical psychology review*, 32(7), 605-617.
- Kofler, M. J., Rapport, M. D., & Matt Alderson, R. (2008). Quantifying ADHD classroom inattentiveness, its moderators, and variability: a meta-analytic review. *Journal of Child Psychology and Psychiatry*, 49(1), 59-69.
- Lawson, K. R., & Ruff, H. A. (2004). Early focused attention predicts outcome for children born prematurely. *Journal of Developmental & Behavioral Pediatrics*, 25(6), 399-406.
- Rapport, M.D., Kofler, M., Alderson, R.M., & Raiker, J. (2007). Attention-Deficit/Hyperactivity Disorder. In M. Hersen & D. Reitman (Eds), *Handbook of psychological assessment case conceptualization and treatment, Vol. 2: Children and adolescents*. Hoboken, NJ: Wiley.
- Richards, J. E., & Casey, B. (1992). Development of sustained visual attention in the human infant. In B. A. Campbell, H. Hayne, & R. Richardson (Eds.), *Attention and information processing in infants and adults: Perspectives from human and animal research* (pp. 30-60). Hillsdale, NJ: Erlbaum.
- Rubin, K. H., & Clark, M. L. (1983). Preschool teacher's ratings of behavioral problems: Observational, sociometric, and socio-cognitive correlates. *Journal of Abnormal Psychology*, 11, 273-286.
- Ruff, H. A., Capozzoli, M., & Saltarelli, L. M. (1996). Focused visual attention and distractibility in 10-month-old infants. *Infant Behavior & Development*, 19(3), 281-293. doi:10.1016/S0163-6383(96)90029-6
- Ruff, H. A., Capozzoli, M., & Weissberg, R. (1998). Age, individuality, and context as factors in sustained visual attention during the preschool years. *Developmental Psychology*, 34(3), 454-464. doi:10.1037/0012-1649.34.3.454
- Ruff, H. A., Lawson, K. R., Parinello, R., & Weissberg, R. (1990). Long-term stability of individual differences in sustained attention in the early years. *Child Development*, 61(1), 60-75. doi:10.2307/1131047
- Schleifer, M., Weiss, G., Cohen, N., Elman, M., Cvejic, H., & Kruger, E. (1975). Hyperactivity in preschoolers and the effect of methylphenidate. *American Journal of Orthopsychiatry*, 45(1), 38.
- Sciberras, E., Efron, D., Schilpzand, E. J., Anderson, V., Jongeling, B., Hazell, P., ... Nicholson, J. M. (2013). The Children's Attention Project: a community-based longitudinal study of children with ADHD and non-ADHD controls. *BMC psychiatry*, 13(1), 18.

ARE THERE GENDER DIFFERENCES IN EMPATHY?

SHEVANTIKA NANDA

University of St. Andrews

Research has shown gender differences in a number of cognitive domains, such as spatial ability and math performance. Research on gender differences in empathy has yielded inconsistent results. Studies have suggested the observed gender differences in empathy may arise from males' reluctance to report empathy instead of a difference in ability. The present research investigated the effect that explicitly informing participants about the nature of an empathy task (empathy condition) or leading them to believe that the task evaluates social abilities (social abilities condition) has on performance on the Interpersonal Reactivity Index (IRI), a standard measure of empathy. Participants (20 males and 20 females) completed the IRI and the Marlowe-Crowne Social Desirability Scale form C, which measures the tendency to respond in a socially desirable way. The scores of females on the IRI were significantly higher than those of males in the empathy condition, and no significant gender difference was found in the social abilities condition. There was no significant difference between conditions for the social desirability score. Together, the results suggest gender differences on self-report measures of empathy do not arise from a difference in abilities.

Empathy is ability to identify what someone else is thinking or feeling and to respond to these thoughts and feelings with an appropriate emotion (Baron-Cohen, 2011). Empathy is thought to comprise two main components, cognitive empathy and emotional (or affective) empathy (Lawrence, Shaw, Baker, Baron-Cohen & David, 2004). Cognitive empathy is the ability to know what another person is thinking and is similar to "theory of mind," and emotional empathy is the ability to experience an emotion similar to that of another individual (Lawrence et al., 2004; Rueckert & Naybar, 2008). Empathy has an impact on social and emotional health across cultures (Cassels, Chan, Chung, & Birch, 2010). It is correlated with prosocial behavior and altruism (Carlo, Hausmann, Christiansen, & Randall, 2003) as well as inhibiting antisocial and aggressive behavior (LeSure-Lester, 2000; Baron-Cohen, 2011). Higher levels of empathy and emotional management are also associated with better relationships with peers (Eisenberg, Miller, Shell, McNalley, & Shae, 1991). Lower levels of empathy have been associated with disorders such as Asperger syndrome and autism (Baron-Cohen & Wheelright, 2004; Baron-Cohen, 2011). Thus,

empathy plays a vital role in an individual's life, as it allows the individual to interact effectively in social situations.

Empathy is thought to be a critical cognitive difference between men and women due to females scoring significantly higher on measures of empathy. The existing research, however, has revealed mixed results regarding evidence for gender differences. Emotional judgment tasks such as "Reading the Mind in the Eyes," which rely on the accurate judgment of emotion from observing the eyes, indicate female superiority (Baron-Cohen, Wheelright, Hill, Raste, & Plumb, 2001). Evaluation of nonverbal data using neuroimaging techniques such as functional magnetic resonance imaging (fMRI) provides less support for gender differences in empathy. In their meta-analysis, Wager et al. (2003) did not find a significant difference in brain activation between men and women in response to emotional stimuli. The use of physiological measures as indices of empathy results in men and women obtaining similar scores (Eisenberg & Lennon, 1983). Thus, in studies that use neuroscientific and physiological methods, gender differences in empathy do not appear. These findings strongly suggest that previously reported

evidence of gender differences in empathy might be influenced by the methods adopted in the studies.

The most convincing evidence for gender differences in empathy is provided by studies using self-report measures to assess empathy (Rueckert, 2011). Women score significantly higher than men on the Emotional Quotient (Baron-Cohen & Wheelright, 2004) and the Interpersonal Reactivity Index (Davis, 1980). Culture and socialization play an important role in the development of empathy (Baron-Cohen, 2005) and thus may explain the discrepancy in findings reporting gender differences in empathy. Eisenberg and Lennon (1983) suggested biases in self-report measures could influence the observed gender differences. The differences may be because men may be reluctant to report empathic experiences due to social expectations. When a measure is thought to assess empathy, it may prompt responses influenced by an individual's identification with gender stereotypes (Michalska et al., 2013). One of the most prevalent stereotypes in society is that women are more caring, people-oriented and empathetic than men (Rueckert, 2011). Therefore, it is possible that when a measure is thought to assess empathy, women feel they must respond more empathically, whereas men feel they must respond less empathically in order to conform to gender roles.

In studies using self-report measures of empathy, gender differences are found to increase between childhood and early adulthood (Mestre et al., 2009; Michalska et al., 2013). However, no significant gender difference in brain activation related to empathy has been found with increasing age (Eisenberg & Lennon, 1983). These findings lend more support to the role of culture and socialization in the observed gender differences in empathy. Previous research has suggested that knowledge of stereotypes in areas of personality traits and achievement increases in adolescence (Signorella, Bigler, & Liben, 1993 in Berk, 2006). Increased awareness of stereotypes may lead to the observed increase in gender differences on self-report measures of empathy. The role of gender stereotypes on other cognitive abilities has been evaluated. Dar-Nimrod and Heine (2006) found that telling participants that there are no

gender differences or that gender differences arise due to differences in experience could attenuate gender differences in math performance. Moe (2009) investigated whether positive beliefs and explanations can influence performance on the mental rotation task, at which males are thought to be superior. It was found that leading women to believe that females perform better than males improved their performance on the task. Thus, changing the information given to participants prior to the task can reduce gender differences in tasks that are stereotypically associated with one gender.

In their experiment, Klein and Hodges (2002) found women performed significantly better than men on a measure of empathic accuracy only when participants were led to believe that the measure was related to empathy prior to completing the task. But when participants were paid in exchange for accuracy, leading participants to believe that the measure was related to empathy resulted in no gender differences. The authors suggest the gender differences observed in the performance of empathic accuracy arise from differences in motivation to perform well. The findings from this study strongly suggest gender differences may be susceptible to test conditions and that the motivation to perform well and may not be due to a difference in ability between the genders.

The current study aimed to investigate the influence of written instructions on gender differences in empathy. Instructions explicitly indicating a measure of empathy and instructions indicating a measure of social abilities were presented to participants prior to completing the Interpersonal Reactivity Index (IRI) (Davis, 1980). To investigate the tendency to respond in a manner considered socially acceptable, a measure of social desirability, Marlowe-Crowne Social Desirability Scale Short Form C (MC-SDS form C) (Reynolds, 1982) was also used. Thus, if participants have a tendency to present themselves in a favorable (or unfavorable) light, their social desirability scores would reflect that. If gender differences in empathy are a result of differences in ability, changing the information about the nature of the task should have no impact on the empathy

and social desirability scores. Based on evidence illuminating the impact of gender stereotypes on performance on self-report measures of empathy, there should be a significant difference between the IRI scores of males and females in the condition in which participants are explicitly told their empathy levels are being evaluated (empathy condition). No gender differences in IRI scores are predicted when participants are told that their social abilities are being evaluated (social abilities condition). As a significant gender difference is predicted in the empathy condition, both males and females in this condition may be responding in a way that they believe is socially acceptable. Thus, for the MC-SDS form C, participants in the empathy condition should display higher social desirability scores than participants in the social desirability condition.

Methods

Participants

40 participants between 18 and 31 years of age ($M = 21.5$, $SD = 2.68$) were recruited through the University of St Andrews Research Participation System (SONA). There were 20 females and 20 males.

Materials

Participants completed the IRI (Davis, 1980), which contained 28 questions divided into four subscales (7 items each) that measure different aspects of empathy. Each question of the IRI including reversed scoring items (items 3, 7, 12, 13, 14, 15, 18, and 19) was assigned a score between 0 and 4. The total score for each subscale within the questionnaire and the overall scores for each participant were calculated. The Empathic Concern (EC) subscale measures an individual's tendency to experience concern for others. The Fantasy Scale (FS) subscale measures an individual's ability to imaginatively be involved with fictitious characters and situations. The Personal Distress (PD) subscale evaluates the tendency of an individual to become distressed as a result of witnessing another individual's distress. Finally, the Perspective Taking (PT) subscale evaluates an individual's ability to adopt the perspective of another person. Participants were required to

respond to each item in the questionnaire on a five-point scale ranging from 0 (does not describe me well) to 4 (describes me very well).

Participants also completed the MC-SDS form C (Reynolds, 1982). This measure evaluates the tendency of individuals to respond in ways that are considered desirable based on culture and social norms. The questionnaire contains 13 statements based on culturally acceptable behaviors that are relatively unlikely to occur (Crowne & Marlowe, 1960). The items are to be answered with either "true" or "false" depending on how well the statement describes the individual. Each item, including the reversed scored items (items 1, 2, 3, 4, 6, 8, 11, and 12), was assigned a score of either 0 or 1. The total score for each participant was calculated.

Priming

Participants were primed using the IRI. In the empathy condition, participants were explicitly told their empathy levels and reactions to emotional situations were being measured. In the social abilities condition, participants were told their social abilities and reactions to social situations were being measured. At the end of the instruction sheet, each participant was required to write his or her age and gender, thereby making gender salient before completing the questionnaire.

Procedure

Participants were randomly assigned to either the empathy condition or the social abilities condition using an assignment system that ensured an equal number of males and females were in each condition. They were then presented with a description of the IRI and instructions on how to complete the questionnaire. The experimenter returned to the room after approximately eight minutes and presented participants with a second set of instructions on how to complete the social desirability scale. The experimenter returned after approximately four minutes, and the participants were given a full debrief explaining the purpose of the study including the predictions of the study. They were then compensated with GBP 1.50 for their time.

Statistical Analysis

The data was checked for normality using SPSS. The Shapiro-Wilk normality tests revealed that only the data of the females in the empathy condition was normally distributed ($w = .790$, $p = 0.11$). To ensure accurate interpretation of parametric tests, the variances for each group of scores were calculated. This revealed that the analysis of variance is likely to be valid, as the largest variance was less than four times the smallest variance (Howell, 1992) (social abilities females = 147.21, social abilities males = 98.44, empathy females = 102.5 and empathy males = 246.27).

A repeated-measures ANOVA was conducted with gender (male/female) and condition (empathy/social abilities) as the between-subjects factors and the subscales (EC, FS, PD, PT) as the repeated measures factors. Where sphericity was violated, appropriate corrections were used. For further analysis of the results, a simple effects analysis was conducted.

The data obtained from the MC-SDS form C was checked for normality, and the Shapiro-Wilk test revealed that the data were drawn from a normal distribution ($w = .972$, $p > .05$). A two-way between-subjects ANOVA was conducted with gender and condition as between-subjects variables and the social desirability score as the dependent variable.

Results

Interpersonal Reactivity Index

Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(5) = 25.64$, $p < .001$). Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.73$). Tests of between-subjects effects revealed a significant effect of gender ($F(1, 36) = 13.19$, $p = .001$) and a significant interaction between gender and condition ($F(1, 36) = 6.97$, $p = .013$). The two-way repeated measures ANOVA revealed a significant main effect of subscale ($F(2.20, 79.29) = 15.89$, $p < .001$). Figure 1 shows the mean scores of participants on each of the subscales within the IRI. However, no significant interaction between subscale and gender ($F(2.20,$

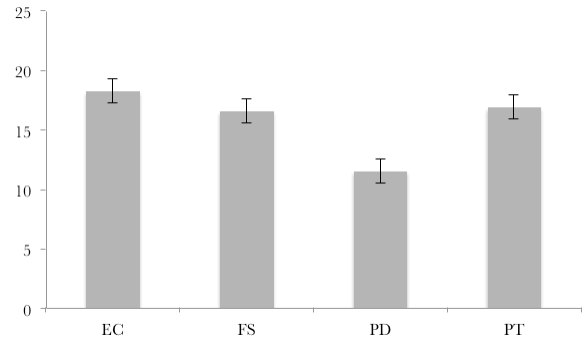


Figure 1: Graph showing the mean for each condition in the IRI with standard error bars

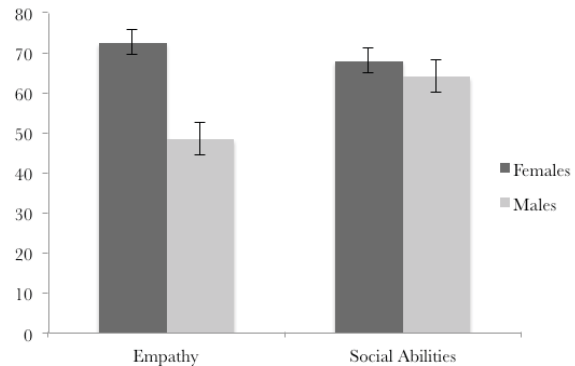


Figure 2: Graph showing the mean scores for each condition on the IRI with standard error bars.

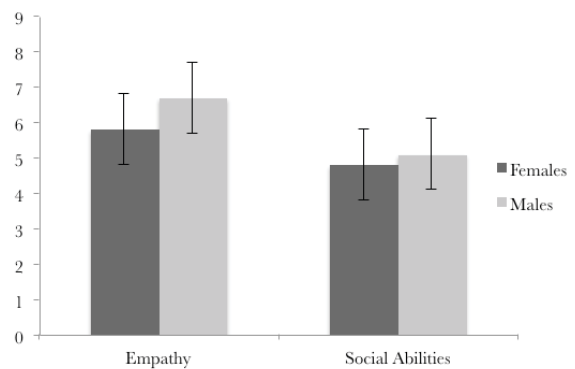


Figure 3: Graph showing the mean social desirability score in each condition on the MC-SDS form C with standard error bars.

79.29) = 1.19, $p > .05$) or subscale and condition ($F(2.20, 79.29) = 2.06, p > .05$) was found. Thus, gender and condition did not influence the scores on the four subscales.

A simple effect analysis revealed no significant difference in scores of males and females in the social abilities condition ($F(1,39) = 0.50, p > .05$). The analysis revealed a significant difference between the scores of males and females in the empathy condition ($F(1,39) = 19.01, p < .001$) with males scoring significantly lower than females. This indicates gender differences in empathy were only present in the empathy condition.

Social Desirability Scale

The two-way ANOVA revealed no main effect of gender ($F(1,36) = 0.39, p > .05$) or condition ($F(1,36) = 0.09, p > .05$). The interaction between gender and condition was also not significant ($F(1,36) = 1.81, p > .05$), indicating there was no difference in social desirability scores in any of the four conditions.

Discussion

The current study aimed to investigate the influence of written instructions on gender differences in empathy and whether participants' scores reflected a tendency to respond in a manner perceived to be socially acceptable. The results revealed that in the empathy condition, males scored significantly lower than females on the IRI, while no significant gender difference was found in the social abilities condition. Thus, the data support the hypothesis that gender differences on self-report measures of empathy do not arise due to a difference in ability. Although significant gender differences have been observed in each of the four subscales of the IRI, the differences on the four subscales in the current study were not influenced by the condition or gender of the participants. This result, however, is consistent with previous research suggesting there is a variation in scores obtained on each subscale (Davis, 1980). In contrast to the prediction that participants in the empathy condition would have significantly higher scores on the MC-SDS form C, no significant difference in social desirability scores

was found between the four conditions. Although the results were not significant, males in the empathy condition scored the highest on the MC-SDS form C, whereas females in the empathy condition scored the second highest.

The results are consistent with previous findings demonstrating that when one gender is thought to perform better due to stereotypes, the information presented prior to the task can reduce gender differences (Moe, 2009). In the current study, information about the task that indicated that the measure evaluated empathy, a trait that is stereotypically thought to be feminine (Rueckert, 2011), resulted in gender differences consistent with earlier studies (Davis, 1980). But when participants were led to believe the measure evaluated an ability that had no gender stereotypes associated with it, gender differences on the same measure diminished. The absence of gender differences in empathy scores in the social abilities condition in the current study suggests a lack of difference in empathic ability between genders. The lower scores of males in the empathy condition compared to both females in the empathy condition and males in the social abilities condition suggest that when it was obvious what behavior trait was being evaluated, males may not have been motivated to perform well.

