

DISSERTATION

CARING MORE ABOUT CARELESS RESPONDING:  
APPLYING THE THEORY OF PLANNED BEHAVIOR TO REDUCE  
CARELESS RESPONDING ON ONLINE SURVEYS

Submitted by

Alyssa D. Marshall

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2019

Doctoral Committee:

Advisor: Gwen Fisher

Co-Advisor: Alyssa Gibbons

Kurt Kraiger

Dan Graham

Chris Henle

Copyright by Alyssa D. Marshall 2019

All Rights Reserved

## ABSTRACT

### CARING MORE ABOUT CARELESS RESPONDING: APPLYING THE THEORY OF PLANNED BEHAVIOR TO REDUCE CARELESS RESPONDING ON ONLINE SURVEYS

Careless responding behavior on online surveys is an insidious problem that can distort research findings in concerning and counter-intuitive ways (McGonagle, Huang, & Walsh, 2016). This study aimed to develop practical strategies for reducing careless responding behavior and to provide theoretical support for the notion that careless responding is a planned behavior affected by motivational processes. This study applied the theory of planned behavior (Ajzen, 1991) to develop three careless responding interventions. One intervention was targeted at each of the theory's antecedent variables – attitudes, subjective norms, and perceived behavioral control. Using a sample of 591 MTurk participants, I examined the interventions' effects on antecedent variables, intentions to respond carefully, and six different measures of careless responding behavior as compared to a control group who received no intervention.

Overall, this study found that the theory of planned behavior does explain significant variance in careless responding behavior. This supports the notion that careless responding is a planned behavior affected by motivational processes. Further, this study found that the perceived behavioral control intervention was effective at reducing careless responding on most metrics, though the attitudes and subjective norms interventions were not. None of the interventions produced measurable effects on antecedent variables or intentions to respond carefully. These findings suggest that the perceived behavioral control intervention may be a beneficial addition to future survey research.

## ACKNOWLEDGEMENTS

I want to thank everyone who has supported and encouraged me during my time as a graduate student. Getting a doctorate is hard work, but it is a whole lot easier when you are surrounded by people who make you laugh, inspire you, and support you.

First, thank you to my advisor, Dr. Kurt Kraiger. I haven't met many grad students who have an advisor that is as knowledgeable, supportive, and considerate as you are. It's hard to pinpoint exactly what makes a mentor great (although I know you're working on it!), but I certainly know you've got it. Your guidance has been invaluable to me over the past five years. Thank you for everything you've done for me and so many other students at CSU.

I'd also like to thank the members of my committee, Drs. Alyssa Gibbons, Dan Graham, and Chris Henle. Your expertise and guidance have strengthened my research in this document and beyond. You have been wonderful teachers and role models, and I greatly appreciate your support.

I'd also like to thank the many friends I made in Colorado along the way. To my cohort – Madison, Dorey, and Colin – I couldn't have done grad school without you. Our photos are perfect, our humor is dark, and our cohort bond is forever. You've all become like siblings to me, and I can't wait to see what you do next. To the rest of the IO students – I've been so lucky to be part of a strong community filled with encouraging and inspiring peers. The people make the place, and you've all made Colorado the absolute best place to be over the last five years. To the many roommates I've had along the way – Crystal, Amy, Madison, and Dorey – thanks for becoming the best friends I could ask for. There has been a lot of bad TV and cheap wine consumed over the years, and you've been the best people to do it with.

Finally and most importantly, thank you to my family for supporting me long before I ever considered grad school. To my parents – you made it so easy for me to achieve anything I desired, and for that I am perpetually grateful. Mom, thank you for being so unflinchingly supportive and selflessly kind. Dad, thank you for showing me the value of education and hard work. To my sisters – I guess you two are okay. No really, thanks for being my first and best friends. You both push and challenge me to be better.

## TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS .....	iii
CHAPTER 1: INTRODUCTION .....	1
Survey Research in Organizations .....	2
Inaccuracy in Survey Responses .....	4
Survey Inaccuracy as Performance .....	10
Effects of Careless Responding .....	11
Measuring Careless Responding .....	14
Causes of Careless Responding .....	23
Applying Motivational Theory to Careless Responding .....	33
The Theory of Planned Behavior .....	35
Present Study .....	37
Attitudes .....	38
Subjective Norms .....	43
Perceived Behavioral Control .....	47
CHAPTER 2: METHODS .....	54
Participants .....	54
Design .....	54
Procedures .....	55
Manipulations .....	56
Measures .....	59
CHAPTER 3: RESULTS .....	65
Demographics .....	65
Descriptive Statistics .....	65
Theory of Planned Behavior Applied to Careless Responding Behavior .....	66
Careless Responding Interventions .....	70
CHAPTER 4: DISCUSSION .....	82
Theory of Planned Behavior Applied to Careless Responding .....	82
Efficacy of Theory of Planned Behavior Interventions .....	83

Careless Responding Metrics.....	93
Future Research .....	101
Practical Implications.....	103
Strengths & Limitations .....	105
Conclusion .....	108
REFERENCES .....	109
APPENDIX A: INTERVENTION MATERIALS .....	124
Study 1 Attitudes Intervention Materials .....	124
Study 2 Subjective Norms Manipulation Materials .....	124
Study 3 Perceived Behavioral Control Manipulation Materials .....	125
APPENDIX B: SURVEY ITEMS .....	126

## CHAPTER 1: INTRODUCTION

Over the last century, survey research has emerged as a predominant research methodology in the social sciences (Rossi, Wright, & Anderson, 1983). Despite the prevalence of survey research, it has not been without controversy. The accuracy of survey data has been a long-standing concern (e.g., Allport, 1927; Cantril, 1944), and the combination of frequent surveys and incentivized participation has led some researchers to express concerns about the quality of survey data being collected in organizational science (e.g., Kraiger, McGonagle, & Sanchez, 2019; Meade & Craig, 2012; Rios, Guo, Mao, & Liu, 2017; Rogelberg, Fisher, Maynard, Hakel, & Horvath, 2001). Much of that concern has focused on the high incidence of careless responding on surveys. Careless responding occurs when participants fail to read or comply with questionnaire instructions or item content despite their ability to do so, but then respond to the items regardless (Bowling et al., 2016). Because careless responding can distort findings of survey research (McGonagle, Huang, & Walsh, 2016), it is important for researchers to develop strategies to combat it.

Careless responding is caused in large part by low motivation to provide accurate and careful responses to survey questions (Rios et al., 2017). However, this low motivation can stem from a number of different sources, including individual differences, survey design factors, and the survey context (Meade & Craig, 2012). Because careless responding is related to individual differences, the removal of careless responses may lead to a restriction of the range of survey participants included in samples (Bowling et al., 2016). This has led some to suggest that developing preventative strategies is preferable to removing careless participant's responses (Bowling et al., 2016; McGonagle et al., 2016).



This research study addresses the need for preventative strategies by testing brief, practical interventions to reduce the incidence of careless responding. In this study, I apply a well-supported motivational theory, the theory of planned behavior (Ajzen, 1991), to design three interventions that target the theory's primary antecedents of behavior – attitudes, subjective norms, and perceived behavioral control. Ultimately, the purpose of this study was twofold. First, it was designed to deepen our understanding of the theoretical basis for careless responding behavior by testing the applicability of motivational theory to careless responding behavior. Second, the study provides and tests practical methods that could be used for the prevention of careless responding behavior in the context of both academic and applied research.

### **Survey Research in Organizations**

A survey is a procedure for systematically collecting data from a sample of a population for the purposes of quantitatively describing that population (Fowler, 2013). Although the term survey is often used colloquially as a synonym for questionnaire, survey research methods actually include a much broader range of data collection techniques, including face-to-face and telephone interviews, mail and online questionnaires, and direct observations by external observers (De Vaus, 2013).

During the decades since the internet revolution, online questionnaires have become an increasingly prominent survey method due to their unique advantages over other techniques (Pew Research Center, 2018). Online surveys allow researchers to collect data in a way that is usually quicker and cheaper than other methods. With online surveys, it is not necessary to spend resources on interviewers to conduct in-person or phone interviews, nor is it necessary to spend time and money sending surveys through the mail to potential respondents. Additionally, data collected online does not require manual entry, and it can typically be analyzed using computer

software with relative ease (Evans & Mathur, 2005; Rogelberg, Church, Waclawski, & Stanton, 2008; Wright, 2005).

Because of these advantages, online survey research has become a popular research method for organizational scientists working in both academic and organizational contexts (Rogelberg et al., 2008). For academics, online survey research has become a way to speed data collection, leading to the production and publication of more research papers in a shorter period of time (Couper & Miller, 2009). For those conducting applied research, online employee surveying provides opportunities to collect data that can be used to assess needs and make changes in a timely fashion (Rogelberg et al., 2008).

Despite the predominance of survey research, its use has not been without controversy. The accuracy of survey data has been a long-standing concern in the social sciences (e.g., Allport, 1927; Cantril, 1944). Critics have focused on a variety of issues, including that samples may not represent their intended populations, that question order and phrasing may impact findings, and that participants may be motivated to respond inaccurately (Krosnick, 1999).

With the emergence of online questionnaires as a dominant survey technique, many of these concerns have been exacerbated. It is not possible to randomly survey internet users, because unlike phone numbers and home addresses, lists of email addresses or other avenues for contacting individuals online are not publicly available. Because this information is not available, online surveys often require the user to visit particular sites or subscribe to mailing lists in order to be asked to participate in a survey. This makes it particularly difficult to collect representative samples of the general population using online questionnaires (Pew Research Center, 2018). Additionally, the rapid proliferation of internet surveys has led to concerns about over-surveying and the effects it may have on response rates and quality (Rogelberg et al., 2008).

In general, online surveys appear to have a lower response rate than paper surveys (Manfreda, Berzelak, Vehovar, Bosnjak, & Haas, 2008), perhaps because they are often perceived as spam by both individuals and automated spam email folders (Evans & Mathur, 2005). When participants do respond to online surveys, researchers are often concerned about the accuracy of their survey responses, because incentives and participation requirements may lead individuals to respond to surveys even if they are not motivated or able to provide accurate responses. Social pressures may also lead participants to respond in ways that are more socially desirable than accurate (Krosnick, 1999).

The combination of frequent work surveys and incentivized survey participation has led some researchers to express concerns about the quality of survey data being collected in organizational science (e.g., Kraiger et al., 2019; Landers & Behrend, 2012; Meade & Craig, 2012; Rios et al., 2017; Rogelberg et al., 2001). Although this problem has begun to attract more attention over the last decade, it remains an underappreciated and under-addressed issue by most organizational researchers (Wertheimer, 2017). To the extent that surveys do generate inaccurate responses, our research findings in organizational psychology may be compromised. The next several sections will outline various forms of inaccuracy in survey responses, why they may occur, and the risks they pose to organizational research.

### **Inaccuracy in Survey Responses**

Inaccuracy in a survey response occurs when a person selects a response option that does not represent their true sentiment. Inaccuracy may be caused by a variety of factors, which can be categorized into two general domains: content-responsive and content-nonresponsive inaccuracy (Nichols, Green, & Schmolck, 1989).

## **Content-Responsive Inaccuracy: Misrepresentation**

The first domain concerns inaccuracies that stem from research participants misrepresenting their true sentiments. Nichols et al. (1989) referred to this as content-responsive faking, because it requires an individual to correctly interpret what an item is asking and then to choose to respond in a way that misrepresents the truth.

Misrepresentation can vary widely in terms of the extent to which the truth has been altered (Johnson, 2005). Some participants may offer responses that stand in direct contrast with their true sentiments, but others may only slightly exaggerate their true feelings or perceptions. In fact, some socially desirable exaggeration may occur unintentionally (Paulhus, 1984).

The direction of the misrepresentation may also vary. Some participants may “fake good” to make themselves, their attitudes, or the situation seem more positive (Johnson, 2005; Nichols et al., 1989). A common example of faking good is socially desirable responding, in which participants overreport attitudes or behaviors that they perceived to be desirable (Krosnick, 1999). However, it can also include flatly lying about attitudes, behavior, or personal characteristics. Survey respondents may also “fake bad” to make themselves, their feelings, and/or the situation appear worse than they are (Johnson, 2005; Nichols et al., 1989).

Misrepresentation is caused by the desire to present a deceptive impression of reality (Sackett, 2011). It is triggered by a participant being more motivated to provide an inaccurate response than an accurate response. As is the case in socially desirable responding, this motivation may stem from a desire to be perceived as better than one actually is. However, misrepresentation may also be motivated by the desire to obtain particular outcomes. In organizational research, this is particularly relevant for surveys that may have serious implications in the eyes of the participant. For example, surveys taken as part of a selection

process or performance appraisal may be particularly vulnerable to misrepresentation if the participant believes their responses could have an impact on their chances of being hired or promoted. For this reason, misrepresentation has become a particular concern on selection assessments (Converse, Peterson, & Griffith, 2009; Sackett, 2011).

However, misrepresentation should also be of concern anytime a participant may reasonably expect their responses to be used to make decisions that impact themselves, their coworkers, or their environment. For example, an organization interested in making changes to an office space may survey employees about their attitudes towards the current environment and potential environmental changes. An employee who would like to see office upgrades may be motivated to misrepresent the extent to which the current office space impedes their work or negatively affects their satisfaction. In other words, employees may be motivated to misrepresent the intensity of their attitudes or outcomes in order to encourage the organization to make positive changes.

### **Content-Nonresponsive Inaccuracy**

The second category of survey inaccuracies occurs because the participant did not read or accurately interpret the item, but nonetheless responded. Nichols et al. (1989) refer to this as content nonresponsivity, because the test taker responded to the item without understanding the content of the question. This can occur for two reasons: linguistic incompetence and careless responding.

**Linguistic incompetence.** Linguistic incompetence refers to a participant's inability to read or interpret an item accurately (Johnson, 2005). This may be due to a limited vocabulary, poor verbal comprehension, or an inability to interpret an item's meaning. Unlike other forms of inaccurate responding, linguistic incompetence is caused by a lack of ability rather than a lack of

motivation. Linguistic incompetence is a particularly pressing concern in samples that may contain individuals with low reading levels, low cognitive ability, or limited language proficiency.

Researchers should take care to combat this form of inaccuracy by developing questionnaires that are appropriate for the reading ability of their least capable respondents (Johnson, 2005). When appropriate and possible, it may be necessary to administer survey questions verbally or in alternate languages. In some domains, specific measures exist that are appropriate for survey respondents who vary in terms of their reading comprehension or cognitive ability. For example, the Faces Scale is a one-item measure of job satisfaction that asks participants to select the facial expression that describes how they feel about their job (Dunham & Herman, 1975). Because the measure is easily understandable even for respondents of low ability, it may minimize response errors due to participants' inability to read or interpret the item.

**Careless responding.** Another form of content-responsive inaccuracy stems from a participant's failure to carefully read or comply with questionnaire instructions or item content despite their ability to do so (Bowling et al., 2016). This type of inaccuracy has been referred to by several names, including careless responding (e.g., Ward & Meade, 2018), insufficient effort responding (e.g., Bowling et al., 2016; Dunn, Heggstad, Shanock, & Theilgard, 2018; McGonagle et al., 2016), inattentive responding (e.g., Fleischer, Meade, & Huang, 2015), and low test-taking motivation (Rios et al., 2017). These terms are often used interchangeably (e.g., Camus, 2015; Brower, 2018; Ward & Meade, 2018; Ward, Meade, Allred, Pappalardo, & Stoughton, 2017), although distinctions are sometimes drawn between them. Careless responding is sometimes considered a broader construct than insufficient effort responding (e.g., Dunn et al.,

2018), although the opposite distinction has also been made (e.g., Huang, Liu, & Bowling, 2015).

It is important to recognize that an individual's test taking motivation may vary widely across situations and even over the course of taking a single survey. However, any level of motivation that does not result in the participant providing an accurate response is problematic. Here, I will use the term *careless responding* to refer to any response that occurs as a result of a participant lacking the motivation to accurately read and respond to an item. On the extreme end, there is *insufficient effort responding*, defined by Dunn et al. (2018) as a situation in which the respondent's goal is to complete the questionnaire in the shortest amount of time with the least amount of effort. However, I also use the term careless responding to refer to situations that are less severe. For example, careless responding can also occur when a participant is initially motivated to respond accurately, but their motivation wanes over the course of a long, uninteresting survey (Meade & Craig, 2012). In this case, they may respond to early items carefully, but their level of motivation falls below the threshold required to provide accurate responses as the survey progresses. A participant might also exert more effort on some types of questions than others. They may pay more attention to shorter questions, those with interesting content, or those that don't require much effort to answer accurately (e.g., multiple choice vs. open ended questions). These individuals' responses to some items would be considered careless, although their overall response pattern would not qualify as prototypical insufficient effort responding.

Like misrepresentation, careless responding is related to participant motivation. Misrepresentation occurs when a participant is very motivated to convey a particular message, but careless responding occurs when the participant is motivated enough to complete the survey,

but not motivated enough to carefully read all the directions, questions, and response options.

Careless responding is particularly concerning in situations in which participants are pressured or incentivized to complete a survey, but there are not any perceived benefits to answering in a careful, accurate manner.

Unfortunately, this scenario is common in both academic and applied research settings. In academic settings, a great deal of organizational research is conducted with student participants who are asked to complete surveys in exchange for required or extra course credit. While students may be very motivated to obtain credit for participating, they typically do not stand to benefit directly from providing accurate, useful information. In most research studies, there is not any substantial effort to motivate students to provide accurate survey responses.

When non-student samples are used, it is common for academic researchers to provide monetary compensation to participants in exchange for their time. These participants are recruited through a variety of means, but one increasingly common recruitment method involves the use of Amazon's Mechanical Turk (MTurk) service to recruit paid research participants (Behrend, Sharek, Meade, & Wiebe, 2011; Cheung, Burns, Sinclair, & Sliter, 2017). On MTurk, participants are paid to complete activities referred to as HITs. Research has found that MTurk samples may contain concerningly high rates of careless responders, with one study finding that up to 42% of respondents fail attention checks (Fleisher et al., 2015). This may be because MTurk participants are incentivized to complete studies as quickly as possible so they can move on to additional paying tasks. Many MTurk workers may complete multiple studies in a sitting, which may increase the likelihood that their attention wanes during participation (Cheung et al., 2017).



In organizations, employees may also be pressured or required to participate in survey research without feeling the need to respond accurately. It is often the case that organizational research has little direct benefit to the respondents. For employees with regular responsibilities and work deadlines, participating in research may take valuable time away from their work tasks. In these cases, employees may not feel the need to respond carefully or accurately. This may be exacerbated in cases where the survey is anonymous, the organization surveys frequently, or the respondent does not believe survey results will lead to positive change.

### **Survey Inaccuracy as Performance**

In many ways, survey inaccuracy maps onto a broader conceptualization of performance. That is, respondents who respond inaccurately on surveys may be considered poor performers. Performance is defined as the behavior an individual engages in (Campbell, McCloy, Oppler, & Sager, 1993; Wildman, Bedwell, Salas, & Smith-Jentsch, 2011). Whether an individual engages in any particular behavior and the extent to which they do so is a function of two factors: can-do and will-do variables (Schmitt, Cortina, Ingerick, & Wiechmann, 2003). The can-do components are those components that are related to the person's ability to engage in the target behavior. Depending on the behavior in questions, these components could include cognitive ability, physical ability, and relevant experiences. They also influence the ability to develop the relevant knowledge and skills required to perform the behavior. The will-do components capture the extent to which the individual is willing to perform the behavior. It includes variables like conscientiousness and integrity, and it ultimately determines an individual's motivation to complete the behavior. Together, these ability and motivational components interact to affect performance. In order to perform a behavior, an individual must be both able and willing to

engage in the behavior. If either component is missing, performance of the behavior cannot occur (Schmitt et al., 2003).

In the context of survey research, the desired behavior is ultimately the participants' delivery of an accurate survey response. If participants provide inaccurate responses, it is because they lack the ability or motivation to do so. Linguistic incompetence represents a lack of ability to respond accurately (Johnson, 2005), while misrepresentation and careless responding represent the lack of motivation to respond accurately (Bowling et al., 2016). In order to ensure that an accurate response is provided, researchers must take care to ensure that participants are both able and motivated to respond accurately.

While there are relatively straightforward ways to ensure a lack of ability does not cause response inaccuracy (e.g., translate surveys, provide alternate response formats, screen out participants who do not speak the survey language), addressing a lack of motivation has been more challenging for researchers. In this study, I explored avenues for addressing participants' motivation to respond accurately on surveys. Specifically, I focused on motivation in low-stakes assessment situations in which participants may not be motivated to respond carefully to surveys.

### **Effects of Careless Responding**

Careless responding is generally regarded as a source of measurement error in survey research (e.g., Berinsky, Margolis, & Sances, 2014; McGonagle et al., 2016; Meade & Craig, 2012; Ward & Pond, 2015). When careless responses are included as part of a sample, they serve as threats to the validity of the inferences made on the basis of a study's findings (Rios et al., 2017). This can affect the reliability of measures and disturb their factor structure (Huang, Liu, et al., 2015), and careless responses can distort the relationship between measured variables (McGonagle et al., 2016; Rios et al., 2017).

Early research on careless responding often characterized careless responses as random responses (e.g., Beach, 1989), because of the perception that unmotivated participants would select response options at random. In accordance with the perception that responses were random, the resulting measurement error introduced was also assumed to be random (Curran, 2016). When random measurement error is introduced, the relationships between the measured variables are necessarily attenuated, thus, there was a belief that careless participant responses necessarily attenuated a study's relationships (Huang, Liu, et al., 2015).

However, in the last several years many authors have refuted the idea that careless responding is necessarily random (e.g., Huang, Liu, et al., 2015; Meade & Craig, 2012; Wertheimer, 2017). Even when participants do not bother to read or interpret items, they will often still respond in non-random patterns (Huang, Curran, Keeney, Poposki, & DeShon, 2012). This may include selecting the same response repeatedly, selecting responses in a particular order (e.g., 1, 2, 3, 1, 2, 3), or selecting items in a way that mimics a plausible response (e.g., varying responses, but sticking mostly to the positive side of the scale; Curran, 2016; Meade & Craig, 2012). In fact, the last tactic may be employed if the test taker is not interested in providing accurate responses but hopes to avoid being detected as a careless responder. This may occur in instances in which the responder is worried about social repercussions (e.g., in organizational contexts where researchers and participants are coworkers) or incentives being withheld (e.g., on MTurk) if they are suspected of not providing valid responses.

Despite the fact that careless responses may not be random, scale means for careless responders tend to look similar to scale means found for truly random responders. That is, mean scale scores for careless responders tend to fall near the center of a response scale (Huang, Liu, et al., 2015). However, the effect that these scores have on the relationships between variables

found in a study can vary (Kraiger et al., 2019). In the case that scale means for attentive participants are also near the midpoint of a scale, the presence of careless responding can have an attenuating effect on relationships between measured variables. In the case that scale means for attentive participants are farther from the center of a response scale, careless responding can inflate relationships between variables (Huang, Liu, et al., 2015; McGonagle et al., 2016).

In addition, the effects of careless responding can also depend on the extent to which careless responding is present in a particular sampling context. Research has found that rates of careless responding range from less than 5% (Meade & Craig, 2012) to nearly 50% (Fleischer et al., 2015; Hauser & Schwarz, 2016). These variations may be due to differences in metrics and criteria used to determine what constitutes a careless response, but they also may be due to differences across studies and samples in underlying levels of participant motivation to respond accurately (Kraiger et al., 2019).

Because careless responding rates are known to vary widely across samples and to either inflate or attenuate correlations depending on the situation, some have begun to speculate that careless responding may play a role in the replication crisis in psychology (Curran, 2016). In some instances, careless responding may lead to spurious correlations that cannot be replicated when variables are measured using different samples or response scales (Huang, Liu, et al., 2015). Some have begun to call for more uniformity in the way that careless responders are identified and screened out of research studies (Curran, 2016; Kraiger et al, 2019; McGonagle et al., 2016). The next section focuses on the various methods researchers have used to measure and identify careless responses.

## Measuring Careless Responding

There are a variety of methods for detecting careless responses, each of which has been developed to detect a specific pattern of careless responding (Curran, 2016). Because patterns of careless responding can look very different, the metrics used to detect them often disagree on whether a single response should be flagged as careless (Meade & Craig, 2012). This has led some authors to recommend using several methods in conjunction to detect careless responding (e.g., Curran, 2016; Meade & Craig, 2012; Niessen, Meijer, & Tendeiro, 2016). Authors have broken these methods into several categories depending on the nature of the method and the way carelessness is indicated (Curran, 2016; DeSimone, Harms, & DeSimone, 2015; Huang et al., 2012; Meade & Craig, 2012).

### Direct Methods

Direct or proactive screening methods require special items to be inserted into a survey prior to administration (DeSimone et al., 2015; Dunn et al., 2018; Meade & Ward, 2012). These items, often referred to as attention check items, are specifically designed to “catch” those who are not responding in a careful manner. That is, these items are designed such that a specific response indicates to the researcher that the participant is paying attention (Dunn et al., 2018). Examples are provided below. However, because of their design, the purpose of these items may be obvious to participants (DeSimone et al., 2015). There is risk that participants who are not motivated to respond accurately to the survey but hope to avoid being identified as careless responders will specifically look for these items and answer them carefully, despite answering the substantive items of the survey inaccurately (Curran, 2016).

**Directed response indexes.** A particularly common strategy for detecting careless responses involves the use of directed response or instructed items (Dunn et al., 2018). This type

of item tells the respondent exactly how to respond to the question. They are often relatively straightforward, e.g., “select ‘somewhat agree’ for this question.” However, some variations on this type of item couch response directions in larger sets of instructions, in an attempt to require a greater degree of attention in order to answer the item correctly. For example, a study may include a paragraph worth of instructions with the last line instructing the participant to disregard the previous instructions and select a specific response or to skip the question entirely (Curran, 2016). Directed response items are often very transparent, in that it is clear to participants that these questions are not used to measure any substantive variables (Curran, 2016). For this reason, some participants may view these items as condescending (Kraiger et al., 2019), particularly when the participants are working adults.

When directed response items are used in a questionnaire, multiple questions are dispersed throughout the survey. By including multiple items, researchers can create an index score measuring the number of directed response items the participant got correct or incorrect (Dunn et al., 2018; Huang et al., 2012). The index score then serves as a measure of the extent to which the participant attended to the survey. A cutoff score is typically used to remove participants who did not get a particular number of the directed response items correct (McGonagle et al., 2016). This is sometimes an arbitrary number of incorrectly answered items, but in other cases the number may be chosen to eliminate a preset proportion of the least attentive participants in the sample (e.g., the lowest 10% of scores on these items). There are not theoretically- or empirically-derived cutoff values that are widely accepted by researchers or implemented across studies (DeSimone et al., 2015), although Curran (2016) recommends that participants be required to answer at least 50% of these items correctly in order to be included in the sample.

**Infrequency indexes.** Rather than using directed response items, a similar technique involves the use of items for which virtually all respondents should provide the same response. On these items, a participant is assumed to be carelessly responding if they do not select the common response (Huang, Bowling, Liu, & Li, 2015). These types of items are referred to as infrequency items (Huang, Bowling, et al., 2015), bogus items (Meade & Craig, 2012), or catch questions (Dunn et al., 2018).

Infrequency items are somewhat less transparent than directed response items, in that they may have some face validity at first glance (Curran, 2016). However, their absurdity may vary. Some items may not clearly stand out as bizarre (e.g., being asked if “I am currently using a computer” on an online survey; Meade & Craig, 2012), while others are obviously ridiculous (e.g., “All my friends are aliens;” Meade & Craig, 2012).

Infrequency indexes are typically created in the same manner as directed response indexes; infrequency items are dispersed throughout a survey, and the number of correctly- or incorrectly-answered items serves as the index score (Dunn et al., 2018; Huang, Bowling, et al., 2015). Cutoff scores on infrequency indexes can be used as criteria to flag and remove participant responses before data analysis (McGonagle et al., 2018), although no widely accepted cutoff standards have been adopted (DeSimone et al., 2015). However, Curran et al. (2016) again recommended that participants be required to answer at least 50% of these items correctly in order to be included for analysis.

**Self-reported indicators of data quality.** Perhaps the most straightforward method for determining if a participant has paid careful attention during a survey is to directly ask them (DeSimone et al., 2015). Self-report items are placed at the end of a survey and ask the participant about the quality of the response they provided (Meade & Craig, 2012). These items

typically ask the participant to report how closely they paid attention, how much effort they exerted, how diligent they were in completing the survey, or how engaged they were with the survey (Meade & Craig, 2012; Wertheimer, 2017). Other metrics ask participants to make a judgement call regarding whether their data should be included for analysis (Meade & Craig, 2018).

Self-reported data quality items have been used as both single- and multi-item measures of careless responding (Meade & Craig, 2012). In the case of single-item indexes, a response that indicates the participant did not provide high-quality data can be used as an exclusion criterion. In the case of multi-item metrics, scale scores can be created to measure overall quality, attention, or effort. Minimum scale scores can be used a criterion for inclusion for data analysis.

### **Post-Hoc Methods**

Post-hoc methods for detecting careless responding involve the examination of response behavior over the course of a survey (DeSimone et al., 2015). These methods are also referred to as reactive (Dunn et al., 2018), archival, or statistical methods (DeSimone et al., 2015). Unlike direct methods, post-hoc methods do not require special attention check items to be inserted into the survey (Meade & Craig, 2012). Rather, they look at participant behaviors and response patterns to substantive items to determine if a participant has responded to the survey in a careful manner. Although these methods may be planned for in advance, they do not necessarily need to be. In most instances, the post-hoc indices can be calculated from participant responses and meta-data collected as part of a typical survey (Dunn et al., 2018). The degree of statistical sophistication required to calculate post-hoc methods varies, with some methods being very straightforward (e.g., total response time) and others requiring greater statistical expertise to



calculate and interpret (e.g., Mahalanobis distance, a statistic that captures the extent to which a response represents a multivariate outlier in the data).

**Response time.** Response time is the total amount of time a participant takes to complete a survey. It is assumed that very short response times are associated with careless responding, because participants are not reading or processing items before answering (Huang et al., 2012). Response time may be the most commonly used metric for eliminating careless responders, because it is typically automatically recorded, easy to understand, and easy to interpret (Curran, 2016). Response time serves as an intuitive metric for catching careless responders that are motivated to finish a survey as quickly as possible.

However, like with other metrics, there are not broadly accepted cutoff values that indicate a participant did not provide careful responses. Response times are related to the number of items on a survey and also to the length, clarity, and content of the instructions and items (Curran, 2016). Further, response time data can be skewed by participants who take long breaks mid survey or who do not immediately submit a survey upon completion (Meade & Craig, 2012). A long survey time may indicate that a participant was being diligent and taking breaks to prevent errors due to inattention, but on the other hand, a long response time may also indicate a participant has sped through a survey but hopes to avoid being detected as a careless responder. As a result, some researchers have used average per-page response times as the careless responding metric (Meade & Craig, 2012), because it is less sensitive to issues caused by leaving a single survey page open for an extended period of time.

Because response times are expected to vary across surveys, populations, and specific participants, it is particularly difficult to establish universal cutoff values for indicating careless responses. Huang et al. (2012) suggest a two-second per item benchmark to serve as a minimum

response time for inclusion. However, this cutoff is somewhat conservative and may err on the side of leaving careless responders in the sample (Curran, 2016).

**Longstring index.** Longstring indexes, or invariant response indexes, typically count the maximum number of times in a row a participant provides the same response (DeSimone et al., 2015). A longstring index specifically tries to identify careless responders who make use of a particular response pattern, straight line responding which involves selecting the same response repeatedly (Niessen et al., 2016). This metric tries to capture overly consistent responses, unlike other metrics (e.g., synonym/antonym metrics, intra-individual response variability, as described below) which try to identify random or inconsistent responses.

Longstring scores should be dependent on the number of items in a survey, the number of responses options, and the homogeneity of items placed near one another in the survey (DeSimone et al., 2015). For this reason, cutoff score recommendations vary widely, and in some instances, recommendations have depended on the specific response option being selected because some responses are more prevalent than others (Huang et al., 2012). Curran (2016) recommended a cutoff value equal to half the number of items in a scale but notes that this may not be appropriate for scales with very homo- or heterogeneous items, as larger or smaller values may be more suitable in those cases due to the probability (or improbability) of careful responders selecting the same response repeatedly.

**Intra-individual response variability.** Similar to the longstring index, intra-individual response variability tries to identify careless responders who answer items in a manner that is more consistent than we would expect (Dunn et al., 2018). However, rather than looking at the longest string of responses, this index considers the standard deviation of a participant's responses to a set of items. It rests on the assumption that an individual's responses should be

varied across surveys that measure multiple constructs, and as such the standard deviation of their responses should be somewhat large (Dunn et al., 2018). Because participants exhibiting long strings of responses will have low standard deviations, this metric will identify those participants while also identifying participants who have low variability in their overall responses, but do not respond with long consecutive strings of the same response (Dunn et al., 2018). This method might particularly diagnostic when participants exhibit long strings of similar responses on item sets that include regular and reversed-scored items.

The creators of this index, Dunn et al. (2018), do not provide cutoff values that would indicate a level of intra-individual consistency that suggests careless responding. However, they do recommend flagging the 10% of participants with the greatest degree of response consistency, as this value is similar to the expected rates of careless respondents in a typical sample.

**Synonyms and antonym indexes.** Unlike longstring and intra-individual response variability indexes which screen for suspiciously high consistency, synonym and antonym indexes indicate participants with low consistency. Synonym metrics examine the extent to which participants respond in a similar manner to similar items, and antonym metrics examine the extent to which participants respond in a dissimilar way to dissimilar items (DeSimone et al., 2015). This type of metric has also been referred to as an inconsistency approach (Huang et al., 2012), although I avoid that phrase because other types of post-hoc indices I will discuss also examine inconsistency in participant responses.

Synonym and antonym pairs can be identified by either their semantic or psychometric properties (Curran, 2016). Semantic synonyms are items that on their face should elicit similar responses, because synonymous language is used in the items (e.g., “I am sad” and “I am unhappy”). Semantic antonyms express linguistically opposite concepts (e.g., “I am sad” and “I

am happy”; DeSimone et al., 2015). The ability to take advantage of semantic synonyms or antonyms post-hoc requires that they were present in the measures used in the initial survey. However, researchers can use semantic pairs as a planned direct measure by purposefully inserting similar or dissimilar items into a survey even if they are not necessary for the measurement of substantive variables. A subset of survey items can also be repeated to create perfectly matched item pairs (Huang et al., 2012).

Synonyms and antonyms can also be identified psychometrically after data collection is complete. Psychometric pairs are identified not based on their content but by strong positive (in the case of synonyms) or negative (in the case of antonyms) correlations between item pairs (Dunn et al., 2018). Meade and Craig recommended considering item pairs with correlations above .60 to be synonyms and pairs with correlations below -.60 to be antonyms.

Once synonym or antonym pairs are identified, they can be used to calculate within-person correlations to serve as the careless responding metric. This is done by creating two vectors of data for each participant, one vector for the first component of each pair and another vector for the second. The number of pairs included can vary, but the magnitude of the intra-individual correlation should depend on the extent to which the pairs are true synonym or antonym pairs (Curran, 2016). Relatively weak intra-individual correlations indicate that a person’s synonym or antonym pairs were not strongly related, and it is assumed that this is because the individual did not respond carefully to the survey (Curran, 2016). Curran (2016) suggested that a negative intra-person correlation on a synonym index or a positive intra-person correlation on an antonym index indicates blatant careless responding, and therefore it should serve as a conservative criterion for removing responses.

**Individual reliability indexes.** Like synonym and antonym indexes, individual reliability indexes capture the extent to which a respondent answers items in an inconsistent manner (Curran, 2016). The most commonly discussed measure of individual reliability is even-odd consistency. This metric involves splitting each scale in a survey into two halves consisting of the even and odd items. Scale scores are calculated for each half of all scales, and then a sample-size-corrected correlation is calculated between the even and odd scale halves for each person (Meade & Craig, 2012; Niessen et al., 2016). A strong within-person correlation indicates that a person responded in a reliable manner, and thus, they are assumed to have responded carefully (Dunn et al., 2018).

Curran (2016) proposed a more sophisticated version of the individual reliability analysis called resampled individual reliability. Rather than dividing scales by even and odd numbered items, this bootstrap method repeatedly samples the item pool to split the items into two halves. By doing so, it removes any quirks associated with an individual split of the items (e.g., randomly ending up with more items with high means in one half and low means in the other; Curran, 2016).

**Mahalanobis D.** Mahalanobis distance or Mahalanobis D captures the extent to which a participant's responses represent an outlier in the data (Curran, 2016; Niessen et al., 2016). Rather than doing this based on an individual item or scale score, it does this on a multivariate basis by taking a set of items into account. That is, the Mahalanobis D metric considers the improbability of the participant's group of answers (Curran, 2016).

For example, consider a dataset with two variables, participants' heights and weights. On a univariate basis, a person who is 6'4" would likely not be considered an outlier, since that is a somewhat common height. Similarly, a person who weighs 100 pounds would likely not be

flagged as an outlier, because that is a relatively common weight. However, if the same individual reports a height of 6'4" and a weight of 100 pounds, they would likely be flagged as an outlier based on the Mahalanobis D metric. On the basis of their entire set of responses, the participant is an outlier. The improbability of these values suggest that they are unlikely to be accurate responses to the height and weight questions.

Statistically, the Mahalanobis D is calculated by determining the distance between a person's item scores and the sample means while taking the correlational structure between items into account. The distance value is largest when the participant's vector of responses is very different from the vector of item means (Niessen et al., 2016). The D value can be calculated for an individual set of items like a scale in a survey or it can be calculated for the entire questionnaire (Meade & Craig, 2012).

Unlike other indexes that measure careless responding, the probability distribution for Mahalanobis D is known; we can obtain *p*-values to describe the probability of an individual obtaining any particular Mahalanobis D value (Niessen et al., 2016). Therefore, a critical *p*-value can be set for determining if a response is improbable enough to be removed from the dataset.

### **Causes of Careless Responding**

While careless responders use a variety of response patterns, they share a common underlying issue, low motivation to respond accurately to survey questions (Rios et al., 2017). Unlike other forms of content nonresponsive survey inaccuracy (i.e., linguistic incompetence), careless responding occurs as a result of low motivation rather than low ability. However, the low motivation associated with careless responding may be caused by a number of different factors associated with the individual participant, the survey itself, the surrounding contextual factors, or some interaction between them.

## Survey Factors

Certain elements of a survey may reduce or drain participant motivation, and thus, increase the frequency of careless responding. These factors include overall survey elements like length and nature of the instructions, as well as factors related to the individual items in the survey.

**Survey length.** Survey length is a frequently-cited cause of careless responding (Brower, 2018; Gibson, 2016; Meade & Craig, 2012). It is assumed that over the course of a lengthy survey participant motivation will wane, and therefore careless responding behavior will increase (Meade & Craig, 2012). This is consistent with previous research that has found that participants are more likely to self-report careless responding towards the end of a survey (Berry et al., 1992), and that participants provide quicker responses with less variation towards the end of a questionnaire (Brower, 2018; Galesic & Bosnjak, 2009).

However, not all studies have found uniform effects for survey length. Gibson (2016) found no differences in rates of careless responding between students in short and long survey conditions. In part, this may be because the effect of survey length presumes that participants are initially motivated, but motivation decreases over the course of the study. However, if participants are so highly motivated that motivation does not drop below the threshold necessary for careless responding, length may not have an observable effect on behavior. Conversely, if participants start with motivation that is low enough that they do not intend to respond carefully from the beginning, no effects of survey length on careless responding will be observed. Thus, this cause of careless responding may only be relevant for those who start a study with moderate motivation but cannot maintain carefully responding throughout a lengthy study.

**Reverse-coded items.** Reverse or negatively coded items, also referred to as reverse worded or reverse keyed items, are worded such that strongly endorsed items indicate low levels of a construct (Kam & Meyer, 2015). While reverse coded items are typically assumed to measure the same construct as items that are worded in a straightforward manner, reviews of the construct development literature reveal that they are often methodologically problematic (Caught, Shadur, & Rodwell, 2000). The inclusion of reverse-code items often leads to poor measure reliability and factor structures where these items load poorly onto their intended factors or load onto a separate factor altogether (Weijters & Baumgartner, 2012; Woods, 2006).

Some authors have suggested that the reason for the poor performance of reverse-coded items may, in part, be due to careless responding (Kam & Meyer, 2015; Weijters & Baumgartner, 2012; Woods, 2006). If participants are reading items carelessly (or not at all) and miss information indicating the negative nature of the item, they are likely to select response options on the opposite end of the scale than they intend. Woods (2006) found that careless responding at rates of at least 10% can lead to incorrect factor assumptions, and Kam and Meyer (2015) found that careless responding contributed to the conclusion that job satisfaction and dissatisfaction were separate constructs.

Like survey length, the effect of negative items on careless responding behavior may be particularly relevant to participants who are somewhat but not strongly motivated to respond carefully. These participants may generally try to read and respond to items accurately, but unexpected switches in the direction of the questions being asked may go undetected. Conversely, very motivated respondents are likely to read and respond to all items carefully, and very unmotivated respondents aren't likely to respond to an item accurately regardless of its direction or consistency with the rest of the survey items. To combat careless responding on



reverse-coded items, Weijters and Baumgartner (2012) recommended alerting respondents to the presence of reverse-coded items in the instructions and using scales that have similar numbers of regular and reverse-coded items to avoid participant assumptions that all items are in worded same direction.

**Warning messages.** Some surveys have attempted to incorporate explicit messages or warnings into the directions in order to motivate participants to respond carefully. Some studies have employed warnings that simply inform the participant that their responses will be statistically evaluated for low quality (Huang, Bowling, et al., 2015; Ward & Pond, 2015), while some warn participants that they will not be compensated for their participation if response quality is low (Huang et al., 2012). Additionally, at least one study provided messages informing participants of additional incentives for careful responders (Gibson, 2016).

These studies have generally found that participants who were warned about careful responding did, in fact, respond more carefully according to some careless-response metrics (Gibson, 2016; Huang et al., 2012; Huang, Bowling, et al., 2015; Ward & Pond, 2015). It appears that warnings were effective for careless responding indicators that are very straightforward, e.g., directed responses (Ward & Pond, 2015), response time (Ward & Pond, 2015), semantic antonyms (Huang et al., 2012), and maximum longstring (Huang, Bowling, et al., 2015; Ward & Pond, 2015). However, they were not effective for less straightforward metrics like even-odd consistency (Ward & Pond, 2015) or Mahalanobis D (Ward & Pond, 2015)

### **Individual Factors**

While survey factors may increase the probability of careless responding, individual factors provide information regarding which individuals may be subject to low motivation to

respond carefully. To the extent that these individual factors vary across sampling contexts, some samples may be prone to higher rates of careless responding.