Females in both the empathy and the social abilities conditions performed at almost the same level (slightly higher in the empathy condition). Males' mean performance was significantly worse than that of females in the empathy condition. In their experiment, Michalska et al. (2013) found that even though males and females did not differ in their hemodynamic and pupil dilation responses to painful stimuli, females reported being significantly more upset than males. With respect to the tendency of participants to display themselves in a socially favorable way, the slightly higher social desirability scores of males and females in the empathy condition lend further support to the role of gender stereotypes in the gender differences observed on self-report measures of empathy. Females in the empathy condition scored the highest on the IRI, and males in the empathy condition scored

the lowest on the IRI. Taken together, the results obtained from the MC-SDS form C and the IRI suggest social expectations regarding gender roles and gender stereotypes may lead women to report higher levels of empathetic behavior and men to report lower levels of empathetic behavior.

In studies that have reported gender differences on self-report measures of empathy, a measure of social desirability has not been included (Mestre et al, 2009, Lawrence et al, 2004; Baron-Cohen & Wheelright, 2004; Davis, 1980). In this study, the MC-SDS form C was included to investigate the likelihood of participants presenting themselves in a socially acceptable manner when it was obvious what behavior was being evaluated. However, MC-SDS form C can be used to control for socially desirable responses in self-report measures. The results illuminate the importance of including a measure of social desirability as scores on self-report measures may be influenced by the tendency of participants to present themselves in a way that is considered acceptable by society, and this may lead to incorrect interpretation of data.

Although the results obtained suggest gender differences on self-report measures of empathy arise from social expectations, the current study had certain limitations. Due to the small sample size within each condition, a reliable correlation between IRI and MC-SDS form C scores was not possible. Future studies may benefit from a larger sample size to allow for analysis of the relationship between empathy scores and socially desirable responses. The current study also did not consider demographic information about participants. It is possible that the tendency to report empathy may vary across cultures due to the variance in gender roles across cultures. Future studies may benefit from investigating whether the observed differences are stronger in cultures that have stronger gender role beliefs. Finally, previous research has found gender role orientation is a better predictor of gender differences in empathy than gender itself (Karniol, Gabay, Ochion & Harari, 1998). An inclusion of a gender role orientation measure such as Bem's gender role orientation inventory (Karniol et al., 1998) may be beneficial for further research. This

would allow one to investigate whether participants who score higher on masculinity obtain lower empathy scores regardless of instruction presented prior to the IRI.

In conclusion, the findings that gender differences do not arise from differences in ability are consistent with studies using implicit measures of empathy (Michalska et al., 2013). The current study suggests gender differences in self-report measures of empathy may be influenced by stereotypes and gender role expectations in society. It is possible that when the nature of the measure is made obvious, individuals may be motivated to perform according to social expectations. Gender differences in self-report measures of empathy previously reported may arise due to a tendency of women to over-report empathic behavior and a tendency of men to under-report it (Wager & Ochsner, 2005). The results suggest males and females can obtain the same level of empathy scores when the stereotype associated with a measure is removed. The results from the current study have important implications for future studies aiming to evaluate empathy, as information about the measure may influence the way participants respond. This study challenges the reliability of previously reported gender differences on self-report measures of empathy and adds to a growing area of literature suggesting no gender differences in ability in a number of cognitive domains (Moe, 2009; Dar-Nimrod & Heine, 2006).

References

- Baron-Cohen, S., & Wheelright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high-functioning autism and normal sex differences. *Journal of Autism and Developmental Disorders*, 34, 163-175.
- Baron-Cohen, S., Wheelright, S., Hill, J., Raste, Y., & Plumb, I. (2001). Reading the Mind in the Eyes test revised version: A study with normal adults and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241-251.
- Baron-Cohen, S. (2005). The essential difference: the male and the female brain. Phi Kappa Phi Forum.
- Baron-Cohen, S. (2011). Zero degrees of empathy: A new theory of human cruelty and kindness. London: Penguin.
- Berk, L. E. (2006). Development of sex differences and gender roles. In *Child Development* (7th ed., pp. 518-555). Boston: Pearson.
- Carlo, G., Hausmann, A., Christiansen, S., & Randall, B.

- A. (2003). Sociocognitive and behavioral correlates of a measure of prosocial tendencies for adolescents. *Journal of Early Adolescence*, 23(1), 107-134.
- Cassels, T. G., Chan, S., Chung, W., & Birch, S. A. (2010). The role of culture in affective empathy: cultural and bicultural differences. *Journal of Cognition and Culture*, 10, 309-326.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science*, 314, 435.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology*, 10(85). Retrieved from http://www.eckerd.edu/academics/psychology/files/Davis_1980.pdf
- Eisenberg, N., & Lennon, R. (1983). Sex differences in Empathy and related capacities. *Psychology Bulletin*, 94, 100-131.
- Eisenberg, N., Miller, P. A., Shell, R., McNalley, S., & Shae, C. (1991). Prosocial development in adolescence: A longitudinal study. *Developmental Psychology*, 27, 849-857.
- Howell, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, California: Duxbury Press.
- Karniol, R., Gabay, R., Ochion, Y., & Harari, Y. (1998). Is gender or gender-role orientation a better predictor of empathy in adolescence? *Sex Roles*, 39, 45-59.
- Klein, K. J., & Hodges, S. D. (2001). Gender differences, motivation and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, 27, 720-730.
- Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the Empathy Quotient. *Psychological Medicine*, 34, 911-924.
- LeSure-Lester, G. E. (2000). Relation between empathy and aggression and behavior compliance among abused group home youth. *Child Psychiatry and Human Development*, 31(2), 153-161.
- Mestre, M. V., Samper, P., Frias, M. D., & Tur, A. M. (2009). Are women more empathic than men? A longitudinal study in adolescence. *The Spanish Journal of Psychology*, 12(1), 76-83.
- Michalska, K. J., Kinzler, K. D., & Decety, J. (2013). Age-related sex differences in explicit measures of empathy do not predict brain responses across childhood and adolescence. *Developmental Cognitive Neuroscience*, 3, 22-32.
- Moe, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences*, 19, 21-27.
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. University of Wisconsin-Madison.
- Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and Cognition*, 67, 162-167.
- Rueckert, L. (2011). Gender differences in empathy. in *Psychology of Empathy*, 221-234.
- Wager, T. D., Phan, K. L., Liberzon, I., & Taylor, S. F. (2003). Valence, gender and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *NeuroImage*, 19, 513-531.
- Wager, T. D., & Ochsner, K. N. (2005). Sex differences in the Emotional Brain. *NeuroReport*, 16(2), 85-87.

EFFECTS OF ALTERING THE VALENCE AND AROUSAL OF MUSIC ON SHORT TERM MEMORY TASK

THAO B. NGUYEN

University of California, San Diego

Additional group project members: Christine Phung, Michihiro Watanabe

Study support: Dr. Julian Parris, Scott Freeman

Correspondence concerning this article should be addressed to Thao B. Nguyen.

Email: thaobt.nguyen@gmail.com

Emotion plays an important role in facilitating memory for particular stimuli; however, it is unclear how this is done. Previous studies have looked at valence and arousal, the two dimensions commonly used to describe emotion, and their independent effects on short term memory. However, there are many contradictory findings in the literature. We hoped to gain a better understanding of the effects of valence and arousal by examining the interaction of these two dimensions on word recall. Participants listened to classical music that had varying levels of valence and arousal and were shown lists of words they had to memorize and later recall. We found an interaction between the levels of valence and arousal within music: participants in the positive valence-high arousal music condition remembered more words than participants in the other three music conditions. These results suggest that it is best to listen to happy music, or other positively valenced and arousing music, when studying or performing other cognitive tasks.

Emotions affect many aspects of our everyday lives including decision making, learning, and perceiving the world around us. Previous research on emotion and memory covers a broad spectrum of topics, though a common aim is to explain why and how certain stimuli are better remembered than others. A study by Rubin and Friendly (1986) examined the properties of words that made them memorable and found that a word's emotionality was one of the best indicators of whether or not the word was remembered. Later studies have expanded on the emotional property of stimuli, and have found two distinct elements that make up emotion: valence and arousal (Kensinger & Corkin, 2004). Valence is a continuous measure of how pleasant or unpleasant an emotion is, whereas arousal is a continuous measure that describes how calm or exciting an emotion is (Kensinger, 2004). For example, surprise would be described as having positive valence and high arousal.

Many studies have manipulated the levels of valence and arousal to better understand the effects of emotion on memory. In one study, Bradley, Greenwald, Petry, and Lang (1992) used pictures with varying levels of valence and arousal to examine the effects of these dimensions on immediate recall tasks. They reported that valence and arousal predicted performance on recall equally well. Furthermore, pleasant and arousing pictures were equally more likely to be recalled during the short-term memory task compared to unpleasant and non-arousing pictures. In conjunction with Bradley et al. (1992), Bolls, Lang, and Potter (2001) tested the effects of valence and arousal on short-term memory using auditory stimuli (i.e., radio advertisements). Arousal was found to be a better predictor of memory than valence, but the interaction between valence and arousal was more significant. When the level of arousal was kept constant across different levels of valence, negative

messages were better remembered, but when arousal was high for positive messages, positive messages were better remembered.

Although there is a great amount of literature supporting the influence of emotional stimuli on memory, some research indicates that music with varying levels of valence and arousal have no effect on short-term memory. Jancke and Sandmann (2010) systematically manipulated the tempo and consonance of one original song to capture the different levels of valence and arousal, and had participants listen to one of the four variations while studying a series of non-words. They reported that performance on the verbal recognition memory task for these non-words did not vary with the different music conditions. In their discussion, they suggest that the lack of effect of music on verbal learning may be attributed to their use of one single song rather than multiple songs.

Most research has shown that varying the levels of valence and arousal differentially affects memory; however, there are inconsistent findings related to music. Our study aimed to reconcile these conflicting findings by incorporating multiple concepts into our design. We utilized both auditory and visual stimuli (i.e., music and lists of words) to examine how music with different levels of valence and arousal can affect the encoding of words and whether emotional words are remembered more easily than neutral words. Rather than manipulating one song as was done in Jancke and Sandmann (2010), we chose four different songs that were unfamiliar to participants and varied in valence and arousal ratings.

In alignment with past studies (Bradley et al., 1992; Bolls et al., 2001), we predicted that arousing songs, rather than non-arousing songs, would lead to improved performance on the short-term memory task, regardless of valence. Moreover, we believed participants in the positive valence-high arousal music condition would perform the best on word recall, and participants in the negative valence-low arousal music condition would perform the worst. Finally, it was predicted that emotional words would be remembered more than neutral words.

Method

Design and participants

The present study was a 2 [valence of music: positive and negative] x 2 [arousal of music: high and low] x 3 [valence of words: neutral, positive, and negative] mixed factorial design. Psychology 111B students and friends of the principal investigators were recruited for this experiment. All participants were undergraduate students at the University of California, San Diego. For the pilot study, there were seven participants (four females and three males) between the ages of 21 and 23. For the actual experiment, there were 34 participants (17 females and 17 males) between the ages of 17 and 25. All participants were required to have Internet access, headphones, and normal or corrected to normal vision and hearing.

Materials

The pilot study took place online, through Qualtrics. Participants were required to have Internet access, a computer, and a pair of headphones to listen to 25 classical songs that played for one minute each. Participants used 7-point Likert scales to rate the valence and arousal of the songs. On the valence scale, 1 corresponded to “most unpleasant,” while 7 corresponded to “most pleasant.” On the arousal scale, 1 corresponded to “most calm,” while 7 corresponded to “most arousing.” Participants were also asked to use one of the six universal emotions (happiness, sadness, anger, surprise, disgust, and fear) to best describe the song they just listened to. After each song, participants were asked if they had heard the song before taking the survey.

After data were gathered for all seven participants, averages were taken for each music clip in order to find the best songs that represented the four music conditions: positive valence-high arousal, positive valence-low arousal, negative valence-high arousal, and negative valence-low arousal. Negative valence and low arousal were defined as having averages of 2.5 or below. Positive valence and high arousal were defined as having averages of 5.5 or above. We verified the valence and arousal ratings with the emotions participants used

to describe the songs. For example, songs that were categorized as positive valence-high arousal were expected to be described as “happy” or “surprise,” whereas songs that were categorized as negative valence-high arousal were expected to be described as “fear” or “angry.” We analyzed the music ratings and found two songs that fit the criteria for the positive valence-high arousal and positive valence-low arousal music conditions. We did not find any songs that perfectly fit the criteria for negative valence-high arousal or negative valence-low arousal, so we chose two songs with lowest ratings on the valence scale for the negative valence music conditions. To account for the low and high levels of arousal for these negative valenced songs, we looked at the emotional descriptions participants chose for these songs. We found that nearly all participants described the negative valence-low arousal song as “sad,” and the negative valence-high arousal song as “fear.” Although these two songs did not fit into our original criteria, a majority of participants used emotions that fall within the dimensions of the negative valence-low arousal and negative valence-high arousal music conditions to describe the songs. We analyzed participants’ responses on the survey and chose songs that all seven participants had never heard of before participating in the pilot study.¹

For the actual experiment, we compiled a list of positive, negative, and neutral words using the Affective Norms for English Words (Bradley & Lang, 1999). Words with average ratings between 1 and 3 on the valence dimension were categorized as negative and words with average ratings between 7 and 9 on the valence dimension were categorized as positive. The positive and negative words also had average ratings between 4 and 6 on the arousal dimension. Words with averages between 4 and 6 on both dimensions of valence and arousal were categorized as neutral. We used Python software to code a function to randomly generate 12 different lists that included 15 words each: five positive words, five negative words, and five neutral words. No words were repeated. We used four songs from the pilot study that best represented the four combinations of valence and arousal for this experiment.

The experiment also took place online through Qualtrics. In an attempt to control for the effects of different volumes among the participants, we asked participants to adjust their volume to a relative rating of 4. They listened to a random music clip for 10 seconds, and were shown a 7-point Likert scale which ranged from the following: almost silent, quiet, a little quiet, just right, a little loud, loud, maximum. Participants were instructed to adjust their volume to “just right,” which corresponded with the number 4 on the scale.

Procedure

Participants signed a consent form and then filled out demographic information that included name, age, gender, and fluency in English. We made sure participants had normal or corrected to normal vision and hearing before they could begin the experiment. After the demographic questions, participants were instructed to put on their headphones and proceeded to volume control. They were randomly assigned to one of the four music conditions: positive valence-high arousal ($n = 10$), positive valence-low arousal ($n = 7$), negative valence-high arousal ($n = 7$), and negative valence-low arousal ($n = 10$). In order to obtain a baseline of memory for each participant, all participants went through a block without listening to music while shown the lists of words. The silence block was also to ensure that playing music during study time would not distract participants and decrease recall. The silence and music blocks occurred in random order for each participant, and each block contained six trials.

For each trial in the silence block, participants were instructed to focus on a blank page shown on the screen for 40 seconds. Afterward, participants were shown a list that contained 15 words (five positive, five negative, and five neutral words in randomized order), and were instructed to memorize as many as they could. The list appeared in the center of the screen. The list was presented for 30 seconds. When the list disappeared, participants were asked to recall and type out as many words as they could remember in any order. Participants were given 30 seconds to type out the words. When

Music Condition	M	SD	n
Silence	37.67	16.76	34
Positive Valence-High Arousal	48.63	15.73	10
Positive Valence-Low Arousal	39	9.78	7
negative Valence-High Arousal	37.29	14.41	7
Negative Valence-Low Arousal	35.71	7.54	10

Table 1. Contrast of Words Recalled by Music Condition

Note: M = mean number of words recalled; SD = standard deviation; n = number of participants in that condition. All participants had a silence condition and were randomly assigned to one of the other four music conditions.

the 30 seconds were over, participants were given 60 seconds to solve two simple addition problems as a distractor task. Participants completed six total trials in the silence block, and each trial contained a different set of 15 words.

Participants listened to the same song during the six trials of the music block. Each trial began with music playing for 40 seconds before participants were shown a list of words. During these 40 seconds, a blank page was presented on the screen. Afterward, participants were shown the list and had 30 seconds to memorize the words while the song continued to play in the background. After the 30 seconds were up, the music stopped and participants were asked to recall and type out as many words as they could remember, in any order. Participants were given 30 seconds to type out the words. Two simple addition problems were presented after the recall task. Participants had up to 60 seconds to solve these problems. The six lists used in the silence block and the six lists used in the music block were different. No words or math problems were repeated.

After participants finished the experiment, they were asked to rate the song to which they listened. They rated the songs according to how unpleasant or pleasant and calm or arousing the song made them feel using 7-point Likert scales. Finally, participants were also asked to describe the song using one of the six universal emotions in order to verify the song ratings from the pilot study.

Results

The mean numbers of recalled words for each condition are shown in Table 1. A matched pairs t-test did not yield a significant difference in

percentage of recalled words between silence and music blocks, $t(27) = 1.51, p = 0.143$ (Figure 1). We also separated the distributions of the percentages of words recalled by music condition and used the outlier box plot to identify statistical outliers. We found two outliers in the negative valence-high arousal music condition and excluded them from further analyses. One participant had a very low average ratio of recalled words of 0.167, while the other had a high average ratio of 0.711. After these two data points were excluded (and the six data points from the t-tests were un-excluded), a two between-subject factors with one repeated factor ANOVA was used to analyze the effects music valence and arousal on the percentage and type of words recalled. There was no main effect of music valence, $F(1, 84) = 3.01, p = 0.087$ (Figure 2a), and no main effect of music arousal, $F(1, 84) = 0.44, p = 0.511$ (Figure 2b) on the percentage of words recalled. There were no differences in the percentage of recalled words among the three types of words (positive, negative, and neutral), $F(2, 84) = 0.29, p = 0.751$ (Figure 3a). In addition, a Least Squares Means contrast showed that emotional words (positive and negative) and neutral words were not remembered differently, $F(1, 84) = 0.26, p = 0.614$ (Figure 3b). The ANOVA did yield a statistically significant interaction (see Figure 4) between valence and arousal in music such that those in the positive valence-high arousal condition remembered the most number of words, $F(1, 84) = 4.16, p = 0.45$.