**Interest in topic.** Participant interest in a survey is an important driver of response behavior (Meade & Craig, 2012). Research has found that participants' interest in survey content is negatively related to careless responding behavior (Brower, 2018). Interest may affect motivation because those who find the study content intriguing are driven to pay attention to items and respond accurately. However, participants may also be interested in a study in a way that leads them to attend to items without necessarily being motivated to provide accurate responses. This is often the case when participants are interested in obtaining a particular outcome on the basis of their responses, but accuracy is either unrelated to the outcome (as is the case when the participant is only interested in completing the survey in order to be compensated) or could cost them the outcome (as is the case when the participant is interested in being hired or promoted). In cases in which participants are interested in obtaining incentives but incentives are not distributed on the basis of accuracy, careless responding is a concern. In the cases in which honesty could cost participants outcomes, faking may be a concern (Meade & Craig, 2012). Low respondent interest may be particularly likely to result in careless responding when survey participation is mandatory or when participants are compensated or otherwise pressured to participate (Meade & Craig, 2012).

Recent research has begun to investigate the relationship between respondent interest in a survey and other survey factors. Brower (2018) found that interest moderated the relationship between survey length and careless responding, such that those with low interest were more likely to carelessly respond on long rather than short surveys. Ward and Meade (2018) tested interventions to reduce careless responding and found that by inducing cognitive dissonance they

were able to increase participant interest and reduce careless responding behavior. Thus, it appears that low interest does contribute to careless responding behavior, and that by manipulating participant interest, it may be possible to reduce careless responding. McGonagle et al. (2016) recommended attempting to pique participant interest by providing information up front about the survey topic and the intended use of the data.

**Attitudes towards surveys.** In addition to interest in the specific topic being studied, participant interest in taking surveys in general may also play a role in determining response quality. Attitudes towards a behavior have been shown to be an important predictor of whether an individual will display that behavior (Armitage & Conner, 2001), consistent with popular theories of reasoned action (Ajzen & Fishbein, 1980) and planned behavior (Ajzen, 1991). Thus, attitudes towards survey responding may predict survey response behaviors.

Rogelberg et al. (2001) found that participants' attitudes towards surveys impacted several outcomes related to careless responding, including following directions and responding to all items. Similarly, Brower (2018) found that their participant interest scale, which included items related to general interest in survey participation, was related to careless responding behavior. Rogelberg et al. (2001) concluded that researchers should promote positive participant attitudes when possible, and that they should even attempt to measure and control for participant attitudes in study analyses.

**Personality.** Recent research has made an effort to focus on the enduring individual differences that may cause some respondents to take less care when completing surveys (e.g., Bowling et al., 2016; Furnham, Hyde, & Trickey, 2015; Ward et al., 2017). Studies have found that those who display careless responding behavior in one study are more likely to display careless behavior on later tasks (Bowling et al., 2016; Camus, 2016). This suggests that careless

behavior is consistent over time and likely stems, and least in part, from stable individual differences.

The five-factor model personality traits – conscientiousness, agreeableness, extraversion, openness to experience, and emotional stability (McCrae & Costa, 1987) – has served as a popular framework for investigating the relationships between personality traits and careless responding behavior. Research has found that four of the five traits – conscientiousness, agreeableness, extraversion, and emotional stability – are related to careless responding behavior (Bowling et al., 2016; Dunn et al., 2018; Ward et al., 2017). Lower scores on each of the four traits is associated with more careless responding behavior. While the negative relationships between conscientiousness and agreeableness with careless responding were predicted based on theory, relationships between extraversion and emotional stability were examined only as exploratory research questions (Bowling et al., 2016; Ward et al., 2017). It is unclear what underlying mechanisms explain those relationships. However, other research has failed to replicate these personality-careless responding behavior relationships (Camus, 2015).

Additional personality differences may also be related to careless responding. Furnham et al. (2015) examined Hogan Personality Inventory traits and found that agreeable, stable, prudent, and ambitious individuals were less likely to respond carelessly, while excitable, cautious, skeptical, and reserved individuals were more likely to carelessly respond. Dunn et al. (2018) also found that an individual's tendency to become bored was related to greater careless responding.

These findings are consistent with broad conclusions that personality traits, especially conscientiousness and to a lesser extent emotional stability, are mechanisms that support motivation by increasing an individual's drive and self-regulatory capability (Diefendorff &

Chandler, 2011; Judge & Ilies, 2002; Kanfer, Frese, & Johnson, 2017). Other personality traits may also be important because of their specific effects on motivation to respond carefully. For example, agreeableness may be important because agreeable individuals are characterized by a desire to help and please others. In the context of completing a survey, agreeable individuals may be driven by their desire to help to respond in a way that is accurate and useful (Dunn et al., 2018).

Because careless responding is related to personality differences in participants, this presents a challenge for researchers conducting survey research. If researchers hope to gather data on participant personality or on attitudes or behaviors influenced by personality, the removal of careless responses may restrict the range of personality present in their sample (Bowling et al., 2016). On the other hand, leaving careless responses in the sample may inappropriately inflate or deflate relationships between variables (McGonagle et al., 2016). This dilemma led Bowling et al. (2016) to conclude that it is preferable to encourage careful behavior through manipulation of survey and contextual factors, rather than to allow for high rates of careless responding and to remove these participants post-hoc.

**Lack of self-control.** Personality, especially conscientiousness, is likely related to motivation because of its association with regulatory processes (Judge & Ilies, 2002). When an individual would like to respond carefully but lacks the self-control to do so, their ability to respond carefully will be compromised. Some research has examined regulatory processes directly to determine their relationship with careless responding behavior (e.g., Barber, Barnes, & Carlson, 2013).

Barber et al. (2013) conducted research to examine the relationship between careless responding, effort exerted on a survey, self-control, and a known cause of self-control depletion

– a lack of sleep. The study found that those who reported a decreased capacity to exert self-control were more likely to respond carelessly, and that this relationship was mediated by the amount of energy the respondent exerted on the survey.

To the extent that survey participants lack self-control, either because of situational factors like sleep deprivation or dispositional factors like low conscientiousness, they may be unable to exert the sustained effort necessary to respond carefully and accurately to a survey. This may serve as a widespread issue for research that samples respondents from populations with low self-control due to their current state (e.g., sleep deprived) or dispositional traits (e.g., low conscientiousness).

### **Contextual Factors**

In addition to survey or participant characteristics, elements of the environment in which a survey is taken may affect individuals' motivation to provide careful, accurate responses.

**Social norms and influence.** Meade and Craig (2012) suggested that participants' perceptions of their social environment may influence the extent to which they answer survey questions carefully. Social norms around appropriate survey behavior may affect the extent to which participants feel compelled to respond carefully and accurately (Meade & Craig, 2012). Some have posited that social norms are likely activated by the extent to which survey respondents interact with those conducting the survey (Ward & Meade, 2018; Ward & Pond, 2015).

However, tests of effects of social influence have been mixed. In one research study, Ward and Pond (2015) included virtual human figures that watched online survey participants while completing a questionnaire. They found that the presence of virtual humans did, in fact, reduce careless responding in conditions in which participants had been warned to respond

carefully. On the other hand, a research study by Ward and Meade (2018) found that including video or written messages from researchers thanking participants for their time and explaining the importance of the research at the beginning of the study did not reduce careless responding. The same study also failed to find reductions in careless responding under conditions that invoked reciprocity norms and perceptions of group gain, two factors that should increase motivation under social exchange theory. The authors suggested that these manipulations may not have seemed authentic to participant, and thus, they lacked the desired effect. More salient manipulations of social influence may need to be used in order to invoke social norms and reduce careless responding (Ward & Meade, 2018).

**Environmental distraction.** Environmental distractions are the elements of a respondent's surroundings that may impede their ability to provide careful responses. This may include any factor that distracts the participant, for example, distractions from other individuals who are present or distractions from television, radio, and other media playing in the background. Any study that allows participants to complete questionnaires in uncontrolled environments (e.g., online survey, mail survey) may be affected by participants attempting to multitask while answering survey items (Meade & Craig, 2012).

Participants using MTurk and other crowdsourcing platforms may be particularly vulnerable to environmental distractions. For example, some MTurk participants may be motivated to complete multiple surveys at once using multiple accounts (Cheung et al., 2017). By switching back and forth between surveys, participants may avoid detection due to short response times. In these cases, participants may be compromised by distraction due to multitasking or frequent task switching. Unfortunately, little research has specifically addressed

the effects of environmental distraction on careless responding behavior or evaluated interventions to prevent distractions (Meade & Craig, 2012).

Though there are a variety of reasons why a person may fail to respond carefully to a survey, their carelessness may not be inevitable. In the next section, I explore how motivational theory may be applied to influence respondents to pay more careful attention over the course of a survey.

### **Applying Motivational Theory to Careless Responding**

By definition, careless responding occurs when participants are not motivated to provide careful responses on surveys (Rios et al., 2017). While other issues also contribute to inaccurate responding (e.g., poor ability read and comprehend questions), these issues are not motivational in nature, and thus are outside the domain of careless responding. Because careless responding is a motivational issue, motivational theory can shed light on the possible behavioral antecedents that may influence motivation to respond carefully, and ultimately, the careful responding behavior itself.

Motivational theory is quite complicated, in part because of the way researchers have used and defined the concept of motivation. In a recent review paper, Kanfer et al. (2017) noted that “motivation is frequently used as an umbrella term meant to capture the dense network of concepts and their interrelations that underlie observable changes in the initiation, direction, intensity, and persistence of voluntary action” (pp. 339). While motivation underlies observable changes, Diefendorff and Chandler (2011) noted that motivation itself is often described as “an unobservable force” (pp. 66).

While motivational forces may not be observable, the behaviors they act on are. Ultimately, motivation determines which behaviors an individual intends to perform, the amount



of effort they put into performance of those behaviors, and the amount of time they are willing to invest in those behaviors (Diefendorff & Chandler, 2011). However, motivation is not the only factor that affects behavioral performance; motivation interacts with abilities and situational constraints that may limit the extent to which an individual is capable of performing the behavior (Schmitt et al., 2003).

Motivation itself is influenced by a variety of individual and situational variables. Correspondingly, a great number of theories have been proposed to explain how these variables affect motivation and performance (Diefendorff & Chandler, 2011; Kanfer et al., 2017). Kanfer et al. categorized these into three general types of theories: content-based approaches, context-based approaches, and process-based approaches. Content-based approaches focus on the needs and desires that inspire behavior, context-based approaches focus on the ways environmental factors affect motivation, and process-based approaches focus on the psychological mechanisms that drive motivation to perform behaviors.

For the purpose of developing interventions to reduce careless responding behavior, some motivational theories may be more applicable than others. Content-based approaches that focus on individual needs and traits may have limited utility within an intervention. That is, while characteristics like need for achievement or conscientiousness may explain careless responding behavior, personality characteristics are relatively stable (Sackett, Lievens, Van Iddekinge, & Kuncel, 2017), and thus may prove relatively difficult to affect via intervention. Similarly, context-dependent approaches may also have limitations when applied to careless responding interventions. While context may affect a person's motivation to respond carefully, it may be difficult to control and manipulate as part of a survey-based research study, particularly one in which participants work in a variety of contexts and responding remotely at their leisure.

Interventions based on contextual approaches may be better suited for survey research conducted in closely controlled laboratory environments.

Process-based theories may be best suited to affecting careless responding behavior. This type of theory describes the mechanisms that produce and sustain motivation towards particular behavioral goals (Kanfer et al., 2017). Popular process theories include expectancy theory (Vroom, 1964), the theory of planned behavior (Ajzen, 1991; Ajzen & Fishbein, 1980), and goal-setting theory (Locke & Latham, 1990).

In particular, the theory of planned behavior may be well-situated to explain motivation to respond carefully and ultimately inform interventions. While expectancy theory, goal-setting theory, and the theory of planned behavior all have empirical support for their general tenants, expectancy theory and goal-setting theory may be less directly relevant to careless responding behavior. For instance, goal-setting theory tends to work best with committed individuals who are focused on challenging goals (Diefendorff & Chandler, 2011). On low-stakes organizational surveys, careful responding may not constitute a challenging or commitment-inspiring endeavor. Similarly, expectancy theory emphasizes the rational, calculating approach individuals take to allocating effort towards goals. For low-stakes goals like careful responding, intentions are unlikely to be made in such a mechanistic manner. The theory of planned behavior takes into account the cognitive biases and social factors that shape motivation towards a wide range of behaviors (Kanfer et al., 2017). For this reason, I focused on the theory of planned behavior as the driving framework behind the careless responding interventions in the present study.

### **The Theory of Planned Behavior**

The theory of planned behavior (Ajzen, 1991) is a popular theory for explaining human behavior. The theory of planned behavior describes the relationships between individual

behavior and its antecedents. The theory posits that three factors – attitudes toward a behavior, subjective norms, and perceived behavioral control – lead to behavioral intentions, and that in turn, behavioral intentions lead to the performance of the behavior (Ajzen, 1991).

Since the initial development of the theory and its precursors (e.g., the theory of reasoned action; Ajzen & Fishbein, 1980), the theory's authors have attempted to broaden the theory by incorporating additional constructs shown by empirical and theoretical research to be related to behavior. These include variables that predict behavioral control (e.g., self-efficacy, locus of control; Ajzen, 2002), the behavior itself (e.g., prerequisite knowledge, behavioral salience, environment constraints, habit; Fishbein & Yzer, 2003; Kasprzyk, Montaña, & Fishbein, 1998), or other factors that affect the entirety of the model (e.g., demographic variables, personality traits, culture; Fishbein & Yzer, 2003; Montaña & Kasprzyk, 2015). Fishbein and colleagues (Fishbein & Yzer, 2003; Kasprzyk et al., 1998; Montaña & Kasprzyk, 2015) refer to this extended theory as the integrated behavioral model.

The theory of planned behavior has been supported by considerable empirical research showing its ability to explain a variety of human behaviors; several meta-analyses have found support for the relationships in the model (Albarracín, Johnson, Fishbein, & Muellerleile, 2001; Armitage & Connor, 2001; Cooke & French, 2008; Hagger, Chatzisarantis, & Biddle, 2002). Further, the model has shown utility as both an explanatory mechanism and as a basis for interventions to change behavior (Fishbein & Yzer, 2003). When applied to work contexts specifically, the theory of planned behavior has been successful at predicting motivation to perform a variety of work-related behavior (Kanfer et al., 2017).

## **Present Study**

Because the theory of planned behavior and its derivatives have been successful at explaining and intervening to change a variety of behaviors, it may be that the theory can also inform research on careless responding. The theory may be particularly relevant in light of research that has proposed relationships between careless responding behavior and theory of planned behavior antecedents, e.g., attitudes as proposed by Rogelberg et al. (2001), social norms as proposed by Ward and Mead (2018) and Ward and Pond (2015), and self-control as proposed by Barber et al. (2013).

Accordingly, I conducted an intervention study to test the effects of the predictors in the theory of planned behavior on careless response intentions and careless responding behavior. The first intervention condition was designed to improve attitudes towards careful responding, the second was designed to improve perceived norms around careful responding, and the final condition was designed to improve perceived control over careful responding.

This study uses a sample of MTurk participants employed in a variety of organizations and job functions. Research suggests that organizational, MTurk, student, and community samples all display similar rates of careless responding behavior (Behrend et al., 2011; Kees, Berry, Burton, & Sheehan, 2017; Kraiger et al., 2019; Necka, Cacioppo, Norman, & Cacioppo, 2016). However, in all types of samples, participants may be motivated to complete surveys quickly and carelessly so that they can move on to other tasks, whether those tasks are other surveys (as on MTurk) or their regular work tasks (as in an organizational context). Therefore, I suggest that the motivation to respond carefully should be affected by the same mechanisms across samples. The effect of the interventions should not depend on the method of participant recruitment.

Though concerns about differences in careless responding rates across samples may not be warranted, recent research has been particularly focused on careless responding behavior in MTurk and other crowdsourced samples (e.g., Behrend et al., 2011; Chung et al., 2017). Because MTurk data collection is becoming more common in organizational psychology and other social sciences (Cheung et al., 2017; Michel, O'Neill, Hartman, & Lorys, 2017), it is important to provide avenues for ensuring participant responses are as trustworthy as possible. This study aimed to address that gap.

Ultimately, the purpose of this study was two-fold. First, I aimed to deepen the understanding of the theoretical basis for careless responding behavior. This study rests on the theoretical assertion that careless responding behavior is affected by individual motivation. If the interventions were successful, they would imply that it is possible to affect careless responding behavior by increasing individual motivation to respond accurately. Second, I aimed to provide practical methods that could be used for the prevention of careless responding behavior. If the interventions were successful, they could be applied to future research studies as a method for minimizing error due to careless responding.

### **Attitudes**

The first intervention condition was designed to improve attitudes towards careful responding behavior. An attitude is defined as a person's overall favorable or unfavorable evaluation of a target (Ajzen, 1991). Although targets can be a variety of different objects – individuals, ideas, behaviors, etc. – in the context of the theory of planned behavior, they typically refer to the behavior in question.

Modern theory generally considers attitudes to have both an affective or emotional component and a cognitive component (Banaji & Heiphetz, 2015). In the context of the theory of

planned behavior, an individual's emotional response to a behavior has been referred to as their experiential attitude. Their cognitive response is referred to as their instrumental attitude (Montano & Kasprzyk, 2015).

Attitudes stem from two sources. The first is an individual's belief that performing a behavior will lead to certain outcomes, and the second is the individual's evaluation of desirability of those outcomes (Fishbein & Yzer, 2003). Under the theory of planned behavior, if a person believes that a behavior will lead to positive outcomes, they are more likely to have a positive attitude towards it; if they believe it will lead to negative outcomes, they are less likely to have a positive attitude towards it. However, the individual's evaluation of the quality of the outcomes also affects attitudes. Assuming an individual believes a behavior will result in a positive outcome, they are more likely to have a positive attitude if they strongly desire the outcome as compared to if they only mildly desire it (Fishbein & Yzer, 2003).

### **Attitudes and Behavior**

There is a great deal of empirical evidence to support the notion that individuals' attitudes towards a behavior are related their performance of the behavior. Several meta analyses have found significant relationships between attitudes and behaviors (Cooke & Sheeran, 2004; Glasman & Albarracín, 2006; Harrison, Newman, & Roth, 2006; Kim & Hunter, 1993a). Further, the attitudes-behavior relationship appears to exist regardless of the target behavior (Cooke & Sheeran, 2004; Kim & Hunter, 1993a). Meta-analyses have found that this relationship exists for job behaviors (Harrison et al., 2006), including employee job performance (Judge, Thoresen, Bono, & Patton, 2001; Riketta, 2008) and turnover (Meyer, Stanley, Herscovitch, & Topolnytsky, 2002). The relationship also holds for health-related behaviors (Albarracín et al., 2001) and behaviors that are relatively new to the individual (Glasman & Albarracín, 2006).

Emerging research suggests that this relationship may also specifically apply to participation in research studies. A recent qualitative research study indicated that those who choose to participate in research have strong positive attitudes towards research participation, science in general, the organization conducting the research, and the group the research may ultimately benefit (Carrera, Brown, Brody, & Morello-Frosch, 2018).

Studies have also investigated the extent to which changes in attitudes lead to changes in behavior. For example, for health-related behaviors, meta-analyses have indicated that interventions that successfully change attitudes also change corresponding behaviors (Derzon & Lipsey, 2002; Sheeran et al., 2016). The effect of attitudes on behavior was small-to moderate (mean  $d = .38$ ; Sheeran et al., 2016).

As is described in the theory of planned behavior, empirical evidence suggests that relationships between attitudes and behavior tend to be mediated by behavioral intentions (Albarracín et al., 2001; Armitage & Connor, 2001; Cooke & French, 2008; Cooke & Sheeran, 2004; Kim & Hunter, 1993b). These findings also extend to intervention research where intentions are measured (Sheeran et al., 2016). However, the majority of intervention studies do not report the effects of the intervention on attitudes, behavioral intentions, or other process variables. Rather, the interventions are assessed solely on the basis of their ability to change behavior (Hardeman et al., 2002; Prestwich et al., 2014).

### **Attitudes Intervention**

In this study, the attitudes intervention condition was designed to reduce careless responding behavior by improving attitudes towards careful responding. In accordance with the theory of planned behavior, it was expected that by increasing attitudes towards careful responding behavior, performance of the behavior would be more likely (Ajzen, 1991). This

would be consistent with other intervention research that has found improving attitudes is effective for improving behavioral performance (Sheeran et al., 2016).

According to theory, attitudes are affected by the belief that performing a behavior will lead to positive outcomes and the extent to which the individual desires those outcomes (Fishbein & Yzer, 2003). Consistent with previous research suggesting compensation is the most common motivator for MTurk workers (Behrend et al., 2011; Paolacci, Chandler, & Ipeirotis, 2010), pilot testing of this study confirmed that compensation was the most common motivator for participants. Therefore, according to the theory of planned behavior, participants' attitudes towards careful responding behavior should depend on the extent to which they believe careful responding will lead to compensation and the extent to which they desire compensation.