Discussion

This study addressed the question of why certain stimuli are better remembered than others. In past research, it has been shown that people remember

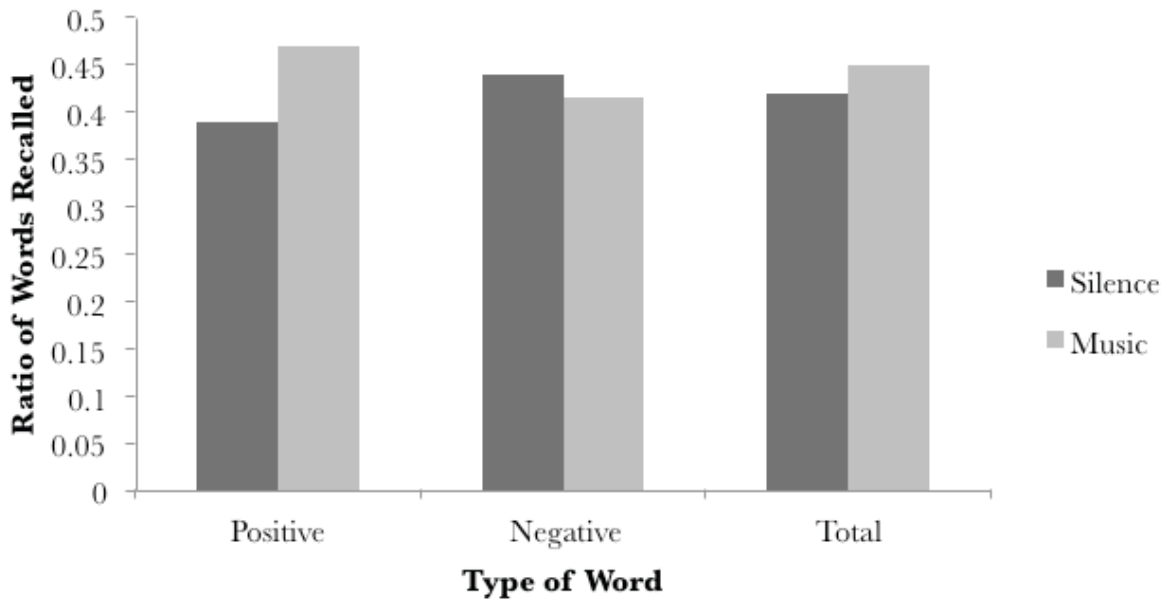


Figure 1. Bar graph showing the comparisons of mean words recalled in silence and music blocks. The word ratios were calculated by comparing the number of words (positive, negative, and overall) each participant recalled to the total number of words (positive, negative, and overall) shown to each participant. For positive words, “music” refers to positively valenced music. For negative words, “music” refers to negatively valenced music. For total words, “music” refers to both positively and negatively valenced music. No significant differences, which suggests that adding music did not distract participants.

emotional stimuli better than neutral stimuli (Brown & Kulik, 1977; Rubin & Friendly, 1986). Furthermore, arousing stimuli are better remembered than non-arousing stimuli; however, the findings on the effect of valence on memory have been contradictory (Bradley et al., 1992; Newhagen & Reeves, 1992). In an attempt to resolve these inconsistent findings Bolls et al. (2001) studied the interaction between valence and arousal. They found that negatively valenced stimuli were better remembered when the level of arousal was not taken into consideration. However, they found that participants remembered positively valenced stimuli more when the level of arousal was considered. Their study showed that stimuli with positive valence-high arousal were remembered the most. There have also been studies on the effects with music during verbal learning. For example, Jancke and Sandmann (2010) manipulated the valence and arousal levels of one song. They found that there were no significant differences in performance on a verbal recognition memory task between the music conditions.

In contrast to Jancke and Sandmann (2010), we found a significant difference between

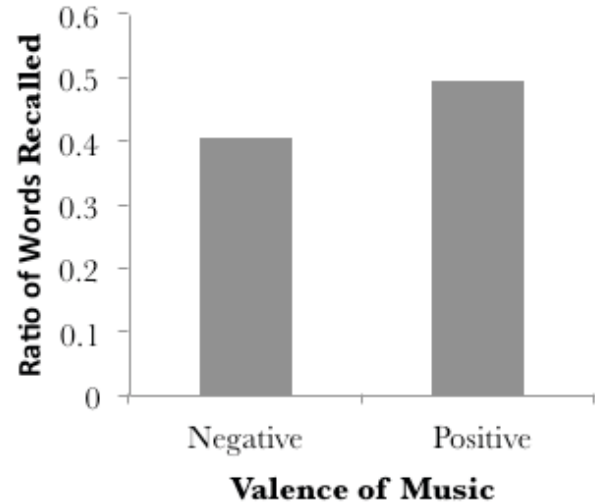


Figure 2a. Mean ratio of words recalled calculated across both levels of music valence, ignoring music arousal. The ratios were calculated by comparing the number of words recalled to the total number of words shown. No significant differences.

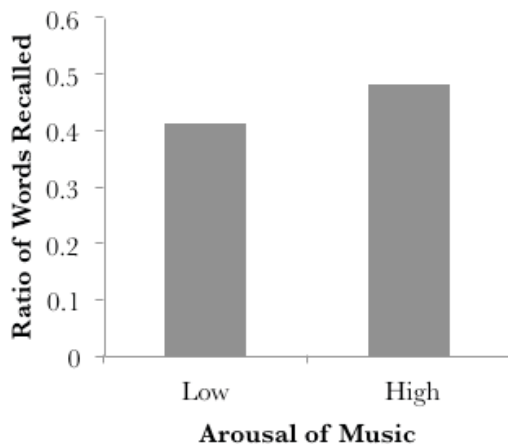


Figure 2b. Mean ratio of words recalled calculated across both levels of music arousal, ignoring music valence. The ratios were calculated by comparing the number of words recalled to the total number of words shown. No significant differences.

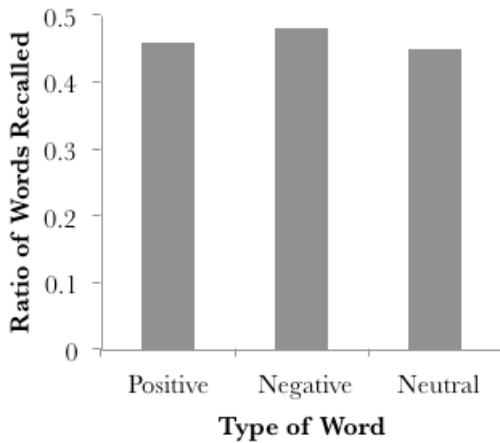


Figure 3a. Mean ratio of words recalled calculated across all levels of word valence. No significant differences.

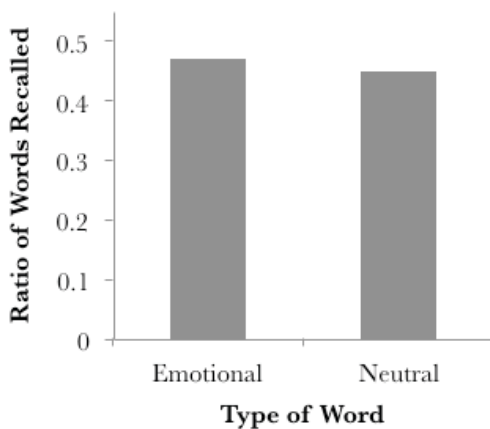


Figure 3b. Mean ratio of words recalled calculated for emotional (positive + negative) and neutral words. No significant differences.

music conditions on short-term memory, and more importantly, an interaction between music valence and arousal. We found that when music arousal was low, there were no differences in the percentage of words recalled between positively and negatively valenced music. However, when music arousal was high, there was a significant difference between positively and negatively valenced music. Our results partly confirm our hypotheses and replicate the findings of Bolls, Lang, and Potter (2001): participants in the positive valence-high arousal music condition performed the best on word recall, while participants in the negative valence-high arousal music condition performed the worst. These results indicate that it is best to listen to happy music, or other positively valenced-high arousing music, when studying and to stay clear of negatively valenced-high arousing music. It could be that emotions with negative valence-high arousal invoke a sense of anxiety and stress, thereby decreasing performance on cognitive tasks. This study can have important implications for students who like to listen to music when studying.

Although we found a significant interaction between music valence and arousal, we did not find any independent effects of music valence and arousal on word recall. Participants in the positively valenced and negatively valenced music conditions (ignoring arousal) did not differ in their performance on word recall. Contrary to our predictions, participants in the low arousing and high arousing music conditions (ignoring valence) did not differ in their performance on word recall. The types of words recalled did not vary as a function of silence or music.

In contrast to previous studies, we did not find that emotional words were better remembered than neutral words. Furthermore, the three types of words (positive, negative, neutral) were not remembered differently. Perhaps we are more likely to remember emotional stimuli than neutral stimuli when it is beneficial to us. For example, in the context of previous studies, participants were asked about significant life events or shown threatening pictures of weapons and mutilated bodies. (Brown & Kulik, 1977; Lane, Chua, & Dolan, 1999). In this

experimental context, however, participants had to remember as many words as they could, and therefore, may have made an effort to remember all the words presented to them equally.

There were many confounds in our study that potentially limited our results. For example, we were required to carry out our experiment on Qualtrics because of our time constraint and limited resources. Qualtrics did not allow us to make all possible randomizations in our study design, such as the presentation of the words in each list. Another limitation with an online experimental design is the lack of control on a participant's study environment. This certainly can affect a participant's performance on the memory task. Although we instructed participants to complete the experiment in an environment free of distraction, there is no guarantee of their compliance. Even though we chose our words from a standardized source (Bradley & Lang, 1999), we did not control for how frequently each word is used and appears in the English language. It could be that more frequently used words are more accessible in one's mind and therefore better remembered than words that are not commonly used in everyday conversation. Furthermore, two of the four music clips (i.e., negative valence-low arousal and negative valence-high arousal) did not meet the rating criteria. Although all participants from the pilot study used the same emotion word to describe these two songs, the music clips may not have represented their respective emotions to all participants since everyone differs in their opinions and criteria for music. Given that everyone has different opinions about music, future studies can examine people's musical preferences to see whether their musical preferences facilitate learning more than positively valenced-high arousing music. Moreover, there was a trend for participants in the positively valenced music condition (ignoring music arousal) to recall more words. Perhaps this trend will become more evident if the study is repeated with longer musical clips and a larger sample size. Our results show it is important to examine both dimensions of valence and arousal simultaneously in order to obtain a clearer understanding of how

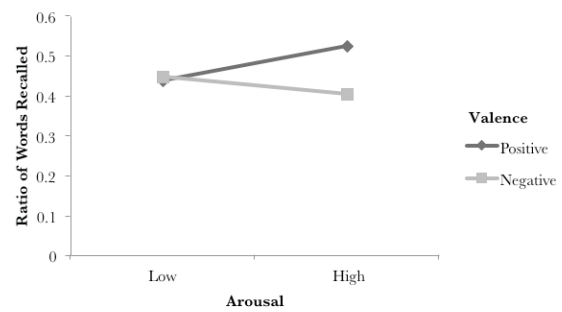


Figure 4a. Statistically significant interaction between music valence and arousal. When arousal is low, the type of valence does not affect the number of words recalled. When arousal is high and valence is positive, participants seemed to remember more words.

emotion affects memory. Some past studies have only looked at one level of a single dimension (i.e., positive stimuli vs. neutral stimuli) or one single dimension (i.e., main effect of valence); however, the interaction between valence and arousal is critical to understanding the complex relationship between emotion and memory.

References

- Bolls, P. D., Lang, A., & Potter, R. F. (2001). The effects of message valence and listener arousal on attention, memory, and facial muscular responses to radio advertisements. *Communication Research, 28*(5), 627-651.
- Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 379-390.
- Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (pp. 1-45). Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73-99.
- Corson, Y., & Verrier, N. (2007). Emotions and false memories: valence or arousal. *Psychological Science, 18*(3), 208-211.
- Jancke L., & Sandmann, P. (2010). Music listening while you learn: no influence of background music on verbal learning. *Behavioral and Brain Functions, 6*(3), 1-14.
- Kensinger, E. A. (2004). Remembering emotional experiences: the contribution of valence and arousal. *Reviews in the Neurosciences, 241-253*.
- Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America, 101*(9), 3310-3315.
- Lane, R. D., Chua, P. M. L., & Dolan, R. J. (1999). Common

effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*, 37, 989-997.

Newhagen, J., & Reeves, B. (1992). This evening's bad news: Effects of compelling negative television news images on memory. *Journal of Communication*, 42, 25-41.

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory and Cognition*, 14(1), 79-94

Footnotes

¹ The four songs used in this experiment were: Johannes Brahms's Hungarian Dance No. 5 (positive valence-high arousal), Mussorgsky's Night on Bald Mountain (negative valence-high arousal), Georges Bizet's Carmen Suite No. 1 - Intermezzo (positive valence-low arousal), and Dvorak's New World (negative valence-low arousal).

GENETICS OF BEHAVIORAL INHIBITION AND APPROACH SYSTEMS: A REVIEW OF THE LITERATURE

SASHA SOMMERFELDT

University of Minnesota

Individual differences in sensitivities of the Behavioral Inhibition System (BIS) and Behavioral Approach System (BAS) have been posited to underlie stable differences in individuals' unique patterns of responding to cues of punishment and reward in their environment. Conceived from the outset as neurological systems, the BIS and BAS have repeatedly been linked to brain structure and function through magnetic resonance imaging and electroencephalogram studies. These significant links to biology suggest an underlying genetic basis for BIS and BAS. Research concerning BIS/BAS in twin pairs as well as BIS/BAS associations with COMT, DRD2, DRD3, DRD4, and 5-HTTLPR genes is reviewed in hopes of elucidating genetic factors underlying personality and psychopathology.

In 1970, Dr. Jeffrey Gray proposed a psychophysiological theory of anxiety and impulsivity as resulting from punishment and reward mechanisms (Gray, 1970). This theory has since been extended into a model of personality referred to as the Reinforcement Sensitivity Theory (RST; Gray, 1970). RST involves three neurological systems thought to underlie emotional behavior: the Behavioral Inhibition System (BIS), the Behavioral Approach System (BAS), and the Fight or Flight System (FFS; Gray, 1987). Individual differences in sensitivities of BIS and BAS have been posited to underlie stable differences in individuals' unique patterns of responding to cues of punishment and reward in their environment.

BIS is thought to underlie trait anxiety. It activates in response to novelty, threat, punishment, and frustrative non-reward (i.e., the omission of anticipated reward). BIS is thus thought to regulate aversive motivation by inhibiting ongoing behavior that may lead to negative or painful outcomes and increase arousal and attention to environmental stimuli (Carver & White, 1994; Takahashi et al., 2007). BIS is also related to negative affect, including fear, anxiety, frustration, and sadness, as these feelings may be experienced as part of the response to aversive stimuli (Carver & White, 1994). BIS has been shown to correlate highly with risk-avoidance (Lorian & Grisham, 2010), neuroticism,

and the processing of unpleasant information (Gomez, Gomez, & Cooper, 2002).

Individuals with higher BIS sensitivity might be more attentive to the presence of uncertainty and sources of potential punishments in the environment. In response to such stimuli, they may be more likely to inhibit behavior, waiting until there is more information and certainty in outcomes before taking action. When a punishment is received, they may experience more negative affect. On the other hand, individuals with lower BIS sensitivity might be less attentive to sources of punishment in the environment and, as a result, may exhibit more risk-taking behavior due to the lack of attention to negative outcomes. Individuals who score low on BIS are less likely to experience negative affect in response to negative outcomes.

BAS, on the other hand, relates to impulsivity, or a proclivity to engage in goal-directed behavior (Carver & White, 1994). BAS activation is thought to regulate responses to reward and non-punishment by motivating goal-directed activity and increasing arousal. BAS is further related to the experience of positive affect, including feelings of hope, elation, and happiness, especially in response to rewards (Carver & White, 1994). BAS has been linked to positive affect, extraversion, and the processing of pleasant information (Gomez et al., 2002).

Individuals with higher BAS sensitivity are thought to be more attentive to sources of reward in the environment and, in response to cues of reward, feel excitement and motivation to pursue the reward, activating behavior to take the steps necessary to obtain it. When a reward is obtained, individuals with high BAS may experience more positive feelings. Individuals with lower BAS sensitivity may be less perceptive of sources of reward in the environment. They might experience less drive and motivation to take steps toward obtaining rewards and may be less likely to feel positive affect when rewards are received.

To assess varying levels of BIS and BAS sensitivity, Carver and White (1994) developed a standardized self-report questionnaire: the BIS/BAS Scales. While traits such as introversion, extraversion, impulsivity, fearfulness, and constraint that are measured through assessments such as Eynsenck's Personality Inventory, the Toddler Temperament Scale, and Cloninger's Temperament and Character Inventory are related to BIS and BAS, the BIS/BAS Scales specifically measure BIS and BAS (Eynsenck & Eynsenck, 1968; Rickman & Davidson, 1994; Fullard, McDevitt, & Caret, 1984; Matheny, 1989; Cloninger, Svrakic, & Przybeck, 1993; Reuter & Hennig, 2005). Items on the BIS/BAS Scales ask how people would respond to certain events on a 4-point Likert scale with anchors of "extremely true of me" to "not at all true of me" (Carver & White, 1994). Items pertaining to the BIS refer to potentially punishing stimuli. BAS includes three subscales: Drive (concerning the persistent pursuit of desired goals), Fun Seeking (focusing on the desire for new rewards and willingness to approach a potentially rewarding event on the spur of the moment), and Reward Responsiveness (concerning positive responses to the occurrence or anticipation of rewards).

Well-balanced BIS and BAS result in adaptive behavior, with decisions and actions appropriately guided by the cues of the environment. Yet extreme sensitivity or insensitivity of either BIS or BAS is maladaptive and linked to psychopathology (Bijttebier, Beck, Claes, & Vandereycken, 2009; Johnson, Turner, & Iwata, 2003). High BIS is associated with anxiety and

depression (Campbell-Sills, Liverant, & Brown, 2004; McFarland, Shankman, Tenke, Bruder, & Klein, 2006; Vervoort et al., 2010), while strong BAS is linked to mania (Urošević, Abramson, Harmon-Jones, & Alloy, 2008; Meyer, Johnson, & Winters, 2001), hyperactive-impulsive symptoms of ADHD (Hundt, Kimbrel, Mitchell, Nelson-Gray, 2008), and substance abuse (Hundt, Kimbrel, Mitchell, Nelson-Gray, 2008; Franken, Muris, Georgieva, 2006; Krmpotich et al., 2013). Inversely, extremely low BAS is associated with anhedonia (Franken, Rassin, & Muris, 2007), the loss of interest or pleasure in usual activities sometimes exhibited in depression, and low BIS is linked to psychopathy (Newman, MacCoon, Vaughn, & Sadeh, 2005). Investigating BIS and BAS thus helps elucidate mechanisms of psychopathology. Furthermore, while the study of psychopathology through syndromes has its benefits, more discrete constructs, such as BIS and BAS, may be better suited to investigations of the biological substrates of maladaptive behavior.

Neurological Correlates

From the outset, BIS and BAS were conceived as neurological systems. Gray's Reinforcement Sensitivity Theory posited that dopaminergic pathways in the limbic circuits make up the BAS, whereas the BIS is comprised of a set of circuits involving the hippocampus, subiculum, septum, and related structures (Gray, 1987). While BAS is more closely associated with dopamine, it has been theorized that BIS is associated with noradrenergic and serotonergic neurotransmitter systems in the brain (Gray & McNaughton, 2000). Modern technologies such as magnetic resonance imaging (MRI) and electroencephalography (EEG) provide opportunities to study the structure and function of the brain in relation to BIS and BAS. Because BIS and BAS are stable personality traits, there should be some relation between brain structure, as well as brain resting-state activity, and BIS and BAS. In addition, because BIS and BAS are sensitive to cues of reward and punishment, there should also be associations between trait measurements of BIS and BAS and brain activation in the context of behavioral tasks.

Structure

Several structural MRI studies have shown associations between BIS and BAS and regional brain volumes. Cherbuin et al. (2008) analyzed hippocampal and amygdalar volumes in relation to BIS and BAS. They found no association between the amygdala and either BIS or BAS; however, they did find a positive association between BIS and hippocampal volume as well as a weak positive association between BAS Reward Responsivity and hippocampal volume. Barrós-Loscertales et al. (2006a) found an association between sensitivity to punishment, a component of BIS measured by the Sensitivity to Rewards and Punishments Questionnaire (SPSRQ; Torrubia, Ávila, Moltó, & Caseras, 2001), and larger hippocampal and amygdalar gray matter volumes. Whereas Barrós-Loscertales et al. (2006b) demonstrated an association between sensitivity to reward (a measure of BAS; Torrubia et al., 2001), and reduced gray matter volume in the bilateral caudate and putamen, left globus pallidus, and right superior frontal cortex. Fuentes, Barrós-Loscertales, Bustamante, Rosell, Costumero, and Ávila (2012) found higher BIS to be associated with reduced right medial orbitofrontal cortex (OFC) and bilateral precuneus volumes. In a longitudinal study, Urošević, Collins, Muetzel, Lim, and Luciana (2012) found that increases in sensitivity to reward across adolescent development were predicted by individual differences in nucleus accumbens (Nacc) and medial OFC baseline volumes. Larger Nacc volume at baseline predicted a more drastic increase in BAS Reward Responsiveness, and larger medial OFC volume at baseline predicted a more drastic increase in global BAS (Urošević et al., 2012). Additional research is needed to replicate these findings across genders and age groups. However, BIS appears to be associated with increased hippocampal and potentially increased amygdalar volumes, as well as decreased volumes of the right medial OFC and bilateral precuneus, while BAS appears to be associated with reduced volumes in the bilateral caudate, putamen, left globus pallidus, and right superior frontal cortex, but increased volumes in the hippocampus, Nacc, and medial OFC.

Function

While gross regional volumetric measurements provide a general depiction of the underlying structure of certain neural systems, these measurements are not directly related to brain function. Thus, studies of neural activity in the brain provide an important perspective in the study of neurological substrates of BIS and BAS. Functional MRI studies analyze the blood-oxygen-level dependent signal (BOLD) as a measure of brain activity, assuming there is increased blood flow to active brain regions (Logothetis, 2003; Ogawa, Menon, Kim, & Ugurbil, 1998). EEG, on the other hand, measures electrical activity on the scalp, which results from the collection of firings from pyramidal neurons with axons perpendicular to the scalp (Kirschstein & Köhling, 2009). Both of these measures can be collected in relation to events in behavioral tasks as well as during resting-state, when the participant is usually instructed not to think about anything in particular, in order to get at the brain's "default" state.

Several event-related fMRI studies have linked BIS and BAS to brain activation. Hahn et al. (2009) found that higher reward sensitivity was associated with increased activation in the ventral striatum and OFC during reward anticipation. Simon et al. (2010) found a positive correlation between BAS and activity in the medial OFC during the receipt and omission of reward, and a negative correlation between BIS and activation in the ventral striatum during the receipt of reward, supporting the idea that individuals with high BIS sensitivity might show a blunted response to rewards. Beaver et al. (2006) found that BAS was positively correlated with activation in the ventral striatum, amygdala, midbrain, OFC, and ventral pallidum in response to pictures of appetizing foods, which would be rewarding stimuli. Beaver, Lawrence, Passmonti, and Calder (2008) found that increased BAS-Drive (Carver & White, 1994) was associated with increased amygdala activation but decreased ventral anterior cingulate cortex (ACC) and ventral striatum activation in response to facial expressions of aggression relative to sad

and neutral expressions. They found that BIS, on the other hand, was associated with increased activation in the dorsal ACC. Additionally, Gray and Braver (2002) found an association between higher BAS and lower working-memory related activation in the caudal ACC and that individuals with higher BAS had better working-memory performance. In an event-related EEG paradigm, Amodio, Master, Yee, and Taylor (2008) found a positive association between BIS and the N200 event-related potential (ERP) on No-Go trials of a Go/No-Go task. The N200 ERP has been linked to the ACC, thus linking BIS with increased ACC activity (van Veen & Carter, 2002).

Resting-state EEG studies have reliably demonstrated a relationship between BAS and left-biased frontal activity (Sutton & Davidson, 1997; Amodio et al., 2008; Coan & Allen, 2003; Harmon-Jones & Allen, 1997; Harmon-Jones & Allen, 1998). Somewhat less reliably, BIS has also been associated with right-biased asymmetrical activity (Sutton & Davidson, 1997). Interestingly, the well-established association between left-biased asymmetrical activity in EEG and BAS has so far failed to be replicated in resting-state fMRI. However, Berkman and Lieberman (2010) did find a positive relationship between BAS and greater left-biased activity in the dorsolateral prefrontal cortex (DLPFC) during approach actions in event-related fMRI. Krmpotich et al. (2013) also found a positive association between BAS and BOLD signal fluctuation in the left DLPFC at rest, but this activity in the left DLPFC was not analyzed in relation to activity in the right hemisphere. Recent PET studies have supported the theory that asymmetrical dopamine signaling from the striatum underlies the asymmetrical frontal activity associated with BIS and BAS (Tomer et al., 2014; Tomer, Goldstein, Wang, Wong, & Volkow, 2008).

BIS and BAS are associated with neural systems involving prefrontal, limbic, and striatal regions, which are heavily influenced by dopaminergic and serotonergic activity. These strong links to biology suggest an underlying genetic basis for BIS and BAS.

Heritability

Studies of monozygotic (MZ) and dizygotic (DZ) twins provide evidence for the heritability of traits. MZ twins share 100% of their genes, while DZ twins share only 50% of their genetic material—the same percentage as any pair of siblings. Thus, if the correlation of a trait in MZ twins is higher than the correlation in DZ twins, it can be deduced that the trait is heritable. The correlation between MZ twin pairs is assumed to be a result of both additive genetic and shared environmental influences. The additive genetic effect is the sum of the average effects of the individual alleles, and shared environmental effects are the environmental influences that make twin siblings similar. In DZ twin pairs, the within-pair correlation is assumed to be due to the sum of half the additive genetic effects and shared environmental influences. Through model-fitting analyses, the extent that a disorder is heritable can be estimated.

Matheny (1989) studied 33 MZ and 32 DZ twin pairs at 12, 18, 24, and 30 months of age. The infants participated in multiple vignette activities designed to assess their behavior in various situations with and without their co-twin and their mother. Videotapes of the vignettes were later coded for the infants' emotional tone for each 2-minute period. Infants also completed the Bayley Scales of Infant Development, including a rating scale of fearfulness of the infant's behavior during Bayley testing (Bayley, 1969). Additionally, mothers of the infants completed the Toddler Temperament Scale, which includes items pertaining to approach and withdrawal in regards to specific activities and events at home. Emotional tone, fearfulness, and (lack of) approach were interpreted as measures of behavioral inhibition. Matheny (1989) found that MZ pair correlations were significantly higher than DZ pair correlations for emotional tone and fearfulness at ages 18, 24, and 30 months, and for approach at ages 12, 18, and 30 months. This consistent pattern of differences supports the hypothesis that there is a strong genetic influence on these aspects of BIS across infant development from 12 to 30 months. Matheny (1989) also found MZ twin pairs shared a more similar pattern of

change over time than DZ twin pairs. Age-to-age changes in behavior were more closely synchronized within MZ pairs, providing evidence for a genetic influence on change in behavior over time. While the measurements employed by Matheny (1989) are associated with certain aspects of BIS and BAS, studies of adults may be better equipped to measure BIS and BAS more specifically.

Takahashi et al. (2007) administered a Japanese version of Carver and White's (1994) BIS/BAS Scales to a sample of 117 adult twin pairs at two separate time points, a little more than 2 years apart. MZ twin-pair correlations for BIS and BAS were higher than correlations for DZ twin-pairs, indicating a moderate genetic influence for BIS and BAS that accounted for approximately one-third of the variance. Because different genes may be influential at different time points (Plomin, 1986), the study employed a longitudinal design to assess changes in genetic influence on BIS and BAS over time. Takahashi et al. (2007) found that genetic influence did not vary across time points and that change in observed temperament was solely due to non-shared environmental influences, thus supporting the hypothesis that BIS and BAS are more genetically stable than other personality traits.

Twin studies support the heritability of BIS and BAS. However, twin studies are not well disposed to illustrate the specific biological mechanisms underlying heritable traits. Candidate gene studies provide more detailed information about the specific biological mechanisms that may underlie the heritability and biological foundations of BIS and BAS.

Genes related to BIS and BAS

COMT

COMT Val158Met is a coding variant encoding the catechol-O-methyltransferase enzyme, which metabolizes dopamine and other catecholamines by inactivating them in the synaptic cleft (Gogos et al., 1998). COMT contains a common functional single nucleotide polymorphism (SNP) in codon 158, resulting from a G-to-A base-pair substitution in the coding sequence of the gene (Lachman et al., 1996). This substitution results in

the gene coding for the production of methionine instead of valine, altering the form of the catechol-O-methyltransferase enzyme. The methionine form of the enzyme is less thermostable than the valine form, and thus has 3 to 4 times lower activity at physiologically relevant temperatures. Met/Met genotypes have low catechol-O-methyltransferase enzyme activity, while Val/Val genotypes have high catechol-O-methyltransferase enzyme activity, and Val/Met genotypes have intermediate levels of the enzyme activity.

Since the catechol-O-methyltransferase enzyme catabolizes dopamine, the COMT gene is of interest in relation to BIS/BAS. Reuter and Hennig (2005) studied extraversion and the COMT Val158Met polymorphism in 363 healthy Caucasian university students of German ancestry. Extraversion was assessed by the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992), which has five scales: Extraversion, Openness to Experience, Conscientiousness, Agreeableness, and Neuroticism. Cloninger's Temperament and Character Inventory (TCI; Cloninger et al., 1993) provided additional assessment of traits related to BIS/BAS and includes measures of novelty seeking, harm avoidance, reward dependence (which includes aspects of both Reward Responsivity and Drive), and persistence. The TCI also includes subdimensions of Novelty Seeking, such as Exploratory Excitability and Impulsiveness. Reuter and Hennig (2005) found that the COMT Val158Met polymorphism had a significant main effect on both Extraversion and Exploratory Excitability, with Val homozygous genotypes scoring significantly higher on Extraversion and Exploratory Excitability scales than heterozygous individuals, or individuals homozygous for Met. These findings support the idea that there is a genetic basis for BAS as the system relates to extraversion and aspects of novelty seeking, however, better support comes from gene studies utilizing the BIS/BAS-specific scales designed by Carver and White (1994).

DRD2

Reuter, Schmitz, Corr, and Hennig (2007) sought to clarify a number of biological relations to

BIS and BAS. In addition to COMT, the researchers also assessed DRD2 genotype and prolactin levels (in a subset of males) and employed the BIS/BAS Scales. Since dopamine inhibits the secretion of prolactin from the pituitary, prolactin concentration is interpreted as being inversely related to the concentration of dopamine (see Ben-Jonathan & Hnasko, 2001). The DRD2 TaqIA polymorphism is correlated with the density of dopamine D2 receptors in the brain (Ritchie & Noble, 2003). Ritchie and Noble (1996) demonstrated that individuals with the A1 allele for DRD2 have a 30-40% reduction in dopamine D2 receptor density compared to individuals homozygous for the A2 allele. The A1A1 genotype occurs in only about 3% of healthy Caucasians (Noble, 2000). Thus, carriers of A1, including heterozygous individuals (A1A2) and individuals homozygous for A1, are often classified as A1+, in contrast to individuals homozygous for the A2 allele, who are classified as A1-. In analyzing interactions with COMT genotypes, carriers of Val (i.e., the combination of heterozygous individuals and those homozygous for Val), will be referred to as Val+, and contrasted to individuals homozygous for Met (Val-).

Reuter et al. (2007) found a significant interaction of COMT and DRD2 on BAS, though a main effect was not noted for either gene. In their sample of 295 German Caucasians, individuals with higher BAS scores were associated with A1+/Val+ and A1-/Val- carriers, whereas low BAS was associated with A1+/Val-, and A1-/Val+ carriers. Additionally, individuals with A1-/Val- and A1+/Val+ genotypes (which predicted higher BAS) had significantly lower baseline prolactin concentrations than individuals with A1-/Val+ and A1+/Val- genotypes (predicting lower BAS scores). Thus, it appears the combination of alleles associated with low receptor density (A1+) and alleles associated with high catechol-O-methyltransferase enzyme activity (Val+), as well as the combination of alleles associated with high receptor density (A1-) but low catechol-O-methyltransferase enzyme activity (Val-) have higher dopamine concentrations and higher BAS scores than individuals with the contrasted genotypes (A1+/Val- and A1-/Val+). This supports

the hypothesis that dopamine activity is associated with BAS (Gray, 1994; Depue, Luciana, Arbisi, Collins, & Leon, 1994).

Montag et al. (2008) tested for associations between the DRD2 TaqIA and the COMT Val158Met polymorphisms, the BIS/BAS Scales, and a psychological measure of emotion processing: the acoustic Affective Startle Reflex Modulation (ASRM) paradigm. The ASRM is a psychophysiological paradigm measuring startle reflex that involves a set of involuntary responses to sudden, intense stimuli and is measured in humans by the amplitude of the eye-blink reflex. It has been shown that the amplitude of the startle reflex is modified by the presentation of affective stimuli of differing emotional valence. Compared to neutral scenes, startle amplitude is potentiated by aversive scenes but attenuated by appetitive scenes (Grillon & Baas, 2003; Bradley, Codispoti, Cuthbert, & Lang, 2001). Additionally, differences in potentiated startle have been shown in individuals scoring low on trait measures of sensation seeking and those scoring high on measures of behavioral inhibition (Lissek & Powers, 2003; Hawk & Kowmas, 2003). Montag et al. (2008) studied 96 healthy females of German origin who were chosen from a larger genetic databank according to their genotype/allele pattern (COMT: Val/Val, Val/Met, and Met/Met; DRD2: A1- and A1+) to form six groups of 16 participants each. Participants watched pictures alternating in valence on a computer screen, with 35 ms startle probes of 106 dB white noise randomly presented during picture presentation. Montag et al. (2008) demonstrated a significant influence of the COMT genotype on the startle reflex in response to the unpleasant picture condition, in which Met/Met genotypes had increased startle compared to either type of Val allele carrying participants. BIS was also negatively correlated with ASRM in the unpleasant picture condition and in the pleasant picture condition. At a trend level, participants with high BIS more often had the Met/Met genotype than either of the Val+ genotypes. Montag et al. (2008) found no significant associations between DRD2 and BIS or BAS.

DRD3

The dopamine D3 receptor has been shown to exert an inhibitory effect on locomotion in rodents (Svensson, Carlsson, & Waters, 1994). Mice with DRD3 knocked out demonstrate increased locomotion in novel environments (Xu et al., 1997). This exploratory behavior in novel environments can be likened to novelty seeking and BAS in humans, highlighting DRD3 as a candidate gene in the study of BAS. A point mutation in DRD3 sometimes results in the substitution of a Serine by a Glycine (Lannfelt et al., 1992). While the in-vivo functional significance of Ser versus Gly alleles is unclear, the receptor products of the two alleles show differential binding affinity for dopamine (Lundstrom & Turpin, 1996).

Henderson et al. (2000) studied associations of BIS, BAS, and COMT and DRD3. Analyses were split into two stages in order to allow replication and to avoid the time and expense of genotyping the entire sample should there be no significant findings in the first set of analyses. In approximately one-third of the sample included in the first analyses (N=862), a significant association between BIS and DRD3 was found. Individuals with at least one Ser allele had higher BIS than individuals homozygous for the Gly allele. A relationship between BIS and DRD3 was not replicated in the second sample (N=1,465). In addition, no association between COMT and either BIS or BAS was demonstrated. The initial positive findings of Henderson et al. (2000) are intriguing, and future research is warranted to determine if group differences between the first and second samples may have played a role in the disparate results. For example, the frequency of other genetic variants may have differed between the samples. Additional candidate genes should be investigated for interactions with DRD3 or contrasting main effects that could obscure the effect of DRD3.