In this study, all participants were aware of the pay rate before agreeing to participate. Therefore, it can be assumed that at least for the majority who are motivated by compensation, participants were motivated to earn the offered pay rate. It was expected that improving attitudes would lead to less careless behavior, because the attitudes intervention would reinforce in participants' minds that it is important to provide accurate responses in order to receive the agreed upon compensation. In the absence of any intervention, participants might be primed to hold positive attitudes towards careless responding, because they believe that careful responding will reduce their overall compensation rate by slowing down their performance. However, by reinforcing their belief that there is a direct relationship between careful responding and receiving the compensation they desire, I believed I could improve attitudes towards careful responding, and ultimately reduce rates of careless responding.

To address beliefs that behavioral performance will lead to compensation, the intervention explained that the researchers required careful and accurate responses. Further, it



reiterated that if a participant responded carefully, they would be compensated for their work. The materials explained that there are no right or wrong answers on the survey, and any answer was acceptable as long as it reflected the respondent's true feelings.

I expected that this information would increase participant's beliefs that they will only be compensated if they respond carefully, and that by doing so, participant attitudes towards careful responding would become more favorable. In accordance with the theory of planned behavior (Ajzen, 1991) and the integrated behavior model (Kasprzyk et al., 1998), I believed that more favorable attitudes would affect the extent to which participants intended to and actually did provide careful responses.

As is the case with most intervention studies (Hardeman et al., 2002; Prestwich et al., 2014), the primary outcome variable of interest was the performance of the behavior in question, careless responding behavior. However, this study also assessed the extent to which the interventions affected the intended attitude (i.e., attitudes towards careful responding behavior) as well as the extent to which it affected the targeted behavioral intention (i.e., intentions to respond carefully). It was expected that the intervention would be successful at improving attitudes, increasing intentions to respond carefully, and reducing careless responding behavior as compared to the control group.

**Hypothesis 1.1:** Those who received an intervention to improve attitudes towards careful responding have more favorable attitudes towards careful responding than those in the control group.

**Hypothesis 1.2:** Those who received an intervention to improve attitudes towards careful responding have stronger intentions to respond carefully than those in the control group.

**Hypothesis 1.3:** Those who received an intervention to improve attitudes towards careful responding are less likely to respond carelessly than in the control group.

### **Subjective Norms**

A second intervention condition targeted another antecedent of behavior, subjective norms. Subjective norms refer to the perceived social pressures an individual feels to perform or not perform a behavior (Ajzen, 1991). Subjective norms are a function of two factors, normative beliefs and motivation to comply. Normative beliefs describe whether the person in question believes referent individuals approve or disapprove of the behavior. Motivation to comply describes the extent to which the person believes it is important to comply with the behaviors others approve of (Montaño & Kasprzyk, 2015).

Normative beliefs can be broken down into two separate categories. Injunctive norms describe the social pressures to perform a behavior based on the individual's belief that it is what other people want them to do. Descriptive norms, on the other hand, describe the social pressure based on the observed or inferred behavior of referent others (Manning, 2009). Referent others are typically considered to be other individuals who occupy the same social position and roles as the person in question (Shah, 1998). Norms have a larger effect on behavior when they refer to behaviors and beliefs of an individual's in-group members. That is, people are more likely to conform to norms if they believe that others in their own social group do as well (Smith & Louis, 2008).

It is important to note that subjective norms may not always be accurate (Reno, Cialdini, & Kallgren, 1993). Presumably there is a relationship between what individuals believe about others' attitudes and behavior how others actually think, feel, and act, although the extent to which a person's subjective norms are accurate for the referent group may vary. Like other

beliefs, perceptions of norms may be influenced by cognitive biases like availability, primacy, and recency effects. In some studies, researchers have attempted to intervene to affect perceptions of social norms with the ultimate aim of affecting behavior (e.g., Cialdini, Reno, & Kallgren, 1990; Middlestadt et al., 1995; Reno et al., 1993).

### **Subjective Norms and Behavior**

Like attitudes, subjective norms have been shown to be related to behavior. Meta-analytic research has found that subjective norms about a behavior are related to performance of that behavior (Hausenblas, Carron, & Mack, 1997; Manning, 2009; Sheeran et al., 2016). Further, both perceived injunctive and descriptive norms were independently related to behavioral performance (Manning, 2009). Individual studies have shown that subjective norms also affect work-related behaviors, such as safety behavior on the job (Fogarty & Shaw, 2010) and job seeking while unemployed (Stutzer & Lalive, 2004).

Additionally, research has examined the extent to which interventions to change subjective norms affect behavior. A meta-analysis by Sheeran et al. (2016) found that interventions which are effective at changing subjective norms around health behaviors lead to small-to-moderate effects on behavior. This pattern held regardless of the individual health behavior in question. Other studies have found that subjective norms interventions are also related to behavioral performance for environmentally-friendly behaviors (e.g., Cialdini et al., 1990; Reno et al., 1993).

In addition to the direct effects of subjective norms on behavior, research has also examined the indirect effects of subjective norms on behavior via increasing behavioral intentions. Meta-analytic research has found support for this mediation relationship across a variety of behaviors (Armitage & Conner, 2001; Manning, 2009), including health-related

behaviors (Cooke & French, 2008; Sheeran et al., 2016), exercise behaviors (Hagger et al., 2002; Hausenblas et al., 1997), and sexual behaviors (Albarracín et al., 2001). Individual studies have found support for mediated relationship between norms about work-related behaviors and the behaviors themselves (Fogarty & Shaw, 2010; Stutzer & Lalive, 2004).

### **Proposed Subjective Norms Intervention**

The subjective norms intervention was designed to affect normative beliefs about careless responding behavior. It aimed to build the perception that referent individuals respond carefully to MTurk surveys and to increase the salience of that perception. The intervention targeted participants' subjective norms by addressing both descriptive and injunctive norms towards careless responding behavior. That is, the subjective norms intervention aimed to affect the extent to which participants believe that other MTurk workers respond with care and the extent to which those workers believe it is important to respond with care.

As described by the theory of planned behavior, it was expected that by affecting the strength and salience of subjective norms around careful responding, rates of careless behavior would decrease (Ajzen, 1991). As is, it is believed that many research participants assume that speedy and careless responding is common among their peers. This perception may be developed through interactions with fellow MTurk workers on online forums and chatrooms. Previous research has found MTurk participants believe that around a third of their fellow MTurk workers complete studies without attending to the instructions and that around a quarter answer questions without thinking about them (Necka et al., 2016). These rates were similar to rates found for student and community samples asked the same questions. To the extent that the intervention was affective at altering these beliefs, participants should have been less likely to engage in these

careless behaviors themselves. This hypothesis was consistent with previous research that found subjective norms interventions to be effective for changing behavior (Sheeran et al., 2016).

To target normative beliefs, the subjective norms intervention condition provided participants with information about referent individuals, in this case other MTurk survey respondents. The intervention described the high rates of careful responding among this group, in an effort to improve descriptive norms about how their peers respond to surveys. It also provided statistics about this group's strong belief that it is important for all MTurk workers to respond carefully in order to receive their agreed upon compensation. In doing so, the intervention also targeted the participants' injunctive beliefs about how others believe they should behave.

By providing this information to the participants, I believed the intervention would bolster the subjective norm that MTurk workers respond carefully to survey questions. As described by the theory of planned behavior (Ajzen, 1991) and the integrated behavior model (Kasprzyk et al., 1998), I believed that stronger subjective norms around careful responding would affect the extent to which participants intended to respond carefully and actually did provide careful responses.

The primary outcome of interest for the subjective norms intervention was behavioral performance, in this case the performance of careless responding behavior. However, I also assessed the extent to which the intervention affected subjective norms and intentions to respond carefully. It was expected that the intervention would be successful at affecting subjective norms, increasing intentions to respond carefully, and reducing careless responding behavior as compared to a control group.

**Hypothesis 2.1:** Those who received an intervention to strengthen subjective norms around careful responding have stronger subjective norms beliefs than those in the control group.

**Hypothesis 2.2:** Those who received an intervention to strengthen subjective norms around careful responding have stronger intentions to respond carefully than those in the control group.

**Hypothesis 2.3:** Those who received an intervention to strengthen subjective norms around careful responding were less likely to respond carelessly than those in the control group.

### **Perceived Behavioral Control**

The final intervention condition aimed to affect participant perceptions that they have control over whether they respond carefully. Perceived behavioral control is the perceived ease or difficulty of performing a behavior (Ajzen, 1991). It is comprised of two components, perceived capacity to perform the behavior and perceived autonomy over performing the behavior. Perceptions of control reflect participants' past experience with the behavior and the obstacles they anticipate encountering if they attempt the behavior again (Ajzen, 1991). Perceptions of control are ultimately determined by beliefs about the presence of facilitators and barriers to behavior and by beliefs about the extent to which these facilitators and barriers will affect the individual's ability to perform the behavior (Montaño & Kasprzyk, 2015). For example, participants may believe they are capable of providing careful responses, but they may believe that regardless of how they perform, researchers will automatically approve or reject their work on MTurk.

The conceptual definition of perceived behavioral control has been a topic of much discussion, in part because of its relationship with similar variables (Ajzen, 2002). Perceived behavioral control is related to several individual characteristics, including self-efficacy and locus of control (Ajzen, 2002). Self-efficacy describes an individual's beliefs about their capabilities to exercise control over their functioning (Bandura, 1993). Locus of control describes the extent to which an individual believes that outcomes in their lives are either external and caused by factors beyond their control or internal and self-determined (Rotter, 1966).

Self-efficacy and locus of control apply to a person's entire range of functioning, while perceived behavioral control is specific to a single behavior (Ajzen, 1991). However, perceived behavioral control for any particular behavior should be theoretically influenced by a person's general self-efficacy and locus of control (Ajzen, 2002). Over the last several decades, there has been some variation in the extent to which authors describe these individual characteristics as part of the perceived behavioral control concept (e.g., Ajzen, 2002; Yzer, 2012), as separate predictors of intentions and behaviors (e.g., Armitage & Connor, 1999), or as components of a larger "personal agency" construct (e.g., Montaña & Kasprzyk, 2015).

In addition to relationships with individual characteristics, perceived behavioral control is assumed to have some relationship with actual autonomy over behaviors. It is assumed that individuals consider the environmental constraints of their situation when conceptualizing their control (Ajzen, 2002). Regardless of an individual's perceived capacity to perform a behavior, if they do not believe they have the autonomy to decide how to perform, they will not perceive control over their behavior (Yzer, 2012).

## **Perceived Behavioral Control and Behavior**

Unlike attitudes and subjective norms, the theory of planned behavior specifies that perceived behavioral control should have some direct effect on behavior, in addition to its indirect effect through changes in behavioral intentions (Ajzen, 1991). This is because perceived behavioral control is related to actual behavioral control. If an individual does not have control over their own behavior, no degree of cognitive change can affect behavior. Even if control beliefs change and intentions improve, if actual behavioral control does not exist, an individual's behavior will remain constant. For this reason, the perceived behavioral control-behavioral intentions-behavior path is said to be moderated by actual control. That is, the path will explain a substantial portion of the variance in behavior under conditions of high actual control, but under conditions of low actual control it will not (Armitage & Connor, 1999; Ajzen, 1991).

There is a great deal of evidence to suggest that perceived behavioral control is related to behavior in most instances, both directly and by way of affecting behavioral intentions. Several meta-analyses have found a positive relationship between perceived behavioral control and performance of the behavior (Albarracín et al., 2001; Armitage & Connor, 2001; Cooke & French, 2008; Hagger et al., 2002; Hausenblas et al., 1997; Notani, 1998). These findings were supported in meta-analyses examining all types of behavior (Armitage & Connor, 2001; Notani, 1998) and specific individual behaviors, including health behaviors (Cooke & French, 2008), exercise behaviors (Hagger et al., 2002; Hausenblas et al., 1997), and sexual behaviors (Albarracín et al., 2001). Within organizational psychology, a related construct, self-efficacy, was found to predict work performance in several meta-analyses (Stajkovic & Luthans, 1998; Judge, Jackson, Shaw, Scott, & Rich, 2007).



Additionally, most meta-analyses support the theory of planned behavior's partial mediation hypothesis for perceived behavioral control. That is, they find that there is a significant direct path between perceived behavioral control and behavior, even when the indirect path through intentions is also included in the model (Armitage & Conner, 2001; Hagger et al., 2002; Notani, 1998). However, at least one meta-analysis found that the relationship between perceived behavioral control and behavior was fully mediated by intentions (Albarracín et al., 2001). That study focused on a single behavior – condom use. It may be the case that for that particular behavior, actual control is high, and thus environmental constraints should not derail the intention-performance relationship. Similarly, only an indirect relationship was found when the model was applied to safety behaviors in the workplace (Fogarty & Shaw, 2010).

Research has also attempted to determine the extent to which perceptions of control can be changed in order to produce changes in behavior. A recent meta-analysis found that interventions that successfully increased perceived behavioral control or self-efficacy led to moderate changes in health behavior (Sheeran et al., 2016). Individual studies have also examined the relationships between work-related behaviors and constructs related to perceived control. For example, interventions to increase self-efficacy led to more job search behaviors among unemployed individuals in several studies (Eden & Aviram, 1993; Vinokur, Price, & Schul, 1995).

### **Proposed Perceived Behavioral Control Intervention**

The final intervention condition in this study attempted to affect individuals' beliefs about their capacity to respond carefully and their control over whether their behavior is judged to be careful by the researchers. In other words, the intervention was targeted towards affecting

both beliefs about the presence of barriers and beliefs about how barriers would affect the individual's ability to achieve the outcomes associated with careful responding.

As described in the theory of planned behavior, it was expected that by affecting beliefs about control over careful responding behavior, rates of careless responding behavior would decrease (Ajzen, 2002). On MTurk, many researchers mention that they conduct attention checks and that failure will result in nonpayment, but they do not provide any information on what is required to pass the attention checks. Necka et al. (2016) noted that this may not actually be beneficial, in that it may have effects on participants' choices to seek or share information online about attention check item content, how to pass the items, and which MTurk requestors have lenient or severe requirements. In fact, when conducting research for a previous study, I was contacted by an MTurk forum moderator and advised that I should stop using stringent attention checks, because MTurk users on the forum were negatively rating my study. I was warned that this may result in fewer MTurk workers participating in my study going forward.

Ultimately, it appears that at least some participants try to affect their personal control by seeking out studies in which they believe their responses will certainly be judged as careful. This could be because they intend to be careful and want their responses to be recognized as such, or it could be because they intend to be careless and they do not want to be detected and refused payment.

In cases where no information is available online (e.g., newly posted surveys) or the MTurk worker does not use online forums, participants may choose to respond carelessly because they assume that there is no relationship between carefulness and payment. In the absence of contradictory information, they may assume they have no control over the outcomes of their behavior. By intervening to increase participants' perceptions of control, I expected to

reduce careless responding within the sample. This hypothesis was consistent with other research that has found that interventions to improve perceived control are effective at changing participant behavior (Sheeran et al., 2016).

The perceived behavioral control intervention provided information designed to affect beliefs about participants' autonomy and capacity. To address autonomy, I provided information that made it clear to participants that compensation is dependent on whether they answered the survey carefully. By doing so, I hoped to avoid the perception that there were impediments to compensation that are not related to careful responding and that were out of participants' control.

Additionally, the intervention addressed participants' capacity by providing information about the barriers to careful responding that might have negatively affected participants' perceived control. I provided information stating that the survey was designed so that all participants would be able to respond carefully and accurately, and I described the features of the survey that facilitated careful responding (e.g., it was as short as possible, provided clear directions, contained simple and straightforward questions). I also provided information about environmental distractions that may have gotten in the way of careful responding (e.g., others in the room, the tv or music playing in the background), and I instructed the participant to remove any barriers that may distract them or to wait to complete the survey until they could provide their undivided attention. I also instructed them to pause and return to the survey later if distractions arose while they were completing the survey.

By delivering this information, I hoped to provide participants with knowledge that increased their perception that they were capable of careful responding. Further, I hoped to provide the information that increased participants' perceptions that they were in control of whether they were awarded compensation in exchange for responding carefully. By increasing

participants' perceptions of control over their responses, I believed that the intervention would increase intentions to respond carefully and reduce the incidence of careless responding, as is consistent with the theory of planned behavior (Ajzen, 1991) and the integrated behavior model (Kasprzyk et al., 1998).

As was the case for the other intervention conditions, the primary outcome of interest for this study was careless responding behavior. However, I also assessed the extent to which the intervention affected perceptions of control and intentions to respond carefully. It was expected that the intervention would be successful at increasing perceived control, increasing intentions to respond carefully, and reducing careless responding behavior as compared to a control group.

**Hypothesis 3.1:** Those who received an intervention to increase perceived control over responding behavior believed they had more control over careful responding than those in the control group.

**Hypothesis 3.2:** Those who received an intervention to increase perceived control over responding behavior had stronger intentions to respond carefully than those in the control group.

**Hypothesis 3.3:** Those who received an intervention to increase perceived control over responding behavior were less likely to respond carelessly than those in the control group.

## CHAPTER 2: METHODS

### **Participants**

Participants were recruited through Amazon's MTurk platform and were paid \$0.50 for their participation. In order to participate, MTurk users were required to be adults employed outside of MTurk. The survey contained measures of variables commonly assessed in organizational research (e.g., job satisfaction, organizational commitment), and therefore participants needed to be employed in order to have an appropriate frame of reference for the survey.

A power analysis was conducted using G\*Power (Erdfelder, Faul, & Buchner, 1996) to determine a minimum sample size required for the study. Pilot testing indicated that the study interventions may have a small-to-medium size effects on attitudes, subjective norms, perceived control, and behavioral intentions. These effects were similar to those found in other theory of planned behavior intervention studies (Sheeran et al., 2016). At 80% power, a sample size of at least 135 participants per group (i.e., 540 participants overall) was required to detect the smallest effects found in the pilot study ( $\eta^2 = 0.02$ ). Based on this estimate, data was collected from at least 135 individuals in each condition for a total of 552 complete responses. No individual was permitted to participate in more than one study condition.

### **Design**

This study employed a between-subjects design with participants randomly assigned to the attitudes intervention, the subjective norms intervention, the perceived behavioral control intervention, or the control group. One-way ANOVAs were conducted to determine if there were differences between groups on behavioral antecedents and intentions. A one-way MANOVA was

conducted to determine if there were differences between groups on careless responding behavior, with one-way ANOVAs also reported to evaluate differences between groups on each individual careless responding metric. Hypotheses were tested by conducting planned contrasts to determine if each intervention group displayed the expected results on the relevant behavioral antecedent, behavioral intentions, and careless responding behavior as compared to the control group.

### **Procedures**

Study participants completed an online survey on Amazon's MTurk platform. The survey was designed to take about 30 minutes. Participants in the control groups began the survey immediately after consenting to participate. Intervention group participants were presented a page of material to read before proceeding to the survey. The material varied based on the intervention condition (see the "Manipulations" section below). After reading the material, intervention group participants were asked to confirm that they had read the information, and they were allowed to proceed.

On the next page, participants were told that before they started the main survey, there were a few questions to answer about their feelings towards taking surveys. They were told that it was very important that they answer these questions honestly and that nothing they say could cause me to reject their work. They then answered single item measures of the three behavioral antecedents (i.e., attitudes, subjective norms, and perceived behavioral control) and behavioral intentions. Next participants proceeded to the main survey, where they completed the filler scales that measured individual personality, job satisfaction, organizational commitment, openness to workplace health programs, openness to organizational change, and trust in management.

Finally, participants were asked to complete some post-survey questions. They were again told that it was very important that they answer these questions honestly and that nothing they say could cause me to reject their work. I then asked about their carefulness on the survey, their experiences on MTurk, and their demographic characteristics. Several manipulation check questions were also asked to determine if participants remembered seeing information designed to influence them to respond carefully, and if so, what type of information was presented and if they believed this information actually led them to respond more carefully.

### **Manipulations**

In each intervention group, participants were presented a page of material before the survey began. Depending on the condition, the material was designed to improve attitudes towards careful responding, increase subjective norms around careful responding, or increase perceptions of control over careful responding behavior.

Participants in the control group were not provided with any informational materials. This option was chosen over other options (e.g., assigning placebo reading materials not designed to change any behavioral antecedent) in order to ensure that control conditions closely resembled a typical survey experience. For these interventions to be considered effective, they should be compared to a researcher's default alternate option, providing no extra information at all. This option was also chosen to avoid actively increasing careless responding rates in the control group by increasing the length of the study and creating a perception that the study included inessential or irrelevant content. Intervention materials can be found in Appendix A.

### **Intervention Materials**

The attitudes intervention materials were designed to affect attitudes by increasing beliefs that responding carefully would lead to desirable outcomes. To do this, the materials explained

why careful responding was important and reiterated that careful responding would lead to compensation and that there were no other reasons why payment would be withheld.

The subjective norms intervention materials were designed to affect subjective norms by influencing normative beliefs and participants' desire to comply with those beliefs. To do this, the materials provided statistics about the high rates of careful responding among MTurk users and provided information about why other MTurk users believe it is important to respond to surveys (i.e., to have work approved and receive compensation, to support scientific research).