DRD4

The DRD4 gene is associated with the function of the dopamine D4 receptor in the brain. Associations between novelty seeking and the long form (seven-repeat allele) of the DRD4

polymorphism have been reported (Benjamin, Patterson, Greenberg, Murphy, & Hamer, 1996; Ebstein et al. 1996). Thus DRD4 is of interest in relation to BIS and BAS. Keltikangas-Järvinen et al. (2003) studied the relation between aspects of novelty seeking and DRD4 in 150 individuals who scored in the top 10% or bottom 10% on the Novelty Seeking scale of the TCI. While the researchers did not find any association between the different DRD4 alleles and novelty seeking overall, they did find significant associations between different DRD4 alleles and subscales of Novelty Seeking — in particular, the Exploratory Excitability and Impulsiveness scales. Keltikangas-Järvinen et al. (2003) found scores on Exploratory Excitability and Impulsiveness were related to two of the shorter alleles (two- and five-repeats), which conflicts with earlier findings of novelty seeking being associated with seven-repeat DRD4 allele (Benjamin, Patterson, Greenberg, Murphy, & Hamer, 1996; Ebstein et al. 1996). Future research may help explain these disparities by using multiple measures of novelty seeking in multiple populations. Novelty seeking is related to BAS but does not encompass the full range of behaviors associated with the system. There is a paucity of research studying associations between DRD4 and BIS and BAS as measured by the BIS/BAS Scales, which warrants future research.

5-HTTLPR

The majority of research into genes related to BIS and BAS have focused on candidate genes related to the dopamine system, but studies of genes related to serotonin are also important in this context, especially in relation to BIS. Whisman, Richardson, and Smolen (2011) studied the relation of BIS and BAS to the serotonin transporter gene (5-HTTLPR) using three-variant genotyping in a sample of 211 undergraduates. Short variants of 5-HTTLPR include 14 copies of a 20-23 base pair repeat unit and are less transcriptionally efficient compared to the long variant, which has 16 copies (Lesch et al., 1996). Recent research suggests that a SNP within the long version is associated with different functional activity (Hu et al., 2005). Most

frequently, this SNP (rs25531) is A; however, when it is G, functional activity is more comparable to the short allele. Thus, participants with the long allele that contained the G rs25531 SNP were grouped with those who had the short allele for three-variant analyses and compared to those with the long allele that contained A at rs25531. Higher BIS was associated with one or two copies of the alleles linked with low expression of serotonin transporters (i.e., short and long-G alleles), which are theoretically coupled with increased serotonin neurotransmission.

Conclusions

Dopamine and serotonin affect behavior through complex systems in the brain. Neurotransmission of dopamine and serotonin is affected by reuptake, enzyme activity in the synaptic cleft, and the activity of other neurotransmitters. In addition, different types of receptors have distinct functions, and the density of various receptor types varies throughout different brain regions (e.g., D3 receptors are more concentrated in the ventral striatum; Piggott et al., 1999). Neurotransmission will have different effects on behavior depending on the area of the brain implicated. The complexity of neurotransmission underlines the importance of studying multiple genes in relation to behavior, as interactions between the biological features that are coded for by different candidate genes can result in very different phenotypes than could be understood by studying single genes in isolation. As demonstrated by Reuter et al., (2007)'s finding of an interaction between COMT and DRD2, a gene variant may have different associations with behavior depending on which variant of another gene is present.

Genetics are not the only factor underlying biology, and experiences throughout life can significantly impact the structure of the brain. Following the diathesis-stress model, the environment will differentially impact an individual's biology depending upon his or her genetic makeup, which can make him or her predisposed to certain effects. Furthermore, the relationship is not 1:1; that is, an individual with genetic makeup A, when exposed

to environmental stressor B, will not always have the outcome C, which can make the detection of interactions more elusive.

Complex traits can be expected to result from complex interactions of environmental factors with a number of genes and their biological consequences. In order to understand the mechanisms of human behavior, we first need valid and reliable measures of discrete aspects of behavior that are representative of the underlying biological mechanisms. BIS and BAS are refined constructs that are advantaged over symptom clusters of psychopathology in their precision. In addition to good psychological measurements, interactions between multiple genes and environmental factors need to be investigated.

Gray's Reinforcement Sensitivity Theory has remained a prominent model of personality, although it has evolved through the years (Gray 1970, 1982, 1987, 1994; Gray & McNaughton, 2000). The theory is well supported by behavioral research, and its integration of biological mechanisms as the source of personality provides a wide range of interesting research inquiries. RST is relevant to a number of disciplines, including psychology, psychiatry, pharmacology, genetics, and neuroscience. By combining methods from these areas of research, the correlates between the BIS and BAS, genetics, neurochemistry, brain structure, personality, and psychopathology may be elucidated. This is a valuable pursuit, as RST has shown promise as a model of personality and psychopathology. By disentangling the many interactions between these psychological and biological phenomena, the etiology of psychopathology may be illuminated, and treatments based on underlying biology and individual genetic makeups may be advanced.

References

- Amodio, D. M., Master, S. L., Yee, C. M., & Taylor, S. E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: Implications for self-regulation. *Psychophysiology*, *45*, 11-19.
- Barrós-Loscertales, A., Meseguer, V., Sanjuán, A., Belloch, V., Parcet, M. A., Torrubia, R., & Ávila, C. (2006a). Behavioral Inhibition System activity is associated with increased amygdala and hippocampal gray matter volume: a voxel-based morphometry study. *NeuroImage*, *33*, 1011-1015.
- Barrós-Loscertales, A., Meseguer, V., Sanjuán, A., Belloch,

- V., Parcet, M. A., Torrubia, R., & Ávila, C. (2006b). Striatum gray matter reduction in males with an overactive behavioral activation system. *European Journal of Neuroscience*, *24*, 2071-2074.
- Bayley, N. (1969). Bayley scale of infant development. New York, NY: New York Psychological Corporation.
- Beaver, J. D., Lawrence, A. D., Passamonti, L., & Calder, A. J. (2008). Appetitive motivation predicts the neural response to facial signals of aggression. *The Journal of Neuroscience*, *28*, 2719-2725.
- Beaver, J. D., Lawrence, A. D., van Ditzhuijzen, J., Davis, M. H., Woods, A., & Calder, A. J. (2006). Individual differences in reward drive predict neural responses to images of food. *The Journal of Neuroscience*, *26*, 5160-5166.
- Ben-Jonathan, N., & Hnasko, R. (2001). Dopamine as a prolactin (PRL) inhibitor. *Endocrine Reviews* *22*, 724-763.
- Benjamin, J., Li, L., Patterson, C., Greenberg, B. D., Murphy, D. L., & Hamer, D. H. (1996). Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nature Genetics*, *12*, 81-84.
- Berkman, T., & Lieberman, M. D. (2010). Approaching the bad and avoiding the good: Lateral prefrontal cortical asymmetry distinguishes between action and valence. *Journal of Cognitive Neuroscience*, *22*, 1970-1979.
- Bjttebier, P., Beck, I., Claes, L., & Vandereycken, W. (2009). Gray's Reinforcement Sensitivity Theory as a framework for research on personality-psychopathology associations. *Clinical Psychology Review*, *29*, 421-430.
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation: Defensive and appetitive reactions in picture processing. *Emotion*, *1*, 276-298.
- Campbell-Sills, L., Liverant, G. I., & Brown, T. A. (2004). Psychometric evaluation of the Behavioral Inhibition/Behavioral Activation Scales in a large sample of outpatients with anxiety and mood disorders. *Psychological Assessment*, *16*, 244-254.
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*, 319-333.
- Cherbuin, N., Windsor, T. D., Anstey, K. J., Maller, J. J., Meslin, C., & Sachdev, P. S. (2008). Hippocampal volume is positively associated with behavioral inhibition (BIS) in a large community-based sample of mid-life adults: the PATH through life study. *SCAN*, *3*, 262-269.
- Cloninger, C. R., Svrakic, D. M., & Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, *50*, 975-990.
- Coan, J. A., & Allen, J. B. (2003). Frontal EEG asymmetry and the behavioral activation and inhibition systems. *Psychophysiology*, *40*, 106-114.
- Costa, P. T., & McCrae, R. R. (1992). The Revised NEO Personality Inventory (NEO-PLR). Odessa, Florida: Psychological Assessment Resources.
- Depue R. A., Luciana M., Arbis P, Collins P. F., & Leon A. (1994). Dopamine and the structure of personality: Relation of agonist-induced dopamine activity to positive emotionality. *Journal of Personality and Social Psychology* *67*, 485-498.
- Ebstein, R. P., Novick, O., Umansky, R., Priel, B., Osher, Y., & Blaine, D. (1996). Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. *Nature Genetics*, *12*, 78-80.
- Eynsenck, H. J., & Eynsenck, S. B. G. (1968). Manual for the Eynsenck Personality Inventory. San Diego, CA: Educational and Industrial Testing Service.
- Franken, I. H. A., Muris, P., & Georgieva, I. (2006). Gray's model of personality and addiction. *Addictive Behaviors*, *31*, 399-403.
- Franken, I. H. A., Rassin, E., & Muris, P. (2007). The assessment of anhedonia in clinical and non-clinical populations: Further validation of the Snaith-Hamilton Pleasure Scale (SHAPS). *Journal of Affective Disorders*, *99*, 83-89.
- Fuentes, P., Barrós-Loscertales, A., Bustamante, J. C., Rosell, P., Costumero, V., & Ávila, C. (2012). Individual differences in the Behavioral Inhibition System are associated with orbitofrontal cortex and precuneus gray matter volume. *Cognitive, Affective, and Behavioral Neuroscience*, *12*, 491-498.
- Fullard, W., McDevitt, S. C., & Carey, W. B. (1984) Assessing temperament in one to three year old children. *Journal of Pediatric Psychology*, *9*, 205-217.
- Gogos, J. A., Morgan, M., Luine, V., Santha, M., Ogawa, S., Pfaff, D., & Karayiorgou, M. (1998). Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 9991-9996.
- Gomez, R., Gomez, A., & Cooper, A. (2002). Neuroticism and extraversion as predictors of negative and positive emotional information processing: Comparing Eysenck's, Gray's, and Newman's theories. *European Journal of Personality*, *16*, 333-350.
- Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behavioral Research and Therapy*, *8*, 249-266.
- Gray, J. A. (1982). Neuropsychological theory of anxiety. New York: Oxford University Press.
- Gray, J. A. (1987). Perspectives on anxiety and impulsivity: a commentary. *Journal of Research in Personality*, *21*, 493-509.
- Gray, J. A. (1994). Three fundamental emotion systems. In: Edman, P., Davidson, R. J. (Eds.). *The Nature of Emotion: Fundamental Questions* (243-247), New York: Oxford University Press.
- Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety*. Oxford, England: Oxford University Press.
- Gray, J. R., & Braver, T. S. (2002). Personality predicts working-memory-related activation in the caudal anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience*, *2*, 64-75.
- Gray, J. R., Burgess, G. C., Schaefer, A., Yarkoni, T., Larsen, R. J., & Braver, T. S. (2005). Affective personality differences in

- neural processing efficiency confirmed using fMRI. *Cognitive, Affective, & Behavioral Neuroscience*, 5, 182-190.
- Grillon, C., & Baas, J. (2003). A review of the modulation of the startle reflex by affective states and its application in psychiatry. *Clinical Neurophysiology*, 114, 1557-1579.
- Hahn, T., Dresler, T., Ehrlis, A.-C., Plichta, M. M., Heinzl, S., Polak, T., . . . Fallgatter, A. J. (2009). Neural response to reward anticipation is modulated by Gray's impulsivity. *NeuroImage*, 46, 1148-1153.
- Harmon-Jones, E., & Allen, J. J. B. (1997). Behavioral activation sensitivity and resting frontal EEG asymmetry: Covariation of putative indicators related to risk for mood disorders. *Journal of Abnormal Psychology*, 106, 159-163.
- Harmon-Jones, E., & Allen, J. J. B. (1998). Anger and frontal brain activity: EEG asymmetry consistent with approach motivation despite negative affective valence. *Journal of Personality and Social Psychology*, 74, 1310-1316.
- Hawk, L. W., Jr., & Kowmas, A. D. (2003). Affective modulation and prepulse inhibition of startle among undergraduates high and low in behavioral inhibition and approach. *Psychophysiology*, 40, 131-138.
- Henderson, A. S., Korten, A. E., Jorm, A. F., Jacomb, P. A., Christensen, H., Rodgers, B., . . . Easteal, S. (2000). COMT and DRD3 polymorphisms, environmental exposures, and personality traits related to common mental health disorders. *American Journal of Medical Genetics*, 96, 102-107.
- Hu, X., Oroszi, G., Chun, J., Smith, T. L., Goldman, D., & Schuckit, M. A. (2005). An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk. *Alcoholism: Clinical and Experimental Research*, 29, 8-16.
- Hundt, N. E., Kimbrel, N. A., Mitchell, J. T., & Nelson-Gray, R. O. (2008). High BAS, but not low BIS, predicts externalizing symptoms in adults. *Personality and Individual Differences*, 44, 565-575.
- Johnson, S. L., Turner, R. J., & Iwata, N. (2003). BIS/BAS levels and psychiatric disorder: an epidemiological study. *Journal of Psychopathology and Behavioral Assessment*, 25, 25-36.
- Keltikangas-Järvinen, L., Elovainio, M., Kivimäki, M., Lightermann, D., Ekelund, J., & Peltonen, L. (2003). Association between the type 4 dopamine receptor gene polymorphism and novelty seeking. *Psychosomatic Medicine*, 65, 471-476.
- Kirschstein, T., & Köhling, R. (2009). What is the source of EEG? *Clinical EEG and Neuroscience*, 40, 146-149.
- Krmpotich, T. D., Tregellas, J. R., Thompson, L. L., Banich, M. T., Klenk, A. M., & Tanabe, J. L. (2013). Resting-state activity in the left executive control network is associated with behavioral approach and is increased in substance dependence. *Drug and Alcohol Dependence*, 129, 1-7.
- Lachman, H. M., Papolos, D. F., Saito, T., Yu, Y.M., Szumlanski, C. L., & Weinshilboum, R. M. (1996). Human catechol-O-methyltransferase pharmacogenetics: Description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics*, 6, 243-250.
- Lannfelt, L., Sokoloff, P., Martres, M.-P., Pilon, C., Giros, B., Jonsson, E., . . . Schwartz, J.-C. (1992). Amino acid substitution in the dopamine D3 receptor as a useful polymorphism for investigating psychiatric disorders. *Psychiatric Genetics*, 2, 249-256.
- Lesch, K.-P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., . . . Murphy, D. L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, 274, 1527-1531.
- Lissek, S., & Powers, A. S. (2003). Sensation seeking and startle modulation by physically threatening images. *Biological Psychology*, 63, 179-197.
- Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23, 3963-3971.
- Lorian, C. N., & Grisham, J. R. (2010). The safety bias: risk-avoidance and social anxiety pathology. *Behaviour Change*, 27, 29-41.
- Lundstrom, K., & Turpin, M. P. (1996). Proposed Schizophrenia-related gene polymorphism: Expression of the Ser9Gly mutant human dopamine D3 receptor with the Semliki Forest Virus system. *Biochemical and biophysical research communications*, 225, 1068-1072.
- Matheny, A. P. (1989). Children's behavioral inhibition over age and across situations: genetic similarity for a trait during change. *Journal of Personality*, 57, 215-235.
- McFarland, B. R., Shankman, S. A., Tenke, C. E., Bruder, G. E., & Klein, D. N. (2006). Behavioral activation system deficits predict the six-month course of depression. *Journal of Affective Disorders*, 91, 229-234.
- Meyer, B., Johnson, S. L., & Winters, R. (2001). *Journal of Psychopathology and Behavioral Assessment*, 23, 133-143.
- Montag, C., Buckholz, J. W., Hartmann, P., Merz, M., Burk, C., Hennig, J., & Reuter, M. (2008). COMT genetic variation affects fear processing: *Psychophysiological evidence. Behavioral Neuroscience*, 122(4), 901-909.
- Newman, J. P., MacCoon, D. G., Vaughn, L. J., & Sadeh, N. (2005). Validating a distinction between primary and secondary psychopathy with measures of Gray's BIS and BAS constructs. *Journal of Abnormal Psychology*, 114, 319-323.
- Noble, E.P. (2000). Addiction and its reward process through polymorphisms of the D2 dopamine receptor gene: a review. *European Psychiatry*, 15, 79-89.
- Ogawa, S., Menon, R. S., Kim, S.-G., & Ugurbil, K. (1998). On the characteristics of functional magnetic resonance imaging of the brain. *Annual Review Biophysics and Biomolecular structure*, 27, 447-474.
- Piggott, M. A., Marshall, E. F., Thomas, N., Lloyd, S., Court, J. A., Jaros, E., . . . Perry, E. K. (1999). Dopaminergic activities in the human striatum: rostrocaudal gradients of uptake sites and of D1 and D2 but not of D3 receptor binding or dopamine. *Neuroscience*, 90, 433-445.
- Plomin, R. (1986). *Development, genetics, and psychology*. Hillsdale, NJ: Erlbaum. Reuter, M., & Hennig, J. (2005). Association of the functional catechol-O-methyltransferase

- VAL158MET polymorphism with the personality trait of extraversion. *NeuroReport*, *16*, 1135–1138.
- Reuter, M., Schmitz, A., Corr, P., & Hennig, J. (2007). Molecular genetics support Gray's personality theory: the interaction of COMT and DRD2 polymorphisms predicts the behavioural approach system. *International Journal of Neuropsychopharmacology*, *9*, 155–166.
- Rickman, M. D., & Davidson, R. J. (1994). Personality and behavior in parents of temperamentally inhibited and uninhibited children. *Developmental Psychology*, *30*, 346–354.
- Ritchie, T., & Noble, E. P. (1996). [3H]Naloxone binding in the human brain: Alcoholism and the TaqIA D2 dopamine receptor polymorphism. *Brain Research*, *718*, 193–197.
- Ritchie, T., & Noble, E. P. (2003). Association of seven polymorphisms of the D2 dopamine receptor gene with brain receptor-binding characteristics. *Neurochemical Research*, *28*, 73–82.
- Simon, J. J., Walther, S., Fiebach, C. J., Friederich, H.-C., Stippich, C., Weisbrod, M., & Kaiser, S. (2010). Neural reward processing is modulated by approach- and avoidance-related personality traits. *NeuroImage*, *49*, 1868–1874.
- Sutton, S. K., & Davidson, R. J. (1997). Prefrontal brain asymmetry: a biological substrate of the behavioral approach and inhibition systems. *Psychological Science*, *8*, 204–210.
- Svensson, K., Carlsson, A., & Water, N. (1994). Locomotor inhibition by the D3 ligand R-(+)-7-OH-DPAT is independent of changes in dopamine release. *Journal of Neural Transmission*, *95*, 71–74.
- Takahashi, T., Yamagata, S., Kijima, N., Shigemasa, K., Ono, Y., & Ando, J. (2007). Continuity and change in behavioral inhibition and activation systems: a longitudinal behavioral genetic study. *Personality and Individual Differences*, *43*, 1616–1625.
- Tomer, R., Goldstein, R. Z., Wang, G.-J., Wong, C., & Volkow, N. D. (2008). Incentive motivation is associated with striatal dopamine asymmetry. *Biological Psychology*, *77*, 98–101.
- Tomer, R., Slagter, H. A., Christian, B. T., Fox, A. S., King, C. R., Murali, D., Gluck, M. A., & Davidson, R. J. (2014). Love to win or hate to lose? Asymmetry of dopamine D2 receptor binding predicts sensitivity to reward versus punishment. *Journal of Cognitive Neuroscience*, *5*, 1039–1048.
- Torrubia, R., Ávila, C., Moltó, J., & Caseras, X. (2001). The sensitivity to punishment and sensitivity to reward questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences*, *31*, 837–862.
- Urošević, S., Abramson, L. Y., Harmon-Jones, E., & Alloy, L. B. (2008). Dysregulation of the behavioral approach system (BAS) in bipolar spectrum disorders: Review of theory and evidence. *Clinical Psychology Review*, *28*, 1188–1205.
- Urošević, S., Collins, P. F., Muetzel, R. L., Lim, K. O., & Luciana, M. (2012). Longitudinal changes in behavioral approach system sensitivity and brain structures involved in reward processing during adolescence. *Developmental Psychology*, *48*, 1488–1500.
- van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, *14*, 593–602.
- Vervoort, L., Wolters, L. H., Hogendoorn, S. M., de Haan, E., Boer, F., & Prins, P. J. M. (2010). Sensitivity of Gray's Behavioral Inhibition System in clinically anxious and non-anxious children and adolescents. *Personality and Individual Differences*, *48*, 629–633.
- Whisman, M. A., Richardson, E. D., & Smolen, A. (2011). Behavioral inhibition and triallelic genotyping of the serotonin transporter promoter (5-HTTLPR) polymorphism. *Journal of Research in Personality*, *45*, 706–709.
- Xu, M., Koeltzow, T. E., Santiago, G. T., Moratalla, R., Cooper, D. C., Hu, X.-T., . . . Tonegawa, S. (1997). Dopamine D3 receptor mutant mice exhibit increased behavioral sensitivity to concurrent stimulation of D1 and D2 receptors. *Neuron*, *19*, 837–848.