The perceived behavioral control intervention materials were designed to affect participants' perceived control over careful responding by changing their beliefs about the presence of barriers and how those barriers would affect their ability to respond carefully. To do this, the materials described how this survey was designed to facilitate careful responding (e.g., as short as possible, clear directions, straightforward questions) and described strategies for avoiding distractions that may prevent careful responding (e.g., turn off tv and music, take a break from responding if you are interrupted or distracted). The materials also reiterated that all work will be compensated so long as participants respond carefully.

### **Intervention Pilot Testing**

All intervention materials were pilot tested prior to data collection, in order to determine if they were effective at changing the relevant behavioral antecedents. Each pilot test condition contained approximately 10 participants.

The first several rounds of pilot testing examined the effects of the first intervention condition, the attitudes intervention, against the control intervention. Initial testing indicated that those who experienced the attitudes intervention had *worse* attitudes ( $M = 4.87$ ) towards careful



responding than those who did not experience the intervention ( $M = 5.35$ ), and they also had lower perception of carefulness norms ( $M = 4.95$  vs.  $M = 5.33$ ) and lower perceived control ( $M = 3.87$  vs.  $M = 4.13$ ).

Because the intervention negatively affected all antecedents not just the targeted antecedent, I suspected that something in the intervention condition may be affecting response patterns. I thought that perhaps by emphasizing the importance of carefulness, intervention participants also took that as an indicator that they should be more honest than they otherwise would have been. A second round of testing emphasized the importance of honesty by including stronger language before the measurement of the antecedents in both intervention and control conditions. After this change, intervention participants still reported worse attitudes than control participants ( $M = 5.14$  vs.  $M = 5.53$ ), and they also reported lower perceived norms ( $M = 4.98$  vs.  $M = 5.50$ ) but higher perceived behavioral control ( $M = 4.31$  vs.  $M = 4.08$ ).

Changes in the honesty wording did not lead to the expected results, so I considered the possibility that the attitudes intervention might simply be ineffective. I decided to launch a pilot of all four conditions to see if the other two intervention conditions were performing as expected. In all three intervention conditions, the effect of the intervention on the relevant antecedent was in the wrong direction as compared to the control group (attitudes:  $M = 4.59$  vs.  $M = 5.22$ ; subjective norms:  $M = 5.04$  vs.  $M = 5.35$ ; perceived behavioral control:  $M = 3.26$  vs.  $M = 3.87$ ).

At this point, it appeared that something about the interventions I had designed was actively reducing perceptions (or at least reported perceptions) of the relevant antecedents. I decided to make two major changes from the originally proposed intervention in order to try to correct this. First, I decided to remove the two test questions that asked participants about the interventions they read and forced participants who answered incorrectly to reread the

intervention materials and correct their answers. It seemed like this could have been annoying participants and leading them to respond less favorably. Second, I decided to change the way the antecedent variables were measured. Rather than measuring them with long scales at the end of the survey, I decided to measure them with single-item scales at the beginning of the survey. The idea was that this might do a better job of capturing their perceptions before beginning to respond to the survey. The scales were shortened to single-item measures in order to avoid overemphasizing the extent to which careful responding was a focus of this study.

With these new revised procedures, I relaunched a pilot of all four conditions. As compared to the control group, both the norms and perceived behavioral control conditions showed effects in the expected directions for the relevant antecedents (subjective norms:  $M = 6.20$  vs.  $M = 5.90$ ; perceived behavioral control:  $M = 6.25$  vs.  $M = 6.00$ ). The attitudes intervention still showed effects in the wrong direction ( $M = 6.13$  vs.  $M = 6.27$ ), so I decided to try an additional tweak to the content of that intervention. The wording of the intervention was altered to remove technical information about how their work would be evaluated, and instead to use more positive wording to describe the benefits of careful responding and how careful responding will definitely lead to approval of their work. Ultimately, this final attitudes intervention group also showed effects in the expected direction ( $M = 6.43$  vs.  $M = 6.27$ ).

In this final pilot, all three intervention groups also scored higher on intentions to respond carefully than the control group. Therefore, I decided to launch the survey using this version of the interventions.

### **Measures**

All items used as part of this study can be found in Appendix B. All measures were collected from participants in all conditions.

## **Behavioral Antecedents and Intentions**

All three behavioral antecedents (i.e., attitudes, subjective norms, and perceived behavioral control) and behavioral intentions were measured using a single-item, self-report measure. The attitudes item read “I think it is important to read and respond to survey questions carefully.” The subjective norms item read “most MTurk workers respond carefully to surveys.” The behavioral control item read “it will be easy to respond carefully to this survey.” The intentions item read “I plan to respond carefully to this survey.” All four items were answered on a seven-point Likert-type scale (1 = *Strongly disagree*; 7 = *Strongly agree*).

## **Careless Responding Behavior**

As described in the introduction, there are a variety of methods for detecting careful responding behavior that either directly measure the behavior or seek to measure it post-hoc based on participants’ provided responses. Consistent with other recent research on careless responding behavior (e.g., Ward & Meade, 2018; Ward & Pond, 2015), this study used a variety of metrics to quantify careless responding. One direct metric was used - the self-reported carefulness indicator. Four post-hoc metrics were also used to determine if individual response patterns indicated careless responding. Each metric was used to target a different type of careless response pattern. Response time captured inappropriately speedy responses, intra-individual response variability captured inappropriately consistent responding, even-odd consistency measured inappropriately inconsistent responding, and Mahalanobis D captured outlier responses.

In addition to metrics that measured individual-level careless responding, I also used one metric to examine group-level careless responding behavior. Following other recent research examining careless responding behaviors across groups (Kees et al., 2017; Kraiger et al., 2019;

Michel et al., 2017), I evaluated the extent to which the groups participating in my studies replicated the estimated population-level correlations between constructs. Samples that replicate known relationships are assumed to have higher quality data with lower rates of careless responding (Michel et al., 2017).

**Self-reported carefulness.** A single, self-report item was included at the end of the survey and asked participants to rate how carefully they responded to the survey. Responses were measured on a 7-point scale.

**Response time per page.** Response time captured the average amount of time an individual spent on each page of the survey. Average response time did not include any time spent reviewing intervention materials.

**Intra-individual response variability.** Intra-individual response variability (Dunn et al., 2018) indicated the extent to which an individual's responses displayed appropriate response variability. It was calculated by taking the standard deviation of responses to items in the filler scales. All calculations were conducted with responses in their original form; no negatively-worded items were reverse coded prior to calculation. Because the filler scales measured multiple constructs with both positive and negatively-worded items, careful responses should have displayed a substantial degree of response variability. Thus, low intra-individual variability indicated careless responding.

**Even-odd consistency.** Even-odd consistency indicated the extent to which participants provided consistent responses to items measuring the same construct. For this metric, all negatively-worded items were reverse coded. For each construct, items were split into two groups based on the order in which they appeared in the survey. Average scores were calculated for each individual on each half of each scale, and then an intra-individual correlation between

scale scores for the first and second halves of each scale was calculated. This intra-individual correlation served as the even-odd consistency metric. Correlations nearer to +1 indicate greater response consistency on internally-consistent scales, and thus, these scores indicated a lesser degree of careless responding. Smaller scores indicated a greater degree of careless responding.

**Mahalanobis D.** Mahalanobis D indicated the extent to which a response pattern represented a multivariate outlier in the data. Larger D values indicated that a response pattern was more of an outlier, and thus suggested that more careless responding occurred.

**Relationship replication.** Relationship replication examined the extent to which correlations between measured variables in the experimental and control conditions replicated the estimated population correlations. Correlations closer to the estimated population values indicated less careless responding in that condition. This was done for three sets of variable pairs that varied in the strength and direction of their relationship estimates.

The first variable pair exhibits a strong, positive correlation in the population, the relationship between job satisfaction and organizational commitment. Correlations were compared against estimates from a large meta-analysis that found that the uncorrected correlation between the two variables was .49 ( $K = 68$ ,  $N = 35,282$ ; Mathieu & Zajac, 1990).

The second pair exhibits a near-zero correlation in the population, the relationship between openness to experience and job satisfaction. Correlations between these variables were compared to results of a large meta-analysis that estimated the uncorrected population correlation was .01 ( $K = 50$ ,  $N = 15,196$ ; Judge, Heller, & Mount, 2002).

The third pair exhibits a moderate negative correlation in the population, the relationship between neuroticism and job satisfaction. Correlations between the variables were compared to

results of a large meta-analysis that estimated the uncorrected population correlation was  $-.24$  ( $K = 92$ ,  $N = 24,527$ ; Judge et al, 2002).

### **Filler Scales**

Several filler scales were selected to comprise the body of the study. When combined, the filler scales consisted of 133 items. For the work-related measures (i.e., all but personality), participants were instructed to answer questions with reference to their primary job not their work on MTurk.

**Personality.** The Big Five Inventory (John, Donahue, & Kentle, 1991) was used to assess Five Factor personality traits. The measure contains 44 items, with eight to ten items in each of the five trait scales: Extraversion, Agreeableness, Emotional Stability, Conscientiousness, and Openness to Experience. Each subscale has shown acceptable internal consistency reliability (typically averaging above  $.80$ ) and three-month test-retest reliability (typically averaging around  $.85$ ; John, Naumann, & Soto, 2008). Participants were shown statements that begin “I am someone who...,” and asked to respond with how much they agree or disagree with each item on a seven-point Likert-type scale ( $1 = \textit{Strongly disagree}$ ,  $7 = \textit{Strongly agree}$ ).

**Job satisfaction.** Job satisfaction was measured using the 20-item short form of the Minnesota Satisfaction Questionnaire (Weiss, Dawis, England, & Lofquist, 1967). The scale measures the extent to which employees are satisfied with various aspects of their job. The measure contains items that capture satisfaction with intrinsic and extrinsic components of the job. Both intrinsic ( $\alpha = .85$ ) and extrinsic ( $\alpha = .81$ ) subscales have displayed strong internal consistency (Hirschfeld, 2000). All items were answered on a 7-point Likert-type scale ( $1 = \textit{Very dissatisfied}$ ;  $7 = \textit{Very satisfied}$ ).

**Organizational commitment.** Organizational commitment was measured using Allen and Meyer's (1990) 24-item organizational commitment measure. The measure was designed to capture three components of organizational commitment: affective commitment, continuance commitment, and normative commitment. All subscales have displayed acceptable internal consistency in previous research ( $\alpha = .73 - .86$ ; Allen & Meyer, 1990). Items were answered on a 7-point Likert-type scale (1 = *Strongly disagree*; 7 = *Strongly agree*).

**Openness to workplace health programs.** Openness to support workplace health programs was measured using a 27-item, two-dimensional scale adapted from Cave and Daigle (2017). The scale's utility dimension measures the extent to which participants believe implementing a workplace health intervention will have a beneficial impact. The appropriateness dimension measures the extent to which participants believe their organization should be offering health interventions to employees. Both subscales displayed acceptable internal consistency in previous research ( $\alpha = .83$  and  $.85$ , respectively; Cave & Daigle, 2017). Items were answered on a 7-point Likert-type scale (1 = *Strongly disagree*; 7 = *Strongly agree*).

**Openness to organizational change.** Openness to organizational change was measured using eight items adapted from Miller, Johnson, and Grau ( $\alpha = .80$ ; 1994). The original scale was developed to evaluate individuals' willingness to support a specific proposed organizational change. In this study, items were adapted to reflect openness towards organizational change in general. Items were answered on a 7-point Likert-type scale (1 = *Strongly disagree*; 7 = *Strongly agree*).

**Trust in management items.** Trust in management was measured using 10 items adapted from Mayer and Gavin ( $\alpha = .76 - .82$ , depending on the measurement source; 2005). Items were answered on a 7-point Likert-type scale (1 = *Strongly disagree*; 7 = *Strongly agree*).

## CHAPTER 3: RESULTS

### **Demographics**

A total of 662 participants began the online survey, though 71 of those participants (10.7%) were prohibited from participating after the screening questions because they did not meet the inclusion criteria (i.e., they were under 18 or not employed outside of MTurk). Of the 591 participants who qualified to participate and began the survey, 552 filled out the survey to completion (93.4%).

On average, participants were 35.2 years old ( $SD = 11.0$ ). They were slightly more likely to be male (53.2%) than female (46.6%). The most common racial or ethnic identity among participants was White (66.1%), followed by Black (11.2%), Native American (5.9%), Asian (5.6%), Hispanic or Latino (4.5%), and other or multiple identities (3.7%). All participants reported attaining at least a high school diploma, 70.8% had attained at least a bachelor's degree, and 21.4% had a graduate degree.

All participants were required to have a job outside of MTurk to participate, and 82.8% of those participants reported that they had a full-time position. The remaining 17.2% of individuals were working part time (i.e., less than 30 hours a week). A majority of participants reported that they supervise others as part of their job (59.1%), while the remaining participants did not (38.6%).

### **Descriptive Statistics**

Descriptive statistics for the behavioral antecedents, behavioral intentions, and careless responding variables can be found in Table 1. Means for attitudes, subjective norms, perceived behavioral control, careful responding intentions, and self-reported careful responding behavior



were all relatively high ( $M = 5.72-6.32$ ;  $SD = 1.02-1.68$ ; measured a 1-7 Likert-type scale) and negatively skewed (Skew = -1.90 to -1.21). The average response time per page of the survey was 26.98 seconds, and the variance in response time across respondents was very high ( $SD = 27.05$ ). The average even-odd consistency score was positive ( $M = 0.45$ ), but this metric also showed a relatively large variance across participants ( $SD = .60$ ). Average intra-individual response variability (i.e., standard deviation of a participant's responses across items) was 1.47 ( $SD = 0.56$ ), and the average Mahalanobis D value was 17.81 ( $SD = 9.74$ ).

Table 1

*Descriptive Statistics for Study Variables*

Variable	<i>M</i>	<i>SD</i>
Attitudes	6.28	1.09
Subjective Norms	5.72	1.31
Perceived Behavioral Control	6.15	1.08
Behavioral Intentions	6.32	1.02
Self-Reported Carefulness	5.87	1.68
Average Response Time Per Page	26.98	27.05
Even-Odd Consistency	0.45	0.60
Mahalanobis D	17.81	9.74
Intra-Individual Response Variability	1.47	0.56

*Note.* N = 552

### **Theory of Planned Behavior Applied to Careless Responding Behavior**

Prior to testing the effects of the interventions deployed in this study, I examined the relationships between theory of planned behavior variables as applied to careless responding behaviors. Table 2 displays a correlation matrix of the relationships between the measured variables in this study.

Table 2

*Correlations Between Variables*

Scale	1	2	3	4	5	6	7	8
1 Attitudes								
2 Subjective Norms	<b>0.26</b>							
3 Perceived Behavioral Control	<b>0.62</b>	<b>0.32</b>						
4 Intentions to Respond Carefully	<b>0.69</b>	<b>0.31</b>	<b>0.63</b>					
5 Self-Reported Carefulness	<b>0.29</b>	0.03	<b>0.29</b>	<b>0.29</b>				
6 Average Page Time	<b>0.20</b>	-0.01	<b>0.15</b>	<b>0.16</b>	<b>0.19</b>			
7 Even-Odd Consistency	<b>0.33</b>	-0.06	<b>0.23</b>	<b>0.34</b>	<b>0.45</b>	<b>0.18</b>		
8 Mahalanobis D	<b>-0.16</b>	<b>-0.14</b>	<b>-0.13</b>	<b>-0.11</b>	0.05	0.02	0.06	
9 Intra-Individual Response Variability	<b>0.32</b>	0.00	<b>0.27</b>	<b>0.34</b>	<b>0.44</b>	<b>0.29</b>	<b>0.70</b>	<b>0.49</b>

Note. N = 552, bolded values are significant at  $p < .01$

**Relationships Between Behavioral Antecedents**

The primary antecedents in the theory of planned behavior model are attitudes towards the behavior, subjective norms around the behavior, and perceived control over the behavior. All three antecedents were significantly and positively correlated with one another. The relationship between attitudes and perceived behavioral control was strong ( $r = .62, p < .001$ ), and the relationships between attitudes and subjective norms ( $r = .26, p < .001$ ) and subjective norms and perceived behavioral control ( $r = .32, p < .001$ ) were moderate.

**Relationships Between Careless Responding Metrics**

The final outcome in the theory of planned behavior model is performance of the behavior, in this case, careless responding behavior. Five individual-level metrics were used to assess careless responding behavior – self-reported carefulness, average time spent per page of the survey, even-odd consistency, Mahalanobis D, and intra-individual response variability. Larger values indicate a lower likelihood of careless responding on all metrics except Mahalanobis D; lower Mahalanobis D values indicate a lower likelihood of careless responding.

There were significant correlations among all careless responding metrics, with exception of Mahalanobis D. Mahalanobis D was only related to one other careless responding metric, intra-individual response variability ( $r = .49, p < .001$ ). However, the relationship between the two metrics was in the opposite direction as expected. That is, greater careless responding as indicated by one metric was correlated with less careless responding as indicated by the other. The non-significant relationships between Mahalanobis D and the other careless responding indicators were also in the opposite direction as was expected.

Correlations among the other careless responding metrics ranged in strength. Page time had small to moderate correlations with the other metrics ( $.18 \leq r \leq .29$ ). The relationships between self-reported carefulness, even-odd consistency, and intra-individual response variability were all moderate to large ( $.44 \leq r \leq .70$ ), and the correlation between even-odd consistency and intra-individual response variability was particularly large ( $r = .70, p < .001$ ).

### **Antecedents and Behavioral Intentions**

As shown in Table 2, each of the three theory of planned behavior antecedents (i.e., attitudes, subjective norms, and perceived behavioral control) was significantly correlated with intentions to respond carefully. The relationship between subjective norms and intentions was moderately-sized ( $r = .31, p < .001$ ), and the relationships between attitudes and intentions ( $r = .69, p < .001$ ) and perceived behavioral control and intentions ( $r = .63, p < .001$ ) were large.

A multiple linear regression analysis was conducted to determine the combined effects of the three antecedents on careless responding intentions. Together, the antecedents significantly predicted intentions to respond carefully,  $F(3, 548) = 223.4, p < .001, R^2 = .55$ , explaining 55.0% of the variance in careful responding intentions. Holding the other antecedents constant, all three predictors explained unique variance in intentions – attitudes:  $\beta = 0.44, t(548) = 12.90$ ,

$p < .001$ ; subjective norms:  $\beta = 0.07$ ,  $t(548) = 2.81$ ,  $p < .01$ ; perceived behavioral control:  $\beta = 0.30$ ,  $t(548) = 8.47$ ,  $p < .001$ ].

### **Behavioral Intentions and Careless Responding Metrics**

The relationship between careless responding intentions and careless responding behavior was evaluated using all five individual-level careless responding metrics. Simple linear regression analyses were conducted to evaluate the impact of intentions on each careless responding metric.

First, I examined the effect of careful responding intentions at the beginning of the survey on self-reported carefulness at the end of the survey. Greater intentions to respond carefully significantly predicted self-reported carefulness,  $F(1, 550) = 51.35$ ,  $p < .001$ . Each one-point increase in intentions was associated with an average 0.48 increase in self-reported carefulness at the end of the study.

Next, average response time per page was evaluated. A greater response time indicates a lower likelihood of careless responding. There was a positive relationship between careful responding intentions and average response time per page,  $F(1, 550) = 15.09$ ,  $p < .001$ . Each one-point increase in intentions was associated with an additional 4.32 seconds spent per page on average.

Mahalanobis D was also evaluated as a careless responding metric. A larger Mahalanobis D value indicates that a response is more of an outlier, and thus more likely to be a careless response. The results indicated that intentions significantly predicted Mahalanobis D scores,  $F(1, 574) = 11.18$ ,  $p < .001$ . Greater intentions to respond carefully were associated with lower average Mahalanobis D scores, i.e., less careless responding.

Intra-individual response variability indicates a person's response variation across diverse constructs with both positive- and negatively-coded items. Greater variability indicates less careless responding. Intentions to respond carefully significantly predicted response variability,  $F(1, 551) = 72.53, p < .001$ , such that greater intentions led to greater variability in participant responses.

Finally, even-odd consistency measures a participant's consistency across items that measure the same construct and are coded in the same direction. Greater consistency indicates less careless responding behavior. Intentions to respond carefully significantly predicted even-odd consistency,  $F(1, 487) = 62.08, p < .001$ , such that greater intentions led to greater consistency in participant responses.

### **Careless Responding Interventions**

#### **Manipulation Checks**

Prior to testing the effects of the interventions deployed in this study, the results of several manipulation check questions were examined. The first manipulation check question asked participants whether they were shown materials designed to convince them to respond more carefully. A chi-square test of independence indicated that there was a relationship between study condition and the response to the first manipulation check question,  $\chi^2(3) = 8.12, p = .04$ . Participants in the control condition were the least likely to respond that they had seen such materials (74.6%), followed by those in the attitudes (76.4%), control (78.9%), and norms (87.5%) interventions. However, a majority of participants in all conditions endorsed the idea that they had been shown materials intended to convince them to respond carefully.

The second manipulation check question was only shown to those participants who responded that they had been shown intervention materials. These participants were asked to

decide if they believed that they had been shown materials intended to convince them that careful responding is important (i.e., attitudes intervention), that other MTurk participants respond carefully so they should too (i.e., subjective norms intervention), or that it would be easy to provide careful responses (i.e., perceived behavioral control intervention). A chi-square test of independence indicated that there was a relationship between study condition and response to the second manipulation check question,  $\chi^2(9) = 32.16, p < .001$ . However, a majority of all four study groups believed that they had been shown materials designed to convince them that careful responding is important. The percentage that endorsed the attitudes option was actually highest in the control group (89.2%), followed by the perceived behavioral control intervention group (81.6%), the norms intervention group (78.9%), and the attitudes intervention group (76.6%). However, those in the norms intervention were more likely to endorse the norms option (19.3%) than those in any of the other three groups (6.9-14.9%), and those in the perceived behavioral control intervention condition were more likely to endorse the perceived behavioral control option (11%) than those in the other groups (0.0-8.4%).

The final manipulation check question was also only shown to participants who believed that they were shown intervention materials. This third question asked participants if they believed the materials they read actually convinced them to respond more carefully (where response options ranged from 1 = “Definitely not” to 7 = “Definitely yes”). There were not significant differences in average response by study condition,  $F(3, 433) = 1.14, p = .33$ , although the mean for those in the norms intervention group ( $M = 5.03$ ) was higher than the mean for the attitudes intervention ( $M = 4.70$ ), perceived behavioral control intervention ( $M = 4.61$ ), or the control group ( $M = 4.65$ ).

## Effects of Interventions on Behavioral Antecedents

The first set of hypotheses – Hypotheses 1.1, 2.1, and 3.1 – stated that the intervention conditions would have a positive effect on the targeted behavioral antecedent as compared to the control group. It was expected that the results of the study would replicate the findings of the pilot study, which found positive effects for all three interventions on the relevant antecedent.

Hypothesis 1.1 concerned the effect of the attitudes intervention on attitudes towards careful responding. First, a one-way ANOVA was conducted to determine if there were significant differences between the four study conditions. The results revealed no significant differences in attitudes across conditions,  $F(3, 548) = .32, p = .80$ . Hypothesis 1.1 was then directly tested by conducting a planned contrast between the attitudes intervention and control group. There were not significant differences between the two groups,  $t(276) = 0.13, p = .45$ . The average attitude score was similar in the intervention ( $M = 6.25$ ) and control groups ( $M = 6.23$ ). Therefore, Hypothesis 1.1 was not supported.

Hypothesis 2.1 concerned the effect of the subjective norms intervention on subjective norms around careful responding. A one-way ANOVA was conducted to determine if there were significant differences in subjective norms between the four study conditions. The results revealed no significant differences in subjective norms across conditions,  $F(3, 548) = 1.05, p = .37$ . Planned contrasts indicated that there were not significant differences between the norms intervention group ( $M = 5.76$ ) and the control group ( $M = 5.65; t(272) = 0.66, p = .25$ ). Hypothesis 2.1 was not supported.

Hypothesis 3.1 concerned the effect of the perceived behavioral control intervention on perceived control over careful responding behavior. A one-way ANOVA was conducted to determine if there were significant differences in perceived behavioral control between the four

study conditions. The results revealed no significant differences in perceived behavioral control across conditions,  $F(3, 548) = 0.29, p = .83$ . Planned contrasts indicated that there were not significant differences between the intervention group ( $M = 6.19$ ) and the control group ( $M = 6.09$ ;  $t(274) = 0.68, p = .24$ ). Hypothesis 3.1 was not supported.

### **Effects of Interventions on Behavioral Intentions**

Hypotheses 1.2, 2.2, and 3.3 stated that the interventions would have positive effects on intentions to respond carefully, as compared to the control group. A one-way ANOVA was conducted to determine if there were significant differences in intentions to respond carefully across the four study conditions. The results indicated that there were no significant differences in behavioral intentions across the groups,  $F(3, 548) = 0.23, p = .87$ .

Hypothesis 1.2. concerned the effect of the attitudes intervention on intentions to respond carefully. Planned contrasts indicated that there were not significant differences between the attitudes intervention group and the control group,  $t(276) = -.48, p = .68$ . In fact, the mean intentions score was slightly higher for the control group ( $M = 6.37$ ) than the intervention group ( $M = 6.32$ ). Hypothesis 1.2 was not supported.

Hypothesis 2.2 stated that the subjective norms intervention would have a positive effect on careless responding intentions as compared to the control group. As was the case with the attitudes intervention, there was not a significant difference between the subjective norms intervention group and the control group,  $t(272) = -.71, p = .76$ . The mean was slightly higher for the control group ( $M = 6.37$ ) than the intervention group ( $M = 6.30$ ). Hypothesis 2.2 was not supported.

Hypothesis 3.2 proposed that the perceived behavioral control intervention would have a positive effect on intentions as compared to the control group. However, there was not a



significant difference between the perceived behavioral control intervention group and the control group,  $t(274) = -.74, p = .77$ . The mean for the intervention group ( $M = 6.28$ ) was slightly lower than the mean for the control group ( $M = 6.37$ ). Hypothesis 3.2 was not supported.

### **Effects of Interventions on Careless Responding Behavior**

The ultimately goal of the interventions was to reduce careless responding behavior, so the final set of hypotheses served as the criterion by which the efficacy of the interventions were evaluated. These hypotheses – Hypotheses 1.3, 2.3, and 3.3 – posited that the interventions would reduce careless responding behavior. The third set of hypotheses was examined using six careless responding metrics.

**Exploratory factor analysis.** First, an exploratory factor analysis was conducted to determine if the five individual-level metrics– self-reported carefulness, average time spent per page of the survey, even-odd consistency, Mahalanobis D, and intra-individual response variability – were explained by a common underlying careless responding factor. The results revealed that a single-factor solution was not a good fit to the data ( $\chi^2(5) = 81.45, p < .001$ ; RMSEA = .16; CFI = 0.90; TLI = 0.81). However, a two-factor solution was a good fit to the data ( $\chi^2(1) = 0.00, p = .99$ ; RMSEA < .001; CFI = 1.00; TLI = 1.01) and a significantly better fit than the single-factor solution ( $\Delta\chi^2(4) = 81.45, p < .001$ ).

Further review of the one- and two factor solutions indicated unexpected factor loadings for one metric in particular. For each of the careless responding metrics except Mahalanobis D, larger values indicate *less* careless responding. For Mahalanobis D, larger values indicate *more* careless responding. Thus, it would be expected that the Mahalanobis D indicator should have factor loadings in the opposite direction of the other indicators. However, this was not the case in either the one- or two-factor solutions. In each case, the Mahalanobis D values loaded onto its

primary factor in the same direction as the other careless responding metrics. This was consistent with correlational findings (presented in Table 2 above) that showed that Mahalanobis D scores were not significantly related to any other careless responding metric, and the non-significant relationships were in the opposite direction as was expected.

Because the Mahalanobis D metric did not appear to be behaving in the expected manner, a second exploratory analysis was conducted that omitted the metric. The remaining four careless responding metrics were a good fit to a one-factor solution ( $\chi^2(6) = 8.49, p = .01$ ; RMSEA = .07; CFI = 0.99; TLI = 0.96). All four careless responding metrics significantly loaded onto the factor (factor loadings  $> .30, p < .05$ ).

**MANOVA.** A one-way MANOVA was conducted to examine the effects of the study interventions on careless responding. The MANOVA was initially calculated using all five individual-level careless responding metrics as dependent variables. The results revealed that study condition did not lead to significant omnibus effects on careless responding metrics,  $F(3, 464) = 1.50, p = .09$ ; Wilk's  $\Lambda = 0.95$ .

Because the exploratory factor analysis showed that Mahalanobis D scores did not load onto the careless responding factor in the expected direction, a MANOVA analysis was also conducted without Mahalanobis D included as a dependent variable. However, this change did not affect the results; study condition did not lead to significant omnibus effects on careless responding metrics,  $F(3, 464) = 1.43, p = .14$ ; Wilk's  $\Lambda = 0.96$ .

Although the omnibus MANOVA was not significant, I did explore the effects of study condition on each careless responding metric individually. Additionally, planned contrasts were conducted to determine if there were significant differences between each intervention group and control group on each careless responding metric. These planned contrasts served as direct tests

of Hypotheses 1.3, 2.3, and 3.3. However, because the MANOVA did not indicate significant overall differences in careless responding rates across conditions, these planned contrasts should be interpreted with caution. The results for each planned contrast are summarized in Table 3.

Table 3

*Results of Planned Contrasts by Dependent Variable*

Planned Contrast	<i>t</i>	<i>df</i>	<i>p-value</i>
<b>Attitudes vs. Control Group</b>			
Self-Reported Carefulness	1.24	276	0.10
Average Page Time	-0.69	276	0.76
Even-Odd Consistency	-0.34	247	0.63
Mahalanobis D	-0.15	290	0.44
Intra-Individual Response Variability	0.38	278	0.35
<b>Norms vs. Control Group</b>			
Self-Reported Carefulness	1.81	272	<b>0.03</b>
Average Page Time	-0.29	272	0.61
Even-Odd Consistency	0.91	247	0.18
Mahalanobis D	-1.29	287	0.09
Intra-Individual Response Variability	-0.07	273	0.53
<b>Behavioral Control vs. Control Group</b>			
Self-Reported Carefulness	2.04	274	<b>0.01</b>
Average Page Time	0.66	274	0.25
Even-Odd Consistency	2.25	245	<b>0.01</b>
Mahalanobis D	-0.33	283	0.37
Intra-Individual Response Variability	1.89	276	<b>0.03</b>

*Note.* Bolded values are significant at  $p < .05$

**Self-reported carefulness.** The first indicator of careless responding behavior examined individually was self-reported carefulness. A one-way ANOVA indicated that there were no significant differences between the groups on self-reported carefulness,  $F(3, 548) = 2.17, p = .09$ . Although there were not significant differences among all four groups, planned contrasts were conducted to explore the possibility of differences between individual intervention groups and the control group. Planned contrasts revealed that there was no significant difference in self-reported carefulness between the attitudes intervention group ( $M = 5.85$ ) and the control group

( $M = 5.58$ ,  $t(276) = 1.24$ ,  $p = .10$ ). However, there were significant differences in careless responding between the subjective norms group ( $M = 5.97$ ) and the control group ( $t(272) = 1.81$ ,  $p = .03$ ), as well as between the perceived behavioral control group ( $M = 6.07$ ) and the control group ( $t(274) = 2.04$ ,  $p = .008$ ). Therefore, with regard to the self-reported carefulness metric, there was support for Hypotheses 2.3 and 3.3, but not for Hypothesis 1.3. However, because the ANOVA did not reveal significant overall differences, these significant contrasts should be interpreted with caution.

**Average response time per page.** The second indicator of careless responding behavior was average response time per page. Results of the one-way ANOVA indicated that there were not significant differences in average response time per page across groups,  $F(3, 548) = 0.59$ ,  $p = .62$ ,  $\eta^2 = .003$ .

Group means indicated that those in the perceived behavioral control intervention group ( $M = 29.36$ ) took the longest average response time per page, followed by the control group ( $M = 27.16$ ), the subjective norms group ( $M = 26.11$ ), and the attitudes group ( $M = 25.29$ ). Planned contrasts indicated that there was not a significant time difference between the control group and the attitudes intervention group ( $t(276) = -0.69$ ,  $p = .76$ ), norms intervention group ( $t(272) = -0.29$ ,  $p = .61$ ), or the perceived behavioral control intervention group ( $t(274) = 0.66$ ,  $p = .25$ ). Thus, there was not support for Hypotheses 1.3, 2.3, or 3.3 with regard to average response time per page.

**Mahalanobis D.** Mahalanobis D was also evaluated as a careless responding metric. A larger Mahalanobis D value indicates that a response is more of an outlier, and thus more likely to be a careless response. A one-way ANOVA was conducted to determine if there were

differences in average D values across study conditions. The results indicated that there were not,  $F(3, 574) = 0.60, p = .61, \eta^2 = .003$ .

The average D value was lowest in the subjective norms intervention condition ( $M = 17.12$ ), followed by the perceived behavioral control condition ( $M = 18.23$ ), the attitudes condition ( $M = 18.40$ ), and the control group ( $M = 18.61$ ). Planned contrasts indicated that there was not a significant difference between the control group and the attitudes intervention group ( $t(290) = -0.15, p = .44$ ), the perceived behavioral control intervention group ( $t(287) = -0.33, p = .37$ ), or the norms intervention group ( $t(283) = -1.29, p = .09$ ). Therefore, there was not support for Hypotheses 1.3, 2.3 or 3.3 with regards to Mahalanobis D metric.

**Intra-individual response variability.** Differences across groups on intra-individual response variability were also examined. On this metric, greater variability indicates less careless responding. A one-way ANOVA indicated that there were no significant differences in response variability across study conditions,  $F(3, 551) = 1.58, p = .19, \eta^2 = .01$ .

Response variability was greatest in the perceived behavioral control intervention group ( $M = 1.56$ ), followed by the attitudes group ( $M = 1.46$ ), the control group ( $M = 1.43$ ), and the subjective norms group ( $M = 1.43$ ). Although there were no significant overall differences between groups, planned contrasts were conducted to explore the possibility of differences between individual intervention groups and the control group. Planned contrasts indicated that there was not a significant difference in response variability between the control group and either the attitudes intervention group ( $t(278) = 0.38, p = .35$ ) or the norms intervention group ( $t(273) = -0.07, p = .53$ ). However, there was a significant difference between the control group and the perceived behavioral control intervention group ( $t(276) = 1.89, p = .03$ ). With respect to intra-individual response variability, there was support for Hypothesis 3.3 but there was not support

for Hypotheses 1.3 or 2.3. However, because the ANOVA did not reveal significant overall differences, the significant contrast should be interpreted with caution.

**Even-odd consistency.** Even-odd consistency was evaluated as the final measure of individual-level careless responding behavior. Greater consistency on this metric indicates less careless responding behavior. A one-way ANOVA found that there were significant differences in even-odd consistency across study conditions across groups,  $F(3, 487) = 2.54, p = .05, \eta^2 = .02$ .

Group means indicated that the average even-odd consistency was highest in the perceived behavioral control intervention group ( $M = 0.58$ ), followed by the subjective norms group ( $M = 0.48$ ), the control group ( $M = 0.41$ ), and the attitudes group ( $M = 0.38$ ). Planned contrasts indicated that there was a significant difference between the control group and the perceived behavioral control intervention group ( $t(245) = 2.25, p = .01$ ). However, there were not significant differences between the control group and either the attitudes intervention group ( $t(247) = -0.34, p = .63$ ) or the norms intervention group ( $t(247) = 0.91, p = .18$ ). Therefore, with respect to even-odd consistency, there was support for Hypothesis 3.3 but not for Hypotheses 1.3 or 2.3.

**Relationship replication.** The final careless responding metric was the extent to which each group replicated estimated population-level correlations from previous research. The first relationship examined was between job satisfaction and organizational commitment, which has an estimated uncorrected population-level correlation of .49 (Mathieu & Zajac, 1990).

In all four groups included in this study, the satisfaction-commitment correlation was positive and significant. The perceived behavioral control intervention group's correlation was closest to the population correlation,  $r = .48, p < .001, 95\% \text{ CI } [.34, .60]$ ; followed by the control

group,  $r = .38, p < .001, 95\% \text{ CI } [.23, .51]$ ; the subjective norms group,  $r = .36, p < .001, 95\% \text{ CI } [.20, .50]$ ; and the attitudes group,  $r = .31, p < .001, 95\% \text{ CI } [.15, .45]$ . The confidence interval for each group except the attitudes group included the estimated population correlation of .49. Because the perceived behavioral control group's satisfaction-commitment correlation was the only correlation that was closer to the estimated population correlation than the control group's, I concluded that there was evidence to support Hypothesis 3.3 but not to support Hypotheses 3.1 or 3.2.

The second relationship examined was between job satisfaction and openness to experience, which has an estimated uncorrected population-level correlation of .01 (Judge et al., 2002). In all four groups included in this study, the satisfaction-openness correlation was positive and significant. The correlation was closest to the estimated population correlation in the control group,  $r = .32, p < .001, 95\% \text{ CI } [.16, .46]$ ; followed by the subjective norms intervention group,  $r = .34, p < .001, 95\% \text{ CI } [.18, .48]$ ; the perceived behavioral control group,  $r = .46, p < .001, 95\% \text{ CI } [.32, .58]$ ; and the attitudes group,  $r = .49, p < .001, 95\% \text{ CI } [.36, .61]$ . No group's confidence interval contained the estimated population value. Because no group was closer to the estimated satisfaction-openness correlation in the population than the control group, there was no evidence to support any of the three hypotheses in question for this second relationship.

The final relationship was the correlation between neuroticism and job satisfaction, estimated to be -.24 in the population (uncorrected  $r$ ; Judge et al. 2002). The satisfaction-neuroticism correlation was negative in all four study groups, though the relationship was not always significant. The correlation was closest to the estimated population correlation in the perceived behavioral control intervention group,  $r = -.27, p < .001, 95\% \text{ CI } [-.42, -.11]$ ; followed by the attitudes intervention group,  $r = -.34, p < .001, 95\% \text{ CI } [-.48, -.19]$ ; the norms intervention

group,  $r = -.14$ ,  $p = .10$ , 95% CI [-.30, .03]; and the control group,  $r = -.11$ ,  $p = .19$ , 95% CI [-.27, .06]. The confidence interval for all four groups included the estimated population correlation. Because each group was closer to the estimated satisfaction-neuroticism correlation in the population than the control group, I concluded there was evidence to support all three of the final hypotheses for this third relationship.



## CHAPTER 4: DISCUSSION

This research study had two major goals. The first was to deepen our understanding of the theoretical basis for careless responding behavior. This study rests on the theoretical assertion that careless responding behavior is affected by individual motivation. By testing the tenants of a well-supported motivational theory, the theory of planned behavior (Ajzen, 1991), this study sheds light on whether it is possible to affect careless responding behavior by increasing individual motivation to respond accurately. Second, this study provides and tests practical methods for reducing careless responding behavior in applied and academic online survey research.

### **Theory of Planned Behavior Applied to Careless Responding**

To the first goal of this study, I applied a well-supported motivational theory, the theory of planned behavior (Ajzen & Fishbein, 1980; Ajzen, 1991), in an attempt to explain careless survey responding. The theory of planned behavior posits that three behavioral antecedents – attitudes towards a behavior, subjective norms around a behavior, and perceived behavioral control over a behavior – each predict intentions to perform a behavior. In turn, intentions then predict whether an individual will perform the behavior in question (Ajzen, 1991).

In this study, all five of the primary theory of planned behavior variables were applied to the problem of careful responding. The relationships proposed by the theory of planned behavior were then explored using the entire sample collected as part of this study. Consistent with empirical research across a wide-variety of other behavioral outcomes (Albarracín et al., 2001; Armitage & Connor, 2001; Cooke & French, 2008; Hagger et al., 2002), I expected to find that

the theory of planned behavior was successful at explaining careless responding behavior on surveys.

I first examined the relationships between the antecedent variables in the model (i.e., attitudes, norms, and behavioral control) with the intermediate outcome, behavioral intentions. The results indicated that individuals who believed it was important to respond carefully, those who thought most others respond carefully, and those who thought it would be easy to respond carefully all planned to respond more carefully themselves. I also examined the relationship between intentions to respond carefully and the ultimate outcome, careful responding behavior. The results showed that those who intended to respond carefully did, in fact, respond more carefully according to all careless responding metrics.

These findings suggest that the theory of planned behavior does, in fact, explain significant variance in careful responding behavior. These findings are consistent with previous research indicating that the theory of planned behavior is able to explain behavior across a variety of outcomes (Albarracín et al., 2001; Armitage & Connor, 2001; Cooke & French, 2008; Hagger et al., 2002). Ultimately, these findings suggest that interventions which successfully affect attitudes, norms, and perceived control or interventions which affect response intentions through other means, may be able to reduce careless responding behavior.

### **Efficacy of Theory of Planned Behavior Interventions**

The second goal of this study was to test practical methods for reducing careless responding behavior. To do this, I designed three interventions based on the theory of planned behavior. Each intervention was intended to affect one of the three theory of planned behavior antecedents – attitudes, subjective norms, and perceived behavioral control.

To test the effects of the theory of planned behavior interventions, I examined the effects of three interventions on behavioral antecedents, behavioral intentions, and careless responding behavior. Ultimately, the third set of hypotheses were designed to directly evaluate the interventions against the underlying goal of this study, to develop methods for successfully reducing careless responding behavior.

### **Behavioral Antecedents**

The first set of hypotheses concerned the effects of the interventions on the theory of planned behavior's antecedents, i.e., attitudes, norms, and perceived control. The hypotheses posited that the interventions would have a positive effect on the relevant behavioral antecedent as compared to the control group. This would be consistent with findings from this study's pilot testing, as well as previous research (Derzon & Lipsey, 2002; Sheeran et al., 2016). However, the results of the study revealed no significant differences between the relevant intervention group and the control group for any of the three behavioral antecedents. Thus, the interventions were not effective at changing attitudes, subjective norms, or perceived behavioral control.

In part, it may have been the case that the study was not able to affect behavioral antecedents because the brevity of the intervention itself. The intervention materials were purposefully developed to be brief in order to increase their practical utility. Brief interventions have been used in other careless responding research studies to mixed effects (e.g., Gibson, 2016; Ward & Meade, 2018; Ward & Pond, 2015), although these studies have exclusively focused on careless responding behavior as an outcome. In my study, which also evaluated process variables, the brief interventions may not have affected behavioral antecedents. Alternatively, the effects on behavioral antecedents may have been smaller than anticipated, and thus they were not able to be detected in a study of this size.