BULLYING, VICTIM, AND AGGRESSOR: PAST EXPERIENCE VERSUS CURRENT BEHAVIOR

FUSHU TAN

University of Oregon

I would like to thank my research advisor, Holly Arrow, for all her extraordinary guidance through this entire process. Thanks to Jordan Pennefather for assisting with data analysis and Mary Hetrick for the use of her data set (Hetrick, 2012). A special thanks to all members of the Groups and War Lab for their assistance in the completion of this research.

Bullying is the most common type of violence in American schools (Swearer & Doll, 2001), and the consequences can persist into adulthood, affecting school achievement, prosocial skills, and psychological well-being for both the victims and the bullies. The current study examined whether past experience with bullying can affect how likely college students are to intervene when someone they know is being bullied. 120 college students (50 males, 70 females) completed a questionnaire that assessed their past experience with physical and relational aggression. They then read a scenario that asked them to imagine someone was spreading vicious rumors about a member of their current campus group. Next, they decided whether they would intervene by contacting the aggressor, the victim, or both. Over 95% of participants reported some past experience as both aggressor and victim. Unexpectedly, males reported significantly more past experience as relational aggressors than females. Over half the participants said they would intervene by contacting both the victim and bully. However, contrary to the hypothesis, past victimization experience did not increase the likelihood of intervening. In fact, past experience scores tended to be somewhat lower for those who intervened. The only significant past experience predictor was that those who chose not to intervene at all tended to have more experience as a bully and less as a victim.

Bullying is not only the most common type of school violence in the United States (Swearer & Doll, 2001) but also a common issue in other countries, such as England, Canada, Japan, Austria, New Zealand, France, and Finland (Smith, et al., 1999; Swearer & Doll, 2001; Ross, 2002). A representative survey conducted for the National Institute of Child Health and Human Development (NICHD) in the United States found that 29.9% American children in grades six through ten had been directly involved in severe and frequent bullying (e.g. physical attack, threat, and emotional violence) in the past 2 months, which included 13% as bullies, 10.6% as victims, and 6.3% as both (Nansel et al., 2001). These numbers do not provide a complete picture of exposure to bullying because a majority

of schoolchildren have experience as “bystanders” who have witnessed aggression among peers without intervening to help their peers (Garbarino & DeLara, 2003). The research reported here investigated whether a person’s past experience as a bully, a victim of bullying, or both affects how likely they are to intervene in a hypothetical bullying incident in the present.

Bullying Behaviors and Roles

Bullying is a specific form of asymmetrical aggression in which one or more powerful individuals repeatedly and intentionally cause pain, psychological stress, and harm to a weaker individual (Olweus, 1993). Bullying is a subtype of aggressive behavior, which Salmivalli (2010)

classifies into three categories: physical bullying, which includes hitting, pushing, and kicking; verbal bullying, which includes name-calling and teasing; and relational bullying (also called indirect bullying), which includes spreading rumors, ignoring peers, and withdrawing friendship (Crick & Grotpeter, 1995; Duffy & Nesdale, 2009; Olweus, 1993).

Bullying is also described as violence in a group context (Sutton and Smith, 1999). Salmivalli and colleagues (1996) identified six roles that children take on during a bullying incident: bully (or ringleader bully), assistant of the bully, reinforcer of the bully, victim, defender of the victim, and outsider. The ringleader bully initiates the bullying. The assistant bully is also active in the bullying process as a follower of the ringleader. The reinforcer supports bullying by laughing, cheering, and providing an appreciative “audience” for the bully. A defender supports, comforts and may protect the victim to discourage the bully from continuing. The outsider does nothing. However, the mere presence of outsiders during a bullying episode plays a distinct role (Salmivalli, 2010) because outsiders may enable bullying situations to continue by not taking action during the event.

Defenders or Outsiders

Challenging the bully’s power by siding with the victim can alter both the severity of bullying and lessen its impact on victims. Research conducted among 573 sixth-grade children indicated that when defenders acted on behalf of the victim, the bullying situation was diffused quickly in comparison to situations with no defenders (Salmivalli, Lagerspetz, Björkqvist, Österman, & Kaukiainen, 1996). Standing up for a victim may also have positive effects on the victim’s adjustment, which improves when the victim has defenders. For example, victims with support exhibited less internalizing disorders than victims without. In a study of children in grades 3 to 5 (Sainio, Veenstra, Huitsing, & Salmivalli, 2011), victims whose classmates defended them were less anxious, less depressed, and more confident than victims without defenders.

Unfortunately, children witnessing bullying rarely take action during a bullying episode

(Whitney & Smith, 1993; Salmivalli, Lappalainen, & Lagerspetz, 1998). Salmivalli’s (2010) review mentions how students often report intentions to defend a victim in a hypothetical situation but fail to defend the victim in real incidents. In one study with Australian middle school-aged children, a video depicting bullying in the presence of bystanders was viewed by 200 late primary (mean age 11.5) and early secondary (mean age 13.5) school students. Questionnaires were employed to assess student attitudes towards victims. Over 43% of students indicated that they would certainly or probably intervene and approximately 23% indicated that they certainly or probably would not. A significant but small correlation of 0.1 ($p < 0.05$) between the social desirability measure and reported intention to support the victim suggests a minor influence on the tendency to report good intentions. The authors propose that in a real bullying situation, unanticipated contingencies might discourage children with good intentions from intervening.

Outsiders and Bystander Inaction

Why don’t outsiders intervene more often? Salmivalli (2010) suggests that this inaction is due to outsiders perceiving other children’s reactions during a bullying episode and following along. As bullying incidents tend to have multiple witnesses, the likelihood of intervening might be reduced by the classical “bystander effect” (Latané & Darley, 1969): helping is less likely when many individuals are witnessing a potentially dangerous or harmful situation. This might be due to the diffusion of responsibility, which occurs when people do not feel personally responsible and fail to act, perhaps hoping someone else will intervene instead. In such situations, bystanders must take four steps before intervening: (1) the bystander must notice something is going on; (2) the bystander must interpret the event as problematic; (3) the bystander must accept responsibility to assist the person in need (rather than waiting for someone else to do so); and (4) the bystander must decide how he or she is going to help (Latané & Darley, 1969).

The bystander can fail to act because of problems at each of the four listed steps. We

analyzed data from a study (Hetrick, 2012) in which group members were informed of a vicious rumor being spread about one of their group members. This ensured that participants noticed that something clearly problematic was happening (steps one and two of the four identified by Latané and Darley). The study focused attention instead on steps three and four: accepting responsibility and deciding what action to take.

Current Research and Hypothesis

For the purpose of this study, we define an outsider as someone who witnesses bullying without taking action. If a group member decides to take action and defend their fellow group member, they shift to a defender role. Group members who fail to act remain outsiders and leave their fellow group member to fend for himself/herself. Fear of becoming the next victim may discourage outsiders from intervening, especially if they have past experience as victims and are traumatized. However, counter-forces such as sympathy for the victim and social norms regarding people's obligation to help each other may encourage outsiders to act (Latané & Darley, 1969). The current research study examines whether the past experience of individuals as a bully, a victim of bullying, or both influences their reaction to a bullying incident in the present. We hypothesized that people who were victims of bullying in the past would be more likely to intervene than those who had not been victimized.

Methods

Participants

This study used an existing data set that was collected from one hundred twenty individuals (50 Males, 70 Females) (Hetrick, 2012). The mean age was 19 years old, (SD=1) with a range of 18 to 29. Of the participants, 78% were Caucasian, 15% were Asian, and the remaining 7% distributed across other races. 118 were undergraduate students and 2 were graduate students. Participants were recruited from the University of Oregon's psychology subject pool, via flyers posted around the university campus. Participants received either

credit for an introduction to psychology class or \$15 for their time. To be eligible for the study, participants had to be a member of a pre-existing group. Potential participants were asked to recruit two other members of their pre-existing group to participate in the study with them. The larger groups included sports teams (N=21), religious organizations (N=9), Greek societies (N=48), and miscellaneous student groups (N=42; e.g. freshman interest groups, Hong Kong student association). The average length of membership was 16 months, (SD=17) with a median of 12 months. Membership length was skewed due to the wide range of responses (1 to 114 months). The gender composition of the 40 three-person groups that participated was 12 all male, 17 all female, 3 majority male, and 8 majority female.

Procedures and Measures

After giving informed consent, participants read a scenario in which someone shared a vicious rumor about a member of their current campus group. Gender specific names for the aggressor and target in the scenario were matched to the gender makeup of the participating group (i.e. members of all-female or majority female groups received a female scenario, and members of all-male or majority male group received a male scenario). After reading the scenario, participants answered questions about whether they would intervene by contacting the aggressor, the victim, or both. They then completed a questionnaire that assessed their past experience with physical and relational aggression. Participants chose from five response options for each of 12 questions ranged from 0 (never happened) to 4 (happened a lot). Aggressor experience and victim experience subscales included six items each: spread rumor, ignore peers, exclude peer group, withdraw friendship, threat with violence, and physical attack. Four about relational aggression and two about physical aggression. See the Appendix for the full scenario and the past experience questionnaire.

Two types of aggression experiences were measured: experience as a victim and as an aggressor. The items for both categories were adapted to better

Forms	Items	Bully		Victim	
		Total %	M (SD)	Total %	M (SD)
Relational	Spread rumor	84.2	1.29 (0.854)	89.2	1.71 (1.048)
	Ignore peers	80	1.19 (0.901)	81.7	1.51 (1.123)
	Exclude peer group	63.3	0.87 (0.829)	74.2	1.43 (1.164)
	Withdraw friendship	38.3	0.48 (0.71)	61.7	1.02 (0.996)
	Overall	66.45	0.958 (0.561)	76.7	1.415 (0.8)
Physical	Threat with violence	23.3	0.34 (0.728)	35.8	0.53 (0.83)
	Physical attack	14.4	0.18 (0.479)	24.1	0.31 (0.605)
	Overall	18.85	0.258 (0.538)	29.95	0.417 (0.646)

Table 1. Prevalence of Involvement by Bullying Assessment Item

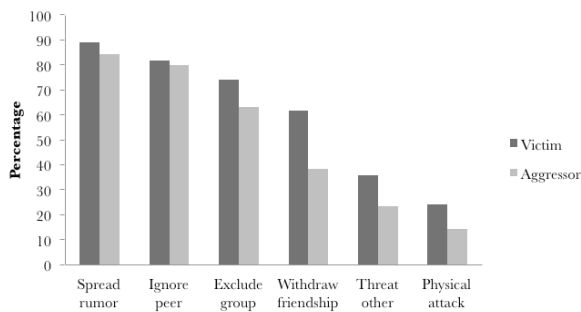


Figure 1. Experience with Specific Bullying Activities

fit the population of our study from the Social and Physical Aggression Personal Experience Interview (SPAPE), originally for children and adolescents who had been targets of aggression (Paquette & Underwood, 1999). Participants were asked to rate how often an event (e.g., “Gossip and rumors were spread about you”) happened to them on a scale of 0 (never happened) to 4 (happened a lot); Cronbach alpha = .70 (n = 6), and how often they instigated an event on a scale of 0 (never) to 4 (frequently), for example, “You spread gossip and rumors about others”; Cronbach alpha = .60 (n = 6).

Participants’ responses about intervening in the bullying scenario were classified into four categories: contact neither, contact aggressor only, contact victim only, and contact both.

Results

Past Experience with Bullying and Victimization

Almost all the participants had some kind of bullying behavior as aggressor and victim. Overall,

118/120 (98%) students reported at least one past instance of bullying behavior, and 119/120 (99%) reported at least one instance of past victimization. The relational bullying scores are higher than the physical bullying scores and some types of behavior were reported more commonly than others. The percentages and means for individual items are shown in Table 1.

At the item level, the percentages of involvement in bullying and victimization are similar for the most common items. For the bullying behaviors, spreading rumors about others (84.2%) and ignoring peers (80%) were the most common types. Likewise, the two most common types of victimization were having rumors spread about them (89.2%) and being ignored by peers (81.7%). For other items, much more victimization than bullying perpetration was reported (see Figure 1). The study also examined the relationship between gender and bullying types. We ran an ANOVA analysis looking at gender differences for the four different subscales and found that males and females reported similar rates of bullying and victimization experience, with an exception that males (M = 1.135, SD = 0.592) were significantly higher on relational aggression than females (M = 0.832, SD = 0.505), $F(1, 119) = 9.087, p = 0.003$ (see Figure 2). No significant differences were observed for the other three subscales, all $F < 1$.

Descriptive statistics and correlations for bullying and victimization experience are reported in Table 2. The internal consistency (Cronbach’s

	1	2	3	4
1. Relational victim experience	—			
2. Physical victim experience	.258**	—		
3. Relational aggressor experience	.404**	.208*	—	
4. Physical aggressor experience	.086	.250**	.161	—

Table 2. Correlations between Bullying and Victimization

Note. $N = 120$. ** $p < 0.01$ level. * $p < 0.05$ level

alpha) of the past experience questionnaire items was moderately high ($\alpha = 0.75$; 12 items). The correlations among subscales indicated a bully/victim pattern in the sample, especially for the relational subscales, $r(118) = 0.404$, $p < 0.01$. A principal component factor analysis of the 12 raw items was conducted to investigate the bully/victim pattern. The four factors extracted were (1) general bully/victim, (2) physical bully/victim, (3) bully not victim, and (4) relational bully. Items for all four factors that loaded at 0.40 or above are shown in Table 3.

Choices about Intervention

Over half the participants (56.67%) said they would intervene by contacting both victim and aggressor. About 20% chose to contact only the current victim, whereas 12.5% said they would contact only the aggressor. Only 10.83% of the

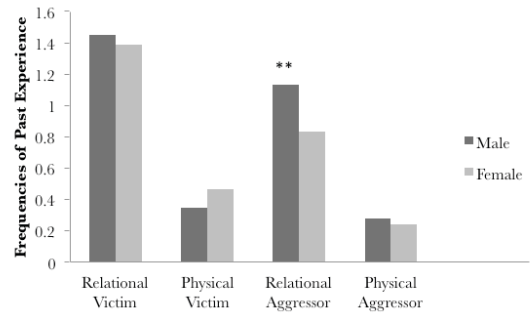


Figure 2. Mean of Past Experience

Note. ** Significantly different, $p < 0.01$.

participants said they would not intervene at all (see Table 4).

Our study tested whether individuals with past experiences as a bully, a victim of bullying, or both would be more likely to intervene when they witness bullying in the present. We ran four one-way ANOVAs examining if past experience of bullying (i.e., physical aggressor, relational aggressor, physical

	Factors			
	G B/V	P B/V	BnV	RB
Gossip spread about you	0.601			
Withdraw friendship	0.602			
Excluded from peer group	0.549			-0.570
Ignored by peers	0.627	-0.452		-0.427
Threatened with physical violence	0.538		-0.615	
Physically attacked	0.465	0.535	-0.461	
Spread gossip about others				0.614
Withdraw your friendship				0.572
Exclude others from group	0.638		0.542	
Ignore other	0.684			
Threaten other with violence		0.685		
You physically hurt other		0.693		

Table 3. Factor Loadings: for 12 Raw Items in Bullying

Note. G B/V = General Bully/Victim. P P/V = Physical Bully/Victim. BnV = Bully not Victim. RB = Relational Bully. Scores below 0.40 were excluded.