In general, most theory of planned behavior interventions in other domains have been lengthier than the ones used in this study and they incorporate a wider variety of intervention materials, targeting multiple behavioral antecedents at once and/or incorporating intervention activities that are not directly tied to the theory of planned behavior (e.g., goal setting, skill development, behavior modeling; Hardeman et al., 2002). Compared to these other interventions, this study was unique in the way it targeted and measured *individual* behavioral antecedents. However, that precision may have minimized impact of the interventions on the measured antecedent variables.

Further, a lack of participant engagement with the study materials could have also contributed to the interventions' lack of effects on behavioral antecedents. Previous research has found that the inclusion of engaging survey instructions can affect some careful responding outcomes (Ward & Meade, 2018). Ward and Meade found that including video-recorded instructions from researchers reduced self-reported careless responding. In this study, the intervention materials were originally included an interactive component – a quiz to assess whether participants had actually read the intervention materials. However, I removed this component after pilot test results suggested that it may have been harming attitudes towards careful responding in the intervention groups. While the testing component did not appear to have a positive effect on antecedents, it could have been the case that alternative interactive materials would have been effective. Different forms of interactive materials (e.g., video or audio recorded instructions, free response quiz questions rather than multiple choice) could have increased the salience of the interventions and induced greater changes in the behavioral antecedent variables.

Alternatively, it may have been the case that the interventions *did* affect behavioral antecedents, but that self-report items were not able to detect these differences. Although the study emphasized the importance of honesty on behavioral antecedent questions, it is possible that participants may not have felt comfortable providing their honest perspective on these items. In that case, differences across intervention and control groups on attitudes, norms, and perceived control would not have been detectable. This explanation seems to be at least partially supported this study's findings that at least one intervention group – the perceived behavioral control group – did respond more carefully than the control group on most metrics, but did not report higher perceived behavioral control.

Regardless of the explanation, the lack of differences across groups was surprising in light of the pilot test results. In the final round of pilot testing, all three intervention groups scored higher on the relevant antecedent than the control group. Although these effects were small in some cases, this study was designed with sample sizes large enough to detect relatively small differences across groups. However, when the full study was conducted, the observed effects were much smaller than the effects observed in pilot testing. Mean differences between the relevant intervention group and the control group did not approach significance for attitudes, norms, or behavioral control. It appears that the small pilot groups ( $n = 10$ ) were not representative of responses in the full study sample.

Ultimately, the findings from the first set of hypotheses show that my interventions were not effective at changing attitudes towards careful responding, subjective norms around careful responding, or perceived control over response behavior. To the extent that future research efforts are interested in affecting the theory of planned behavior's antecedent variables, different or stronger manipulations are recommended.

## **Intentions to Respond Carefully**

The second set of hypotheses examined the effects of the interventions on behavioral intentions. It was expected that each intervention would be successful at increasing intentions to respond carefully as compared to the control group, consistent with previous research applying the theory of planned behavior interventions in other contexts (Derzon & Lipsey, 2002; Sheeran et al., 2016). However, the results of this study revealed that there was not a significant difference between any intervention group and the control group on reported intentions to respond carefully. Thus, it appears the interventions were not effective at changing intentions to respond carefully.

As with the first set of hypotheses, it could have been the case that this study was not able to detect changes in intentions to respond carefully because the interventions were too brief or not engaging. Alternatively, the interventions may have had an impact on careful responding intentions, but the self-reported intention measure was not able to capture differences in intentions because participants did not feel comfortable providing honest responses. This explanation seems to be at least partially supported this study's findings that at least one intervention group – the perceived behavioral control group – did respond more carefully than the control group on most metrics, but did not report greater intentions to respond carefully.

Regardless, this study does not provide support for the notion that the interventions employed in this study are effective at changing intentions to respond (or not respond) carefully, at least not as measured here. To the extent that future studies aim to affect response intentions, these interventions (as carried out here) are not a viable method for doing so.

## Careless Responding Behavior

The final set of hypotheses examined the effects of the interventions on careless responding behavior. It was hypothesized that each intervention group would display less careless responding behavior as compared to the control group, consistent with previous research showing that theory of planned behavior interventions have been successful at changing behavior across a variety of contexts (Derzon & Lipsey, 2002; Sheeran et al., 2016). Here, careless responding behavior was measured using multiple metrics intended to capture various types of careless response patterns, as common in careless responding research (e.g., Kraiger et al., 2019; Ward & Meade, 2018; Ward & Pond, 2015).

The results of the omnibus analysis indicated that there were not overall differences in careless responding behavior across conditions. However, evidence from planned contrasts did reveal reductions in careless responding behavior for some intervention groups on some metrics. Because the omnibus test was not significant, these findings should be interpreted with caution.

Compared to the control group, the attitudes intervention group did not display less careless responding on most metrics. The only support found for this intervention was the replication of one of the three population-estimated correlations, the neuroticism-satisfaction correlation. Similarly, the subjective norms intervention did not display less careless responding on most metrics than the control condition. However, it did lead to greater self-reported carefulness and a closer replication of the population-estimated correlation between neuroticism and satisfaction.

Unlike the other two interventions, planned contrasts indicated that the perceived behavioral control intervention was effective at reducing careless responding behavior according to most careless responding metrics. There was less careless responding in the perceived

behavioral control intervention condition than in the control group according to the self-reported carelessness, intra-individual response variability, and even-odd consistency metrics, and the intervention group did replicate the population-level estimates of two out of three relationships (i.e., the satisfaction-commitment and neuroticism-satisfaction correlations) more closely than the control group. Overall, most of the evidence from planned contrasts provided support for the perceived behavioral control intervention. These findings are consistent with other theory of planned behavior interventions which find that experimentally-induced changes in self-efficacy (a component of behavioral control) have a larger impact on behavior than changes induced to attitudes or norms (Sheeran et al., 2016). However, because the overall omnibus test for differences in careless responding across conditions was not significant, these findings should be considered exploratory in nature.

Taken together, these results suggest that the perceived behavioral control intervention may have been effective at reducing careless responding, whereas the attitudes and subjective norms interventions were not. There are several plausible explanations for this. First, it may be the case that the particular perceived control intervention used in this study may simply have been more convincing to participants than the other interventions. It is plausible that participants may have believed that this study was designed to be as easy as possible to a greater extent than their counterparts believed that their individual response quality was important or that most of their peers respond carefully.

Further, while the perceived control intervention provided specific, actionable suggestions for participants about how to control their response quality, the attitudes and norms interventions focused on providing information that was intended to be persuasive rather than actionable. It may be the case that the actionable information is generally more effective for



changing behavior than is persuasive information. Additionally, the persuasive information used in the attitudes and norms intervention focused on the importance of careful responding in order to achieve societal outcomes (i.e., better working conditions for employees) in addition to individual outcomes (i.e., approval of work and compensation). It may have been the case that the focus on societal outcomes detracted from the salience of the individual-level outcomes associated with careful responding. If so, this would not have occurred in the perceived control intervention which did not refer to other MTurk workers or employees.

Alternatively, it may have been the case that the perceived behavioral control intervention included a broader range of information than did the other two interventions. All three interventions included information about how to obtain individual outcomes, that is, that work would be approved if the participant responded carefully. Beliefs about outcomes are associated primarily with the formation of attitudes (Kasprzyk & Montaño, 2015), so it may have been the case that, to some extent, the perceived behavioral control intervention served as a combination of an attitudes and behavioral control intervention.

Regardless of the explanation, this study did find tentative support for the perceived behavioral control intervention employed in this study. Broadly, these findings suggest that theory of planned behavior interventions may be effective at reducing careless responding behavior. Researchers who are interested in reducing careless responding behaviors within their samples should continue to consider and explore theory of planned behavior-based interventions, especially those targeting perceived control.

### **Manipulation Checks**

It is important to note that the manipulation check questions used in this study did not display the expected results. Although the results showed significant differences in whether a

participant believed they were shown manipulation materials across groups, a majority of all groups, including the control group, believed they were shown materials designed to convince them to respond carefully. Further, among those who believed they were shown manipulation materials, a majority of all four groups believed the materials were designed to convince them that careful responding was important (as opposed to careful responding being common or easy).

There are two potential reasons why these manipulation checks might have failed. First, there is the possibility that the manipulation check questions were confusing to participants. It may have been the case that participants were confused about what materials the manipulation check questions were referencing. Participants in all four groups were shown an informed consent prior to agreeing to participate in the survey on MTurk. Although the informed consent did not discuss response behavior, it did state the espoused purpose of the study (i.e., to learn about how employee personality affects the ways we view and interact with our work environment). For control group participants this may have been confused with an attempt to convince them to respond carefully.

Further, all four groups saw language encouraging them to respond honestly at two separate points in the study. First, they saw this wording the beginning of the survey when they were asked to answer behavioral antecedent and intentions questions, and then it was used again after the main survey as part of the self-reported careful response item (i.e., “This is for research purposes only, so we want you to be completely honest. Nothing you say could cause us to reject your work. Please let us know what you really think.”). To the extent that the participants perceived this request for honesty as an appeal for carefulness, it may explain why a large portion of them believed they had seen materials designed to influence them to respond carefully.

If it was the case that many of the participants believed that the informed consent or honestly wording was designed to influence them to respond carefully, it may explain their responses to the second manipulation check question. A majority of all four groups believed the materials were designed to convince them that careful responding was important as opposed to being common or easy. While the informed consent and honestly instructions don't state that careful responding is important, it may be more reasonable to assume that a researcher would include that information to convince participants that careful responding is important than to convince them that it is common or easy.

It does appear that the results of my manipulation check questions were similar to the results found in other careless responding research. My first question resembled an item assessed in multiple previous studies, "The instructions for this survey asked me to answer honestly and accurately" (Ward & Meade, 2018; Ward & Pond, 2015). In those papers, the authors were either not able to detect differences across groups (Studies 1 and 3 in Ward & Meade, 2018; Ward & Pond, 2015) or reported small effect sizes ( $\eta^2 = .02$  in Study 2 in Ward & Meade, 2018). My second manipulation check question was somewhat similar to another item assessed by Ward and Meade (2018), "I could guess the purpose of this study." Again for this item, the researchers were either not able to detect differences across groups (in Study 1 and 3) or found a small effect size ( $\eta^2 = .02$  in Study 2). If this explanation – that the questions were, in fact, confusing or unclear to participants – is true, then the manipulation check items used in this study may not have been a particularly sensitive way of determining if participants did or did not read the intervention materials.

However, it may also be the case that these surprising results are representative of true effects. That is, it may be the case that a sizable number of participants did not read instructions

and intervention materials, and therefore were not sure if they had seen information designed to convince them to respond accurately or what the topic of the information was. Alternatively, it may be the case that participants were reading the intervention materials, but that they simply did not retain the information long enough to recall it again at the end of the survey.

The results of this study provide mixed support for these two explanations. On one hand, at least one intervention group, the perceived behavior control group, showed reductions in careless responding behavior on most metrics. This would suggest that that intervention was at least somewhat effective, and that quality manipulation check items should have reflected this. However, the other two interventions were not effective, even though the study showed that their target behavioral antecedents were related to careless responding behavior. It could be the case that the interventions did not work because the participants were not attending to, retaining, or internalizing the information from the intervention materials.

Ultimately, it may be the case that both explanations played a role in producing the unexpected manipulation check results. These interventions were intentionally designed to be practical and brief, but that also means that they are relatively weak manipulations. It is likely that some participants were able to ignore or speed through the intervention. It may be the case that stronger manipulations would be more effective, and that this would be reflected in the manipulation check questions. However, it may also be the case that the manipulation check questions themselves were somewhat unclear, and that some participants providing incorrect manipulation check responses did, in fact, diligently read the intervention materials.

### **Careless Responding Metrics**

As part of my investigation into the theoretical understanding of careless responding behavior, I also examined the performance of the careless responding metrics used in this study.

This included both the relationships among the various careless responding metrics and the individual performance of specific metrics.

### **Relationships Between Careless Responding Metrics**

Previous research has found that there are different types of careless responders identified by various response patterns (Wertheimer, 2017). Different careless responding metrics are designed to identify responders that match different response patterns (Curran, 2016; Dunn et al., 2018; Meade & Craig, 2012). To the extent that different types of careless responders are included in the same sample, careless responding metrics should not necessarily be expected converge. That is, different metrics should identify different participants as having responded carelessly if the metrics are meant to detect different careless response patterns. Research using both simulated (DeSimone et al., 2018; Meade & Craig, 2012) and actual data (Dunn et al., 2018; Kraiger et al., 2019; Meade & Craig, 2012, Wertheimer, 2017) have supported the notion that careless responding indicators do not always converge.

However, the present study found that there were significant correlations between all of the careless responding indicators with the exception of one, Mahalanobis D, which was not related to most other metrics. Further, exploratory factor analyses found that the remaining four careless responding indicators were a good fit to a single careless responding factor. In part, the presence of these relationships may be due to a methodological difference between this study and those conducted in the past. Unlike most studies (Curran, 2016), the present study did not take a black and white approach to careless responding. Rather than setting arbitrary cut scores on the continuous careless responding metrics to indicate that someone either is or is not a careless responder, this study used the original continuous careless responding scores to indicate the extent to which a participant responded carefully. By not restricting variance in the careless

responding scores, it may have been easier to detect relationships between the careless responding metrics in this study. This may explain why other authors have been able to find convergence at high levels of careless responding, but not more moderate levels (DeSimone et al., 2018).

While significant relationships were found between most careless responding metrics used in this study, this is not to say that they were redundant. The relationships ranged from relatively weak to very strong. It appears that the various metrics are able to capture unique variance in response behavior, and thus, it is appropriate to use more multiple indicators of careless responding behavior, as others have recommended (e.g., DeSimone et al., 2018, Dunn et al., 2018).

In fact, correlations across careless responding behavior should be expected to vary depending on the particular metrics in question and their theoretical relationship with one another. Not all careless responding metrics are designed to catch careless responders with the same types of response behaviors and patterns. Among others, there are careless responding metrics designed to indicate responses with atypical consistency/variability (e.g., even-odd consistency, intra-individual response variability, longstring index), outlier responses (e.g., Mahalanobis D), implausible or incorrect responses (e.g., infrequency index, directed response indexes), and inappropriately speedy responses (e.g., response time; Curran, 2016; Meade & Craig, 2012). Careless responding metrics designed to capture the same types of response patterns or behaviors should be highly related to one another, as should metrics designed to capture response behaviors that tend to co-present in the same response. Whether metrics should or should not be expected to be highly correlated should depend greatly on the authors' choice of careless responding metrics.

That said, it may be less fruitful to examine multiple indicators that are very highly correlated (e.g., intra-individual response variability and even-odd consistency) than those that are less related if the purpose examining the metrics is to make decisions about which participants to exclude. If metrics are highly correlated, they are more likely to flag the same responders as careless, and thus be largely redundant.

Overall, the results of this study suggest that there is convergent validity evidence for most of the careless responding metrics used in this study when they are measured as continuous indicators of careless responding behavior. However, most correlations were moderately-sized, indicating that the metrics are not entirely redundant. It seems that it is appropriate to use multiple careless responding metrics to capture various types of careless responders.

### **Mahalanobis Distance**

While most careless responding metrics in this study were correlated with one another, there was one metric that failed to show significant correlations with most other careless responding metrics – Mahalanobis D. Further, exploratory factor analyses indicated that Mahalanobis D values did not load onto careless responding factors in the direction that was anticipated. This finding was consistent with previous research showing no significant relationship between Mahalanobis D scores and most other careless responding indicators (Kraiger et al., 2019; Ward & Meade, 2018; Ward & Pond, 2015). This study also found that no intervention had an impact on Mahalanobis D scores, including the perceived behavioral control intervention which had a significant effect on most other measured indicators of careless responding.

Unlike the other post-hoc metrics used in this study (i.e., response time, intra-individual response variability, and even-odd consistency) which are calculated in relation to an objective

ideal response (e.g., long response time, high response variability across diverse items, high consistency across theoretically related items), Mahalanobis D scores are calculated based on the relationship between an individual response and the entire sample of responses. That is, low D scores indicate that a response is more like other responses in the sample. It is assumed the those with low scores are more careful because most respondents are careful. To the extent that the underlying belief that most participants are very careful is inaccurate, lower D scores may not actually indicate more careful responding. Thus, this metric may be particularly problematic in samples with high rates of careless responding. If this were the case in my sample, it could explain the lack of relationships between Mahalanobis D scores and other careless responding metrics. I recommend caution when using the Mahalanobis D metric in samples that are expected to show substantial rates of careless responding.

### **Average Response Time Per Page**

Like the Mahalanobis D metric, average response time per page showed no differences across conditions in this study. However, response time did show the expected correlations with other careless responding metrics though they were mostly small to moderate in size.

The lack of differences in response time across conditions may, in part, be due to the frequency with which response time is used as an approval criterion on MTurk. It is common for researchers on MTurk to explicitly warn participants that they will not be compensated if they respond to the survey too quickly. Thus, participants on MTurk be particularly aware that they should not complete surveys too quickly, regardless of their intentions to respond or not respond carefully (Cheung et al., 2017). Therefore, this may be a particularly difficult metric to affect even if an intervention does have an impact on other indicators of careless responding. If this is the case, it may mean that response time is not a valid indicator of careful responding behavior at



least in MTurk samples. Therefore, I would not recommend relying solely on response time as a careless responding indicator in MTurk samples, as it may not be useful in detecting experienced careless responders.

### **Replication of Population Correlations as Careless Responding Metrics**

Unlike the other careless responding metrics examined in this study, estimated population correlations were examined at the group level. While careful responders should be expected to replicate true relationships more closely than careless responders do regardless of the relationship in question, the effects of careless responding will have on the magnitude and direction of the relationship will vary based on the nature of the specific variables involved in the correlation (McGonagle et al., 2016). That is, in some instances, the careless responders should be expected to inflate the relationships, and in other instances they should be expected to deflate the relationships. For this reason, my findings for this indicator of careless responding warrants additional discussion.

On one of the correlations examined in this study – the relationship between neuroticism and job satisfaction – there was a closer replication of the population effect in every intervention group as compared to the control group. In the population, the neuroticism-satisfaction correlation is estimated to be moderate and negative (Judge et al., 2002). Because this correlation is negative at a population-level, it may be particularly sensitive to distortion via straight line (e.g., selecting the same response repeatedly) or positive-side (e.g., alternating between positively-valenced response options) careless responding. Those types of careless responding can result in positive relationships between variables that are negatively correlated among careful responders (McGonagle et al., 2016). Because there is such a wide disparity in correlations between these types of careless responders and their careful counterparts, it may be the case that

even minor improvements in carefulness will result in visible changes in relationships between constructs that are negatively correlated at the population level.

For another of the examined correlations – the relationship between openness to experience and job satisfaction – there was not a closer replication of the estimated population-level correlation for any of the three intervention groups as compared to the control group. In the population, this correlation is estimated to be near-zero (Judge et al., 2002). However, in my study, I found significant, positive correlations in all four study groups, and no group's confidence interval included the .01 uncorrected population estimate. This may be because the studies included in meta-analyses could have included some data screening for carelessness prior to analysis that would have removed the straight line or positive-side careless responding that can inflate correlations between unrelated variables (McGonagle et al., 2016). This may partially explain the larger correlations found for all groups in my study.

Additionally, it may be the case that the correlations in my study shared more common method variance than did the studies included in the Judge et al. (2002) meta-analysis. In my study, both openness to experience and job satisfaction were measured via self-report at the same time point and using similar 7-point response scales. When true relationships between personality and attitude constructs are zero, it is estimated that their actual observed correlation will be inflated to .17 due to systematic and random measurement error (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). If the true relationship is slightly larger than zero, as Judge et al. (2003) estimated, observed correlations may be even larger than the .17 estimate. In my study, it may have been the case that shared systematic measurement error inflated correlations in all groups.

While no conditions' confidence interval included the population estimate, the group that came the closest was the control condition. This result was somewhat surprising, considering that one group – the perceived behavioral control group – responded more carefully than the control group according to most individual-level metrics. It may be the case that groups with more careless responders are less susceptible to some types of common method variance that inflates correlations for careful responders. For example, those who answer questions without regard to the question's content will provide answers that are not affected by their positive or negative affectivity or their tendency to embrace or avoid extreme responses. In other words, it may be the case that there was more truly random responding occurring in the control group, leading to a smaller correlation between these two variables in that condition.

These results could also be an artifact of the MTurk sample used in this study. It is plausible that those who are willing to complete paid online research studies may be more open to experience or have less satisfying (and lower paying) jobs on average than the broader population. It could be the case that the relationship between these two variables is different for the MTurk population, and that a more careful MTurk sample would not be more likely than a less careful MTurk sample to replicate population estimates.

Finally, the last correlation examined in this study – the correlation between job satisfaction and organizational commitment – showed different results across intervention groups. While there was a closer replication of the population estimate in the perceived behavioral control intervention group as compared to the control group, the same was not true for the other two interventions. This generally aligned with findings from other individual-level careless responding metrics (i.e., intra-individual response variability, even-odd consistency, and

self-reported carefulness) showing that the perceived behavioral control intervention did work for reducing careless responding behavior, while the other two interventions did not.

In general, I recommend taking great care when using the replication of estimated population correlations as a means for evaluating careless responding in a given sample. It is important to consider how the response patterns of careless responders should be expected to differ from that of careful responders for the particular variables being measured (McGonagle et al., 2016), as well as how common method variance should be expected to effect relationships between various types of variables (Podsakoff et al., 2003). All findings should be interpreted in light of the particular relationships being evaluated.