Mean

Contact Action	n (%)	Relational Victim Experience	Physical Victim Experience	Relational Aggressor Experience	Physical Aggressor Experience
Contact Neither	13 (10.83)	1.6731	0.3846	1.2692	0.3846
Aggressor Only	15 (12.50)	1.65	0.2333	0.9167	0.1
Victim Only	24 (20.00)	1.4375	0.4792	0.8437	0.3125
Contact Both	68 (56.67)	1.3051	0.4412	0.9485	0.25
Total	120 (100)	1.4146	0.4167	0.9583	0.2583

Table 4. Past Experience Scores with Intervention Choices

victim, and relational victim) was associated with current behaviors of contacting (i.e., contact neither, victim only, aggressor only, and contact both). We found that nothing was significant, $p > 0.1$ for all tests. The pattern of means was actually opposite that predicted for three of four past experience subscales (see Table 4). For relational victim, relational aggressor, and physical aggressor, the highest past experience scores were for contact neither. For physical victim, the highest past experience scores were for contact the victim only.

Multinomial logistic regressions were performed to investigate whether the four past experience factors from the factor analysis predicted intervention. The past experience variables were continuous and the current contacting behaviors were categorical. The logistic regression compared those who contacted neither with those who intervened in some way (victim only, aggressor only, or both) and found that only one of the four regressions was significant, indicating that participants who chose not to take any action tended to score higher on a “bully not victim” factor (with higher scores on aggression and lower scores

for victimization) than those who intervened, ($M = 0.563$, $SD = 0.043$), $B = -0.770$, $Wald(1) = 4.58$, $p = 0.032$. This suggests that “pure” bullies are less likely to intervene than those who score low on this factor. General bully/victim experience regression was close to significant, ($M = 0.483$), $p = 0.064$, showing that participants who have experience both as a bully and a victim in the past are less likely to intervene. The other two regressions (physical bully/victim and relational bully) provided no significant evidence of association between the two past experience factors and current intervention (see Table 5).

Discussion

Main Findings and Alternative Explanations

The hypothesis that past victimization experience would increase the likelihood of intervening was not supported. In fact, some indications of past experience decreased the likelihood of intervening. Participants who did not intervene at all had higher past experience on three of the four subscales. This may be due to the fear of becoming the next victim. Garbarino

	Contact Neither	Aggressor Only	Victim Only	Contact Both
1. G B/V	0.4837	-0.0222	-0.0288	-0.0774
2. P B/V	-0.1054	-0.4108	0.1216	0.0679
3. BnV	0.5635*	-0.9895	-0.22293	-0.0072
4. RB	0.5635*	-0.9895	-0.22293	-0.0072

Table 5. Intervention Choices with Past Experience Factors

Note. G B/V = General Bully/Victim Experience. P B/V = Physical Bully/Victim Experience. BnV = Bully not Victim. RB = Relational Bully. * $p < 0.05$ level.

and DeLara (2002) mentioned in their book that most students at school watch the bullying of their peers with a sense of helplessness, frozen in fear and guilt, but do nothing to help, because they do not want to be the next target. This fear and guilt may be especially acute for those who have been traumatized. In other words, victimization may encourage people to choose the outsider role instead of becoming a defender.

Furthermore, some common personality characteristics found for victims might make them less likely to intervene. Craig (1998) found that victims are usually very shy, quiet, lonely, sensitive, and have lower levels of self-esteem. These characteristics may both predispose them to victimization compared to more confident and extroverted individuals and also indicate a relative lack social skills and inclination to intervene when others are victimized. A study conducted by McClure and Shirataki (1989) identified victims as having an anxious personality pattern combined with poor social communication and problem solving skills. Compared with non-victimized peers, victims are more withdrawn, depressed, worried, and fearful of new situations (Kumpulainen et al., 1998). Unfortunately, we were unable to test personality characteristics as a separate predictor of intervention because we did not collect personality data on our participants.

Over 95% of participants reported some past experience with bullying behaviors as both aggressor and victim. This does not necessary mean that they were both bully and victim in the past. Actually, most participants scored quite low on the past experience scale. For example, spreading rumors and ignoring peers were the most common behaviors for both victim and aggressor, but these behaviors still happened only once or twice in the past for the majority of students. This falls well short of Olweus's (1993) requirement that bullies repeatedly "cause pain, stress, and harm to a weaker individual."

Limitations and Future Research

One limitation already mentioned is that we failed to measure personality characteristics

that might make people hesitant to intervene. For example, people who are shy and quiet may not have the skills needed for effective intervention, whether or not they were bullied in the past. Measuring personality factors in our future research would help us to investigate this alternative explanation. Moreover, college students may not be an ideal group to investigate past bullying experience in our study, because children with lots of experience as victims and/or bullies may be less likely to go to college. Reid (1983) mentioned that both victims and bullies may have problems related to school, such as poor attendance, running away from home, and even committing suicide (Prewitt, 1988). Since our study has limited variance in bullying experience and scores were very low for almost all participants, we may consider including data from other populations in our future research, such as high school students.

Conclusions

In talking informally to high school teachers, educational researchers, and parents of middle school kids, I discovered that they were not surprised by the results because they did not think the hypothesis would be supported. Some of them believed that school intervention programs were making the situation much better than before. Several people even thought that the lower physical bullying scores in the study were due to the efficacy of school prevention programs.

School bullying is not only a problem in the United States but also a common issue in the world. The prevalence of bullying tends to peak in early adolescence, and it is during that period children are especially vulnerable to the psychological harm of bullying (Rigby, 1999). Researchers in previous studies have discussed much about the negative consequences of school bullying. Few have studied how the past bullying experiences of students might affect their responses to a bullying incident when they are older. The results of this research indicate that straightforward predictions about the impact of past experience as a victim are complicated by the fact that many victims also have experience acting as bullies.

References

- Alsaker, F. D., & Vakanover, S. (2001). Early diagnosis and prevention of victimization in kindergarten, In J. Juvonen & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 175-195). New York: The Guilford Press.
- Beatty, L. A., & Alexeyev, E. B. (2008). Bullying in kindergarten and prevention. In D. Pepler & W. Craig (Eds.), *Understanding and addressing bullying: An international perspective*. PrevNet publication series, Vol. 1 (pp. 230-248).
- Bjorkqvist, K., Ekman, K., & Lagerspetz, K. (1982). Bullies and victims: Their ego picture, ideal ego picture and normative ego picture. *Scandinavian Journal of Psychology*, 23, 307-313.
- Boulton, M. J., & Smith, P. K. (1994). Bully/victim problems in middle-school children: Stability, self-perceived competence, peer perceptions, and peer acceptance. *British Journal of Developmental Psychology*, 12, 315-329.
- Boulton, M. J., & Underwood, K. (1992). Bully/victim problems among middle school children. *British Journal of Educational Psychology*, 62, 73-87.
- Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development*, 79, 1185-1229.
- Craig, W. M. (1998). The relationship among bullying, victimization, depression, anxiety, and aggression in elementary school children. *Personality and Individual Differences*, 24, 123-130.
- Crick, N. R., Bigbee, M. A., & Howes, C. (1996). Gender differences in children's normative beliefs about aggression: How do I hurt thee? Let me count the ways. *Child Development*, 67, 1003-1014.
- Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and socialpsychological adjustment. *Child Development*, 66, 710-722.
- Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child Development*, 73, 1134-1142.
- Duffy, A. L., & Nesdale, D. (2009). Peer groups, social identity, and children's bullying behavior. *Social Development* 18, 121-139.
- Elsaesser, C., Gorman-Smith, D., & Henry, D. (2013). The role of the school environment in relational aggression and victimization. *Journal of Youth & Adolescence*, 42, 235-249.
- Garbarino, J. & DeLara, E. (2002). Emotional violence can kill. And words can hurt forever (pp. 1-15). New York, NY: Free Press.
- Gilmartin, B. G. (1987). Peer group antecedents of severe love-shyness in males. *Journal of Personality*, 55, 467-489.
- Goldbaum S., Craig W. M., Pepler D., Connolly J. (2007). Developmental trajectories of victimization: Identifying risk and protective factors. In J. Zins, M. Elias, C. Maher (Eds), *Bullying, victimization, and peer harassment: A handbook of prevention and intervention* (pp.143-60). New York, NY: Haworth Press.
- Gropper, N., & Froschl, M. (2000). The role of gender in young children's teasing and bullying behavior. *Equity & Excellence in Education*, 33, 48-56.
- Hetrick, M. (2012) *Groups and relational aggression: The role of group membership in Defending the Victim*. University of Oregon.
- Hugh-Jones, S., & Smith, P. K. (1999). Self-reports of short- and long-term effects of bullying on children who stammer. *British Journal of Educational Psychology*, 69, 141-158.
- Karatzias, A., Power, K. G., & Swanson, V. (2002). Bullying and victimisation in scottish secondary schools: Same or separate entities?. *Aggressive Behavior*, 28, 45-61.
- Kumpulainen, K., Ra'sa'nen, E., Henttonen, I., Almqvist, F., Kresanov, K., Linna, S. L., Moilanen, I., Piha, J., Puura, K., & Tamminen, T. (1998). Bullying and psychiatric symptoms among elementary school-age children. *Child Abuse & Neglect*, 22, 705-717.
- Lagerspetz, K., & Bjorkqvist, K. (1994). Indirect aggression in boys and girls. In L. R. Huesmann (Ed.), 1212 Card, Stucky, Sawalani, and Little *Aggressive behavior: Current perspectives* (pp. 131 - 150). New York: Plenum Press.
- Lagerspetz, K., Bjorkqvist, K., Berts, M., & King, E. (1982). Group aggression among school children in three schools. *Scandinavian Journal of Psychology*, 23, 45-52.
- Latané, B., & Darley, J. M. (1969). Bystander apathy. *American Scientist*, 57(2), 244-268.
- Loeber, R., & Dishion, T. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin*, 94, 68-99.
- Macklem, G. (2003). Age and gender variables in bullying. *Bullying and teasing: Social power in children's groups* (pp. 49-74). New York, NY: Kluwer Academic/Plenum.
- McClure, M., & Shirataki, S. (1989). Child psychiatry in Japan. *Journal of the American Academy of Child and Adolescent Psychiatry*, 28, 488-492.
- Nansel, T. R., Overpeck, M., Pilla R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *JAMA*, 285, 2094-100.
- Olweus, D. (1991). Bully/victim problems among schoolchildren: Basic facts and effects of a school based intervention program. In D. J. Pepler & E. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp.411-449). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Olweus D. (1993). *Bullying at school what we know and what we can do*. Cambridge, MA: Blackwell, 1993.
- Olweus, D. (1995). Bullying or peer abuse at school: Facts and interventions. *Current Directions in Psychological Science*, 4, 196-200.
- Olweus, D. (1997). Bully/victim problems at school: Knowledge base and an effective intervention program. *The*

- Irish Journal of Psychology*, 18, 170-190.
- Owens, L., Shute, R., & Slee, P. (2000). "Guesswhat I just heard!": Indirect aggression among teenage girls in Australia. *Aggressive Behavior*, 26, 67-83.
- Paquette, J. A., & Underwood, M. K. (1999). Gender differences in young adolescents' experiences of peer victimization: Social and physical aggression. *Merrill-Palmer Quarterly*, 45(2), 242-266.
- Peeters, M., Cillessen, A. H. N., & Scholte, R. H. J. (2010). Clueless or powerful? Identifying subtypes of bullies in adolescence. *Journal of Youth and Adolescence*, 39, 1041-1052.
- Pepler, D. J., & Sedighdeilami, F. (1998). Aggressive girls in Canada: Should we worry about them? Presented at Investing in children: A National Research Conference.
- Prewitt, P. W. (1988). Dealing with Ijime (bullying) among Japanese students. *School Psychology International*, 9, 189-195.
- Reid, K. (1983). Retrospection and persistent school absenteeism. *Educational Research*, 25, 110-115.
- Rigby, K. (1999). Peer victimization at school and the health of secondary school students. *British Journal of Educational Psychology*, 69, 95-104.
- Rigby, K., & Johnson, B. (2006). Expressed readiness of Australian schoolchildren to act as bystanders in support of children who are being bullied. *Educational Psychology*, 3, 425-440.
- Roland, E. (1980). Terror i skolen. Stavenger, Norway: Rogalandsforskning.
- Ross, D. M. (2002). Bullying. In J. Sandoval (Ed.), *Handbook of crisis counseling, intervention, and prevention in schools* (pp.105-135). Mahwah, NJ: Erlbaum.
- Sainio, M., Veenstra, R., Huitsing, G., & Salmivalli, C. (2011). Victims and their defenders: A dyadic approach. *International Journal of Behavioral Development*, 2, 144-151.
- Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior*, 15, 112-120.
- Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior*, 22, 1-15.
- Salmivalli, C., Lappalainen, M., & Lagerspetz, K. (1998). Stability and change of behavior in connection with bullying in schools: A two-year follow-up. *Aggressive Behavior*, 3, 205-218.
- Sanders C., & Phye, G. (2004). What is bullying? Bullying: implications for the classroom (pp. 1-13) San Diego, CA: Elsevier/Academic Press.
- Slee, T. (1994). Situational and interpersonal correlates of anxiety associated with peer victimization. *Child Psychiatry and Human Development*, 25, 97-107.
- Slee, T. (1995). Bullying in the playground: The impact of inter-personal violence on Australian children's perceptions of their play environment. *Children's Environments*, 12, 320-327.
- Smith, P. K., Morita, Y., Junger-Tas, J., Olweus, D., Catalano, R., & Slee, P. (1999). The nature of school bullying: A cross-national perspective. New York: Routledge.
- Sutton, J., & Smith, P. K. (1999). Bullying as a group process: An adaptation of the participant role approach. *Aggressive Behavior*, 25, 97-111.
- Swearer, S. M., & Doll, B. (2001). Bullying in schools: An ecological framework. *Journal of Emotional Abuse*, 2, 7-23.
- Wang, J. , Iannotti, R. J., & Nansel, T. R. (2009). School bullying among adolescents in the united states: Physical, verbal, relational, and cyber. *Journal of Adolescent Health*, 45, 368-375.
- Whitney, I., & Smith, P. K. (1993). A survey of the nature and extent of bullying in junior/middle and secondary schools. *Educational Research*, 35, 3-25.
- Volk, A. , Camilleri, J. , Dane, A. , & Marini, Z. (2012). Is adolescent bullying an evolutionary adaptation?. *Aggressive Behavior*, 38, 222-238.
- Xie, H., Farmer, T. W., & Cairns, B. D. (2003). Different forms of aggression among inner-city African-American children: Gender, configurations, and school social networks. *Journal of School Psychology*, 41, 355-375.

WHEN TO TAKE NOTES: DURING LECTURE OR DURING PODCAST?

**WESTON URIBE¹, TOSHIYA MIYATSU², CAROLE L. YUE¹,
AND ROBERT A. BJORK¹**

¹*University of California, Los Angeles;* ²*Washington University in St. Louis*

This research was supported by Grant 29192G from the James S. McDonnell Foundation. We thank Mo Ally for help scoring and members of the Bjork lab for helpful comments regarding this research.

Contact: Wesluri@g.ucla.edu

Modern technology gives students many study options, including using audio podcasts during study sessions. Although students often take notes during lectures, taking notes during a subsequent study session—rather than during initial learning—may be more beneficial. This is not only because the burden of hurriedly taking notes in class can cause students to miss key ideas during lectures, but also because students may organize the material more effectively (Mayer, 2005). This experiment utilized a between-groups design, with timing of note-taking manipulated between groups. Undergraduates first viewed a multimedia science lesson, followed by studying with an audio podcast of the same lesson. Participants took notes during either the initial lecture or during the study session. Our results show that students who take notes during a subsequent audio podcast version of lecture recall more idea units than those students who take notes during lecture. These results have the potential to demonstrate how students' study strategies should be altered at different points in the learning process and suggest how available technology can be used to optimize learning.

Note-taking is one of the most common activities in the learning process and many studies have demonstrated the benefits of note-taking for learning (e.g., Di Vesta & Gray, 1972; Kiewra & DuBois, 1991). Although the primary benefit of note-taking is generally considered to be its external storage function, evidence suggests that the initial act of recording notes also benefits learning (Kiewra & DuBois, 1991; Peper & Mayer, 1978). One reason for the latter benefit is that when students record notes in class, they must actively engage with the material in order to integrate, organize, and synthesize a spoken lecture into an abbreviated written form (Kiewra, 1989). These types of generative activities promote meaningful learning even without reviewing the notes (Ausubel, 1968; Wittrock, 1974).

To test these benefits, Peper and Mayer (1978, 1986) showed participants a videotaped lecture on computer programming, statistics, or automobile mechanics, during which half the participants took notes and half did not. Note-

takers performed better than non-note-takers on tests of problem solving transfer, while non-note-takers performed better than note-takers on more detailed, technical questions. In addition, note-takers were more likely to include information from other parts of the lecture in their responses to a particular question. These results suggest that note-takers are more likely to focus on higher-order relationships in a lesson, while non-note-takers are more likely to focus on the details.

In addition to the type of test used to measure learning, another variable worth investigating is working memory capacity (WMC), which is a measure of the amount of information that an individual can actively hold in the mind and manipulate. Note-taking during a lecture requires the learner to select and record key points while simultaneously comprehending new information, placing significant demands on his or her WMC (Kobayashi, 2005; Piolat, Olive & Kellogg, 2005). Thus, despite the potential benefits of generative

note-taking, students with low working memory capacity may be at a disadvantage if the cognitive processing required to integrate and organize the information exceeds their ability to perform those functions. In previous studies concerning note-taking, learners only had the opportunity to hear the lecture one time. However, the relatively new educational practice of podcasting allows students to listen to a lecture multiple times, which may help overcome any potential disadvantage of low WMC.

An audio podcast is an audio recording that has been converted into an instantly streamable online version. Podcasts in the realm of education are usually accessible only to class participants, and they are most commonly live recordings of lecture material. Podcasting is a relatively recent phenomenon, with the first audio feeds appearing in July 2003. By mid-2005, there were approximately 10,000 different podcasts available on a wide range of topics (Hew, 2009).

From the learner's perspective, the main advantages of podcasting are the simplicity and convenience that it offers. While lectures are given within a constrained time period and location, learners can listen to podcasts at their own pace and in a location of their choice (Hew, 2009). Additionally, podcasting may function as an external storage device—rather than relying only on the notes they take in class, students have the option of replaying the exact narrative of a lecture.

Since the phenomenon of podcasting is relatively new, research on the topic is scarce. Much of the work that does exist demonstrates how to create effective podcasts (e.g., Geoghegan & Klass, 2005; and Jham, Duraes, Strassler & Sensi, 2008), but research on how students should be utilizing this new technology is severely lacking. One exception examined the learning outcomes associated with podcasting and found that most students who use podcasts think they are helpful, but podcasts actually have little advantage over a written transcript of the lecture. From a practical standpoint, however, full transcripts of lectures are unlikely to be provided in most classes. The process of transcribing a lecture would be time-consuming, and students may be less likely to read through pages of transcripts than listen to an audio recording.

To fully understand the role of podcasting in the learning process, it should be assessed in conjunction with other typical study activities, such as note-taking. Some students may choose to take notes during the initial lecture, and some may choose simply to listen during the initial lecture and then take notes while studying with a podcast. It is possible that the timing of note-taking influences the effectiveness of the generative activities thought to occur while students record notes.

The Current Study

The current study explores how note-taking during different encoding opportunities could affect learning. Participants experienced two study sessions of scientific content, resembling a student attending lecture then studying with a podcast of the lecture. The final test included a free recall test to assess memory for details, as well as a transfer test to assess participants' comprehension of major concepts and their ability to make inferences using those concepts. We hypothesized that, by taking notes during only one study session, it may be possible for students to receive the problem-solving transfer benefit of note-taking as well as the fact-retention benefit of non-note-taking. Specifically, taking notes during only the podcast review, may allow participants to acquire a more broad and organized understanding of the material during lecture before engaging in the generative processes associated with note-taking. Along this line of reasoning, we predict that optimal timing of note-taking (i.e., during a podcast study session) may combine with the generative benefits of note-taking to produce greater overall learning. Furthermore, this benefit may be even greater for people with low WMC—although people with high working memory are linked to better performance on a number of measures (Unsworth & Engle, 2007), shifting the timing of note-taking to the podcast study session may allow low-working memory individuals to overcome any difficulties they usually face.

Method

Participants

A total of 79 undergraduate students from the University of California, Los Angeles

participated in the study. Seven participants were eliminated due to prior knowledge (i.e., scoring above a 7 on the pretest), and seven were eliminated due to technical errors, leaving 65 participants in the final analysis (45 female, average age: 20.1 years). Participants were enrolled in various psychology courses and were given extra credit for participating in the study.

Design

The experiment utilized a between-groups design, with timing of note-taking manipulated between groups. Half the subjects took notes during the initial lecture (lecture notes group), while half took notes during the podcast review (podcast notes group).

Materials and Procedure

Participants took a pre-test on the computer to assess their pre-existing knowledge of astronomy. They were told that the test was self-paced, and they could leave items blank if they did not know the answer.

After completing the pre-test, participants put on headphones and watched a four-minute animated, narrated presentation on the life cycle of a star. Participants in the lecture notes group received a blank sheet of paper and a pen and were told to take as many notes as possible during the lecture. On the other hand, participants in the podcast notes group were not given paper or a pen and were instructed to simply pay attention and try to remember as much information as possible. Both groups were instructed to try to remember as much information as possible, for they would be tested at the end of the experiment. After the initial presentation, the experimenter collected the notes from the lecture notes group. Then, both groups were given a math distractor task for five minutes.

After the initial distractor task, both groups listened to the podcast version of the lecture. They were told that they would be experiencing the same material that was covered in the first presentation; however, this time there would be no video. Participants in the lecture notes group were told that they would not be taking notes, and that they should

just try to listen to best of their ability. Participants in the podcast notes group were now given a blank sheet of paper and a pen and were told to take as many notes as possible. Again, both groups were instructed to remember as much information as possible, and that the test was coming up shortly. After the podcast version of material ended, the notes taken by participants in the podcast notes group were collected. Then, both groups worked on a different math distractor task for five minutes.

After the second distractor task, participants began the final test. The first question was a free recall question that required participants to type all the information that they could remember from the lecture. They were informed that they would have 10 minutes to complete this first question, so they were encouraged to continue writing as much as possible during this entire time. After the time was up, the participants then continued with the rest of the test, which consisted of five transfer questions. Transfer questions tested information that was not directly stated in the material, but the participant should have been able to infer a correct answer from the presented information, thus demonstrating a deeper level of understanding and an ability to integrate abstract concepts effectively. For example, one transfer question read, "Imagine that two clouds of dust and gas become protostars of identical mass at the same time. What could cause these stars to enter the red giant phase at different times?" One possible correct answer was that these two stars could have different levels of external gravity affecting them, which participants could infer from other facts presented in the lesson.

After the final test, the participants completed a demographic questionnaire and a running span task (Broadway & Engle, 2010) to measure working memory capacity. The task took approximately five minutes, and then participants were debriefed and released.

Results

Scoring

Three independent raters scored the free recall and transfer tests for 32 subjects. Free recall responses received one point for each idea unit

recalled, up to a possible total of 24 points. Each transfer question received one point for a correct answer, for a possible total of 5 points. Interrater reliability was .93; discrepancies were discussed among raters, and the remaining data were scored by one of the original raters.

Scores from the running span task were divided into two groups using a median split, creating a high WMC group and a low WMC group.

Recall

Figure 1 presents the proportion of correctly recalled idea units as a function of time of note-taking and WMC. We analyzed recall using a 2 (time of note-taking: lecture versus podcast) x 2 (WMC: high versus low) analysis of variance (ANOVA). As seen in Figure 1, the high WMC group tended to score better on the final recall test ($M = .56, SD = .13$) than the low WMC group ($M = .43, SD = .15$), $F(1,61) = 16.1, MSE = .28, p < .01$.

In support of our hypothesis, we also found a main effect of time of note-taking, $F(1,61) = 7.69, MSE = .14, p = .007$. The podcast notes group recalled more idea units ($M = .54, SD = .15$) than the lecture notes group ($M = .45, SD = .15$). Contrary to our prediction, WMC and time of note-taking did not interact, $F(1,61) = .488, p = .488$.

Transfer

Figure 2 presents the proportion of correctly answered transfer questions as a function of time of note-taking and WMC. A 2 (time of note-taking: lecture versus podcast) x 2 (WMC: high versus low) analysis of variance (ANOVA) revealed that the high WMC group scored better on the transfer test ($M = .57, SD = .24$) than the low WMC group ($M = .38, SD = .26$), $F(1,61) = 8.65, MSE = .56, p = .005$. However, there was no effect of time of note-taking on transfer score, $F(1,61) = .90, p = .35$, and no interaction between WMC and time of note-taking, $F(1,61) = .04, p = .84$.

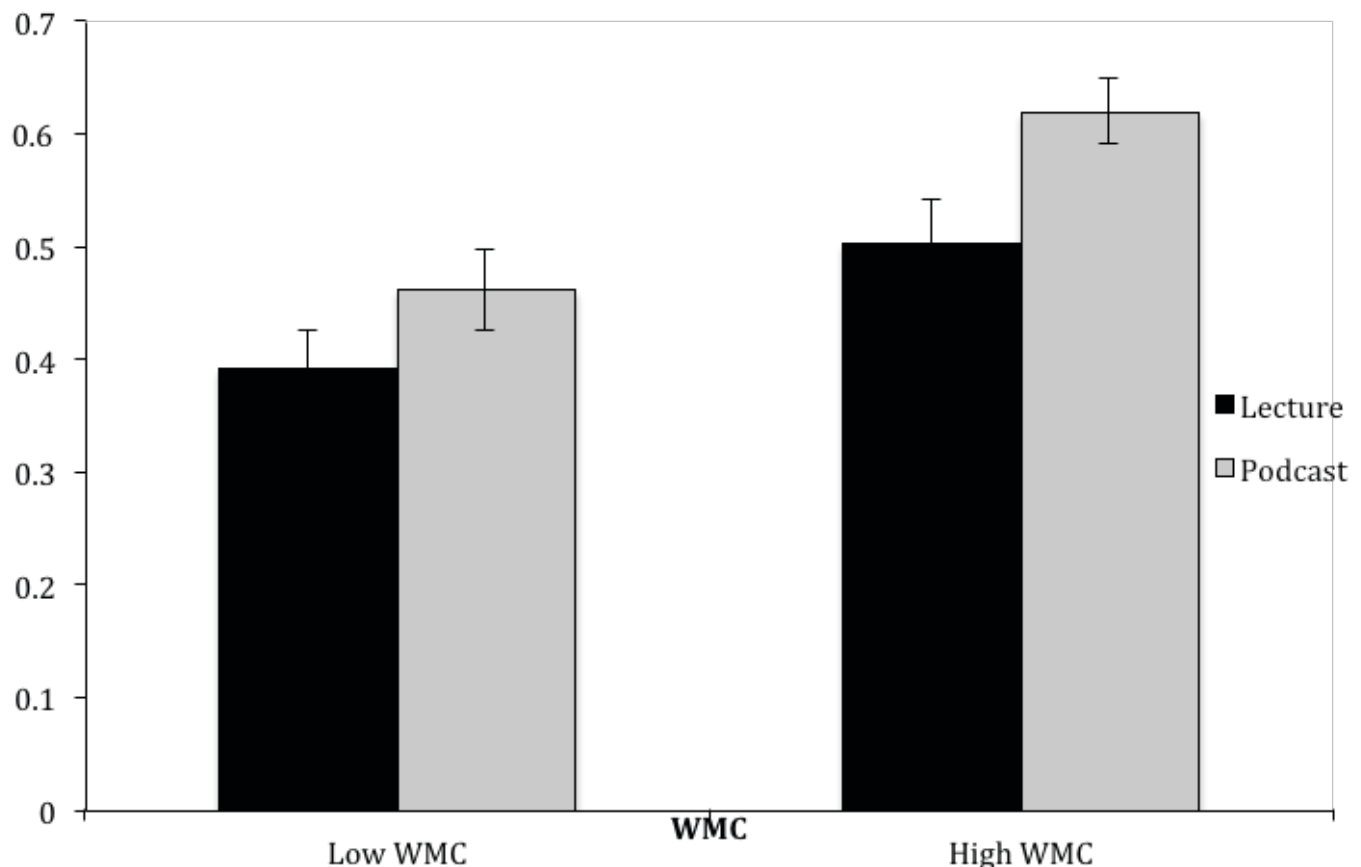


Figure 1. Proportion correct on the final free recall test as a function of time of note-taking and working memory capacity (WMC).

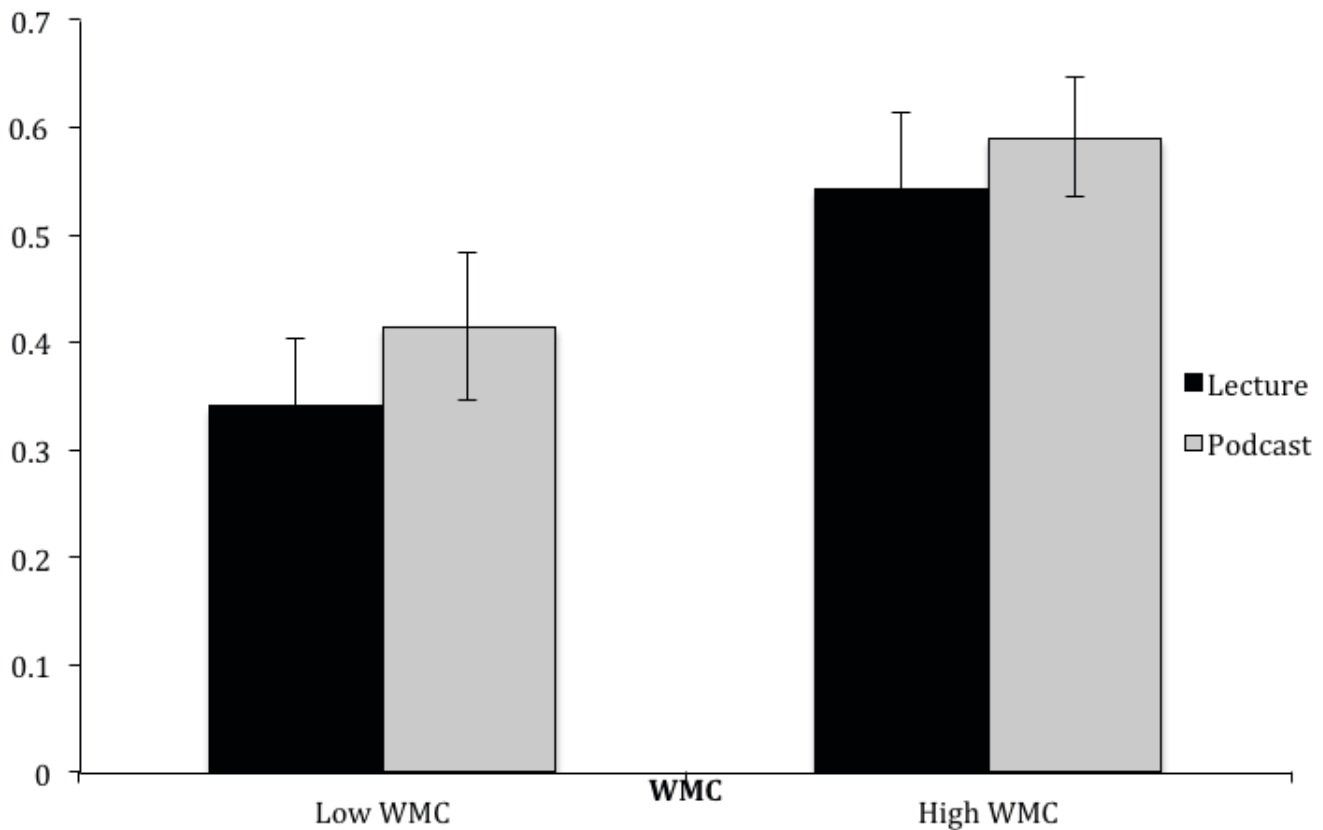


Figure 2. Proportion correct on the transfer test as a function of time of note-taking and working memory capacity (WMC).

Individual Differences in Note-taking Behavior

Using the running span score as a continuous variable, correlational analyses indicated that WMC was positively correlated with the number of idea units contained in participants' notes, $r = .40$, $p = .001$, but there was only a marginal relationship between WMC and the number of words present in the notes, $r = .24$, $p = .06$. Furthermore, students who wrote down more idea units in their notes tended to score better on the final recall test ($r = .43$, $p < .001$), but not the transfer test ($r = .15$, $p = .22$).

Discussion

In the present study, we found that, unsurprisingly, people with high WMC performed better on final recall and transfer tests than people with low WMC (cf. Unsworth & Engle, 2007). A more novel finding, and one that supports our initial hypothesis, showed that participants who took notes during the podcast study session scored higher

on a final free recall test than participants who took notes during the lecture study session. This benefit is likely a result of the increased efficacy of generative activities during a second study session. Specifically, because note-taking requires comprehension, selection of information, and written production processes, performing all these actions simultaneously may cause cognitive overload (Kellogg, 2005), especially during the initial lecture. When students take notes with a podcast after the lecture, however, having processed the information once should lessen the cognitive demand for comprehension. Thus, a podcast study session should be a more optimal note-taking opportunity. Furthermore, having learned the material once provides a unique opportunity for the podcast notes group to make connections between their existing knowledge and what they are processing while taking notes and allows them to understand the material holistically. Therefore, these students are able to be more efficient note-takers and remember more information by utilizing the podcast session

in such a manner. Although we initially predicted that people with low WMC would benefit more than those with high WMC from being able to take notes during a later study session, that was not the case; rather, it appears that note-taking during the podcast study session is equally beneficial to both high and low WMC groups.

We did not find a difference in transfer scores as a result of the timing of note-taking, although the data are numerically in the hypothesized direction. It is possible that the timing of note-taking has less of an effect on inferential abilities than it does on factual memory. Thus, if transfer skills receive the same benefit regardless of the timing of note-taking, then learners may wish to take notes during a podcast study session in order to optimize both retention and transfer, especially if both types of skills will be tested.

Conclusion

Our results suggest that if students were to take notes either during lecture or during a podcast study session, the students who took notes during the podcast study session would likely receive higher grades on a subsequent test of that material. However, many students may choose to take notes during both the initial lecture and the podcast review. Thus, a logical next step would be to examine the three following conditions: two groups that take notes both in the lecture and in the podcast session (one being allowed to add to their previous notes, and the other having to create two separate set of notes for each of the two study sessions), and a third group that does not take notes in any study session.

We also see great potential in conducting experiments that manipulate the ability for participants to pause and rewind the presentation in the podcast session. This function of podcasts is another reason that they are a useful note-taking tool. This advantage may benefit those in the podcast session by making them even better note-takers, and therefore, even better test takers than those taking notes in the lecture session.

This line of research has both theoretical and practical implications—it provides insight on

the cognitive processes that influence how typical study strategies, such as note-taking and studying with podcasts, interact. Furthermore, it offers practical suggestions for how students can improve their learning.

References

- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart, & Winston.
- Broadway, M. J., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42*, 2, 563-570. DOI: 10.3758/BRM.42.2.563.
- Di Vesta, F. J., & Gray, G. S. Listening and note taking. *Journal of Educational Psychology, 63*, 8-14.
- Geoghegan, M. W., & Klass, D. (2005). Podcast solutions. Apress.
- Hew, K. F. (2009). Use of audio podcast in K-12 and higher education: A review of research topics and methodologies. *Educational Technology Research and Development, 57*, 3, 333-357. DOI: 10.1007/s11423-008-9108-3.
- Jham, B. C., Duraes, G. V., Strassler, H. E., & Sensi, L. G., (2008). Joining the Podcast Revolution. *Journal of Dental Education, 72*, 278-281.
- Kiewra, K. A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review, 1*(2), 147-172.
- Kiewra, K. A. & DuBois, N. F. (1991). Note-taking functions and techniques. *Journal of Educational Psychology, 83*, 2, 240-245.
- Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology, 30*, 242-262.
- Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.
- Peper, R. J., & Mayer, R. E. (1978). Note taking as a generative activity. *Journal of Educational Psychology, 70*, 4, 514-522.
- Peper, R. J., & Mayer, R. E. (1986). Generative effects of note-taking during science lectures. *Journal of Educational Psychology, 78*, 34-38.
- Piolat, A., Olive, T., & Kellogg, R. T. (2005) Cognitive Effort during Note Taking. *Applied Cognitive Psychology, 19*, 291-312. DOI: 10.1002/acp.1086.
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological review, 114*(1), 104.
- Witrock, M. C. (1974). Learning as a generative process. *Educational Psychologist, 11*, 87-95.