### **Future Research**

This study found that a motivational theory, the theory of planned behavior (Ajzen, 1991), was able to explain significant variance in careless responding behavior on surveys. In doing so, I found support for the notion that careless responding is at least in part a planned behavior, and that it is affected by motivational processes. This opens the door for future research to explore additional theories and variables that may explain additional variance in careless responding behavior. As is the case with other behaviors (Diefendorff & Chandler, 2011; Kanfer et al., 2017), it is likely that many motivational theories could shed light on the drivers of careless responding behavior.

In this study, interventions were developed in an attempt to affect the individual tenants of the theory of planned behavior. Although all three tested interventions had effects on at least one careless responding metric, none of the interventions were effective at reducing careless responding behavior across all metrics. It may be the case that alternate interventions targeted at the same behavioral antecedents could be more effective spurring behavior change among survey

participants. While my interventions were designed to be as easy to use as possible, other theory of planned behavior interventions are often designed to be lengthy and interactive in order to have the greatest chance at effecting behavior (Hardeman et al., 2002). It may be the case that careless responding interventions with these design features would have a greater effect on response behavior, as well as significant effects on behavioral antecedents and intentions. Future research should manipulate interventional characteristics to determine the best way to design theory of planned behavior interventions targeted at careless responding.

Additionally, future research should consider combining interventions targeting the three antecedents of the theory of planned behavior. My study attempted separated out the interventions to test the tenants of the theory of planned behavior with greater precision. However, theory of planned behavior interventions aimed at changing other types of behaviors typically target multiple antecedents at once (Hardeman et al., 2002). In the careless responding domain, it may be the case that a combined approach would have cumulative effects on response behavior. This idea is supported by the current study's finding that all three antecedent variables explained unique variance in careless responding intentions, and that careless responding intentions, in turn, predicted all careless responding behavior according to all metrics. These relationships are consistent with those found in meta-analytic research on the theory of planned behavior (Armitage & Connor, 2001). Therefore, additional research into the cumulative effects of the interventions tested in this study is warranted.

More broadly, I encourage additional research on the performance of the careless responding metrics used in this study. Based on the results of this study, I recommended caution when using the Mahalanobis D metric in samples with high rates of careless responding, as well as caution when using the response time metric in samples with experienced MTurk workers.

Additional research should further investigate the performance of these metrics in samples with varying levels of survey experience and intention to respond carefully. It may be the case that some careless responding metrics have varying validity based on sample characteristics.

Further, I join other researchers (e.g., Kraiger et al, 2019; McGonagle et al., 2016) in encouraging future research efforts to focus on providing clarity around which careless responding metrics should be used and how. While many researchers recommend using multiple careless responding metrics in any given study (e.g., Dunn et al., 2018; Huang, Bowling, et al., 2015; Meade & Craig, 2012), there is much less guidance on which specific metrics should be used and in combination with what other metrics. Additionally, for most metrics there is not clear guidance on how these metrics should be used (Curran, 2016; Kraiger et al, 2019; McGonagle et al., 2016). That is, are there particular cut scores at which participants should be removed from datasets? When multiple metrics are used, how should multiple scores be jointly considered to determine if participants' responses should be excluded from data analysis? Future research should attempt to address these gaps by clarifying the theoretical and empirical basis of each individual metric, as well as the relationships between various metrics.

### **Practical Implications**

The interventions developed and tested in this study were designed to be brief, non-invasive, and easy to add to the beginning of any survey. From a practical standpoint, there appears to be little downside to including the interventions tested in this study at the beginning of a survey. The individual interventions take only about an additional 30 seconds to read, and therefore do not have a sizable impact on overall survey time. Overall, the interventions appear to have little measurable effect on attitudes, norms, perceived control, or behavioral intentions before a study begins. However, at least one intervention – enhancing perceived control – led to

a pattern of more careful responding as compared to the control group. It appears that the benefits of using this intervention outweighs any minor drawbacks.

In light of these findings, I recommend that academic and applied researchers consider employing the perceived control intervention as part of the instructions on future surveys. The results of this study suggest that by telling participants that a survey was designed to be easy and giving them additional recommendations on how to make it even easier for themselves, we may be able to reduce careless responding among participants.

While I recommend the continued testing and application of my perceived behavioral control intervention in future research, in an organizational, community, or academic setting, there are additional considerations researchers may wish to reflect on before employing the interventions used in this study. First, these interventions were developed for use with MTurk workers, and the language used in the interventions reflects this. For use with non-MTurk participants, all three interventions would likely require some modification to ensure they are appropriate. For example, all three interventions use language that discusses the approval of work as the outcome of interest for participants. In other samples, this language would need to be modified to reflect an accurate outcome (e.g., awarded compensation, awarded extra credit, shared opinions). In addition, the language in the subjective norms intervention would need to be modified to reflect an accurate peer group (e.g., other participants, other students, other employees) rather than MTurk workers. Of course, as changes are made to interventional materials, testing should be conducted to ensure that interventions have the desired effects on careless responding behavior in the new sample.

Further, the nature of this sample was such as the survey administrator, I had had no previous interaction with the participants prior to data collection. In organizations and other

surveying contexts, this may not necessarily be the case. In practice, researchers should keep in mind that data collection does not take place in a vacuum. To the extent that organizations or research teams have eroded trust with survey participants in the past - perhaps by not implementing changes based on survey results, designing difficult to understand surveys, or promoting a culture in which responding to surveys is not valued – these interventions may have limited effects. Alternatively, building up perceptions that survey responses are important and easy to provide through other pre-survey communications, organizations may be able to bolster the effects of these interventions.

### **Strengths & Limitations**

There were a number of unique strengths and limitations associated with this study. First, I employed a study design that allowed me to collect pre-survey data (i.e., behavioral antecedents and response intentions), main survey data (i.e., filler measures for the calculation of careless responding metrics), and post-survey data (i.e., self-reported carefulness). Unlike some theory of planned behavior interventions which only collect antecedent data after the behavior has already been performed (or not performed; Hardeman et al., 2002), the current study design allowed me to measure variables at times that corresponded to their temporal order in the theory of planned behavior. Thus, I was able to examine relationships between variables measured before responding on the main survey began (i.e., behavioral antecedents and response intentions) and response behavior on the main survey itself.

Additionally, I employed an experimental design that allowed me to examine the effects of manipulating individual tenants of the theory of planned behavior. Although it may not have been possible to completely isolate information that would affect attitudes, norms, or perceived control, I was able to tailor the development of each intervention towards a single antecedent.



This type of design is somewhat unique, in that many theory of planned behavior interventions that aim to affect multiple antecedents do so simultaneously in the same interventional conditions (Hardeman et al., 2002). By targeting the individual antecedents, I was able to test the effects of various theory of planned behavior interventions separately.

However, there were also some disadvantages associated with this study's methods. First, this study utilized a sample of MTurk participants. Research generally suggests that MTurk samples are demographically similar to other types of samples (Berinsky, Huber, & Lenz, 2012; Casler, Bickel, & Hackett, 2013; Kraiger et al., 2019; Paolacci, et al., 2010) and that their careless responding rates are similar to those collected via other means (Kraiger et al., 2019; Paolacci et al, 2010). However, additional research should confirm that the findings of this study extend to organizational, student, community, or other online samples. As mentioned in the practical implications section, there may be some additional organizational-level concerns (e.g., organizational support for surveying, results of past survey efforts) that should be investigated prior to employing careless responding interventions in the context of any individual organization.

Additionally, this study used pairwise comparisons to draw conclusions about the effects of the individual interventions as compared to the control group. In some instances, there were not significant omnibus tests for the effect of condition on careless responding, but individual pairwise comparisons on the same dependent variable were significant. In these cases, it is possible significant findings may be due to Type 1 error as additional tests increase familywise error rates. In light of this risk, this study's findings should be interpreted cautiously. Additional research should attempt to confirm and replicate this study's findings in order to provide more robust support for the conclusions discussed here.

Another potential limitation in this study was the use of straightforward, single-item measures for behavioral antecedent, intentions, and self-reported carefulness. Single-item measures were chosen deliberately, in an attempt to avoid alerting participants (particularly control participants) that the purpose of this study was to examine careless responding behavior. While single-item measures are appropriate to use when the situation demands it, they may display poorer psychometric properties than multi-item scales (Fisher, Matthews, & Gibbons, 2015). It may have been the case that by using only one item, these scales were not able to capture the true essence or breadth of intended constructs. This study did not employ common measure development techniques (e.g., consulting outside subject matter experts to confirm content validity, collecting convergent and discriminant validity evidence to confirm construct validity) that could have shed light on the quality of these single-item measures.

On a related note, this study did not focus on exploring the relationships between measures of the theory of planned behavior's process variables (i.e., attitudes, norms, behavioral control, and intentions) in the context of careless responding behavior. It is possible that the items I used to measure these process variables are all indicative on a single underlying factor rather than four separate constructs. It could be the case that some overarching factor associated with being a good survey participant is actually the true predictor of careless responding behavior, as opposed to the individual antecedents in the theory of planned behavior. This notion may align with previous research suggesting individual difference variables predict careless responding (Bowling et al., 2016). A more thorough measure development process could be employed to more rigorously examine the relationships between each process variable and measures of theoretically distinct constructs. Future research should conduct a more thorough investigation into the psychometric quality of these measures or otherwise consider replicating

these findings with alternate measures that have established reliability and validity evidence to support their use.

### **Conclusion**

The aim of this study was to deepen the theoretical understanding of careless survey responding and to test several practical interventions designed to reduce careless responding behavior. This study found that the theory of planned behavior's tenants do explain significant variance in careless responding. This supports the notion that careless responding is a planned behavior affected by motivational processes.

Further, this study found that the perceived behavioral control intervention was effective at reducing careless responding behavior on most metrics as compared to a control group who received no intervention. However, the attitudes and norms interventions were not successful at changing response behavior. These findings suggest that the perceived behavioral control intervention may be useful for discouraging careless responding on future survey research.

## REFERENCES

- Albarracín, D., Johnson, B. T., Fishbein, M., & Muellerleile, P. A. (2001). Theories of reasoned action and planned behavior as models of condom use: A meta-analysis. *Psychological Bulletin*, 127, 142-161.
- Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology*, 63, 1-18.
- Allport, F. H. (1927). Self-evaluation: a problem in personal development. *Mental Hygiene*, 11, 570-583.
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40, 471-499.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32, 1-20.
- Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, N.J.: Prentice Hall.
- Banaji, M. R., & Heiphetz, L. (2010). Attitudes. In S. T. Fiske, D. T. Gilbert, G. Lindzey, M. R. Banaji, & L. Heiphetz (Eds.), *Handbook of Social Psychology*. Hoboken, NJ: John Wiley & Sons.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28, 117-148.

- Barber, L. K., Barnes, C. M., & Carlson, K. D. (2013). Random and systematic error effects of insomnia on survey behavior. *Organizational Research Methods, 16*, 616-649.
- Beach, D. A. (1989). Identifying the random responder. *Journal of Psychology: Interdisciplinary and Applied, 123*, 101-103.
- Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*, 800.
- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis, 20*, 351-368.
- Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science, 58*, 739-753.
- Berry, D. T., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment, 4*, 340-345.
- Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology, 111*, 218-229.
- Brower, C. K. (2018). *Too long and too boring: The effects of survey length and interest on careless responding* (Doctoral dissertation, Wright State University).
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.
- Cantril, H. (1944). *Gauging public opinion*. Princeton, NJ: Princeton University Press.

- Camus, K. A. (2015). *Once careless, always careless? Temporal and situational stability of insufficient effort responding (IER)* (Doctoral dissertation, Wright State University).
- Carrera, J. S., Brown, P., Brody, J. G., & Morello-Frosch, R. (2018). Research altruism as motivation for participation in community-centered environmental health research. *Social Science & Medicine*, 196, 175-181.
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29, 2156-2160.
- Caught, K., Shadur, M. A., & Rodwell, J. J. (2000). The measurement artifact in the organizational commitment questionnaire. *Psychological Reports*, 87, 777-788.  
doi:10.2466/pr0.2000.87.3.777
- Cave, K. & Daigle, K. C. (2017). *Openness to workplace health interventions: A scale development*. Unpublished manuscript.
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58, 1015-1026.
- Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, 32, 347-361.
- Converse, P. D., Peterson, M. H., & Griffith, R. L. (2009). Faking on personality measures: Implications for selection involving multiple predictors. *International Journal of Selection and Assessment*, 17, 47-60.

- Cooke, R., & French, D. P. (2008). How well do the theory of reasoned action and theory of planned behaviour predict intentions and attendance at screening programmes? A meta-analysis. *Psychology and Health, 23*, 745-765.
- Cooke, R., & Sheeran, P. (2004). Moderation of cognition-intention and cognition-behaviour relations: A meta-analysis of properties of variables from the theory of planned behaviour. *British Journal of Social Psychology, 43*, 159-186.
- Couper, M. P., & Miller, P. V. (2008). Web survey methods: Introduction. *Public Opinion Quarterly, 72*, 831-835.
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19.
- Derzon, J. H., & Lipsey, M. W. (2002). A meta-analysis of the effectiveness of mass-communication for changing substance-use knowledge, attitudes, and behavior. In W. D. Crano, M. Burgoon, & S. Oscamp (Eds.), *Mass media and drug prevention: Classic and contemporary theories and research*, (pp. 231-258).
- DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36*, 171-181.
- De Vaus, D. (2013). *Surveys in social research*. Crows Nest, Australia: Routledge.
- Diefendorff, J. M., & Chandler, M. M. (2011). Motivating employees. In S. Zedeck (Ed.), *Handbook of industrial and organizational psychology* (pp. 65–135). Washington, DC: American Psychological Association.
- Dunham, R. B., & Herman, J. B. (1975). Development of a female Faces Scale for measuring job satisfaction. *Journal of Applied Psychology, 60*, 629-631.

- Dunn, A. M., Heggstad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology, 33*, 105-121.
- Eden, D., & Aviram, A. (1993). Self-efficacy training to speed reemployment: Helping people to help themselves. *Journal of Applied Psychology, 78*, 352-360.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). G\*POWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*, 1-11.
- Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research, 15*, 195-219.
- Fishbein, M., & Yzer, M. C. (2003). Using theory to design effective health behavior interventions. *Communication Theory, 13*, 164-183.
- Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology, 21*, 3-21.
- Fleischer, A., Mead, A. D., & Huang, J. (2015). Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology, 8*, 196-202.
- Fogarty, G. J., & Shaw, A. (2010). Safety climate and the theory of planned behavior: Towards the prediction of unsafe behavior. *Accident Analysis & Prevention, 42*, 1455-1459.
- Fowler, F. J. (2013). *Survey research methods*. Thousand Oaks, CA: Sage.
- Furnham, A., Hyde, G., & Trickey, G. (2015). Personality and value correlates of careless and erratic questionnaire responses. *Personality and Individual Differences, 80*, 64-67.
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*, 349-360.



- Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, 132, 778-822.
- Gibson, A. (2016). *Stop with the questions already! The effects of questionnaire length and monetary incentives on insufficient effort responding* (Masters thesis, Wright State University).
- Hagger, M. S., Chatzisarantis, N. L., & Biddle, S. J. (2002). A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: Predictive validity and the contribution of additional variables. *Journal of Sport and Exercise Psychology*, 24, 3-32.
- Hardeman, W., Johnston, M., Johnston, D., Bonetti, D., Wareham, N., & Kinmonth, A. L. (2002). Application of the theory of planned behaviour in behaviour change interventions: A systematic review. *Psychology and Health*, 17, 123-158.
- Harrison, D. A., Newman, D. A., & Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Academy of Management Journal*, 49, 305-325.
- Hausenblas, H. A., Carron, A. V., & Mack, D. E. (1997). Application of the theories of reasoned action and planned behavior to exercise behavior: A meta-analysis. *Journal of Sport and Exercise Psychology*, 19, 36-51.
- Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48, 400-407.

- Hirschfeld, R. R. (2000). Does revising the intrinsic and extrinsic subscales of the Minnesota Satisfaction Questionnaire short form make a difference? *Educational and Psychological Measurement, 60*, 255-270.
- Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology, 30*, 299-311.
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*, 99-114.
- Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology, 100*, 828-845.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory: Versions 4a and 54*. Berkeley, CA: Institute of Personality Assessment and Research.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of Personality: Theory and Research, 3*, 114-158.
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*, 103-129.
- Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology, 92*, 107-127.
- Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*, 797-807.

- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127, 376-407.
- Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, 18, 512-541.
- Kanfer, R., Frese, M., & Johnson, R. E. (2017). Motivation related to work: A century of progress. *Journal of Applied Psychology*, 102, 338-355.
- Kasprzyk, D., & Montaño, D. E. (2007). Application of an integrated behavioral model to understand HIV prevention behavior of high risk men in rural Zimbabwe. In I. Ajzen, D. Albarracin, & R. Hornik (Eds.), *Prediction and change of health behavior: Applying the reasoned action approach*. Hillsdale, N.J.: Erlbaum.
- Kasprzyk, D., Montaño, D. E., & Fishbein, M. (1998). Application of an integrated behavioral model to predict condom use: A prospective study among high HIV risk groups. *Journal of Applied Social Psychology*, 28, 1557–1583.
- Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, 46, 141-155.
- Kim, M. S., & Hunter, J. E. (1993a). Attitude-behavior relations: A meta-analysis of attitudinal relevance and topic. *Journal of Communication*, 43, 101-142.
- Kim, M. S., & Hunter, J. E. (1993b). Relationships among attitudes, behavioral intentions, and behavior: A meta-analysis of past research, part 2. *Communication Research*, 20, 331-364.

- Kraiger, K., Sanchez, D. R., & McGonagle, A. K. (2019). *What's in a sample? Comparison of effect size replication and response quality across student, MTurk, and Qualtrics samples*. Manuscript submitted for publication.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50, 537-567.
- Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, 8, 142-164.
- Locke, E.A., & Latham, G.P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice.
- Manfreda, K. L., Berzelak, J., Vehovar, V., Bosnjak, M., & Haas, I. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 50, 79-104.
- Manning, M. (2009). The effects of subjective norms on behaviour in the theory of planned behaviour: A meta-analysis. *British Journal of Social Psychology*, 48, 649-705.
- Mayer, R. C., & Gavin, M. B. (2005). Trust in management and performance: who minds the shop while the employees watch the boss? *Academy of Management Journal*, 48, 874-888.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- McGonagle, A. K., Fisher, G. G., Barnes-Farrell, J. L., & Grosch, J. W. (2015). Individual and work factors related to perceived work ability and labor force outcomes. *Journal of Applied Psychology*, 100, 376-398.

- McGonagle, A. K., Huang, J. L., & Walsh, B. M. (2016). Insufficient effort survey responding: An under-appreciated problem in work and organisational health psychology research. *Applied Psychology, 65*, 287-321.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455.
- Meyer, J. P., Stanley, D. J., Herscovitch, L., & Topolnytsky, L. (2002). Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of Vocational Behavior, 61*, 20-52.
- Michel, J. S., O'Neill, S. K., Hartman, P., & Lorys, A. (2018). Amazon's Mechanical Turk as a Viable Source for Organizational and Occupational Health Research. *Occupational Health Science, 2*, 83-98.
- Middlestadt, S. E., Fishbein, M., Albarracin, D., Francis, C., Eustace, M. A., Helquist, M., & Schneider, A. (1995). Evaluating the impact of a national AIDS prevention radio campaign in St. Vincent and the Grenadines. *Journal of Applied Social Psychology, 25*, 21-34.
- Miller, V. D., Johnson, J. R., & Grau, J. (1994). Antecedents to willingness to participate in a planned organizational change. *Journal of Applied Communication Research, 22*, 59-80.
- Montaño, D. E., & Kasprzyk, D. (2015). Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. K. Glanz, B. K. Rimer, K. Viswanath (Eds.), *Health behavior: Theory, research and practice* (pp. 95-124). San Francisco, CA: Jossey-Bass.
- Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PloS One, 11*(6). <https://doi.org/10.1371/journal.pone.0157732>

- Nichols, D. S., Greene, R. L., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology, 45*, 239-250.
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality, 63*, 1-11.
- Notani, A. S. (1998). Moderators of perceived behavioral control's predictiveness in the theory of planned behavior: A meta-analysis. *Journal of Consumer Psychology, 7*, 247-271.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411-419.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.
- Prestwich, A., Sniehotta, F. F., Whittington, C., Dombrowski, S. U., Rogers, L., & Michie, S. (2014). Does theory influence the effectiveness of health behavior interventions? Meta-analysis. *Health Psychology, 33*, 465-474.
- Pew Research Center (2018). *Collecting survey data*.  
<http://www.pewresearch.org/methodology/u-s-survey-research/collecting-survey-data/>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003) Common method biases in behavioral research: A critical review of literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.
- Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The trans-situational influence of social norms. *Journal of Personality and Social Psychology, 64*, 104-112.

- Rich, B. L., Lepine, J. A., & Crawford, E. R. (2010). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal*, 53, 617-635.
- Riketta, M. (2008). The causal relation between job attitudes and performance: A meta-analysis of panel studies. *Journal of Applied Psychology*, 93, 472-481.
- Rios, J. A., Guo, H., Mao, L., & Liu, O. L. (2017). Evaluating the impact of careless responding on aggregated-scores: To filter unmotivated examinees or not? *International Journal of Testing*, 17, 74-104.
- Rogelberg, S. G., Church, A. H., Wacławski, J., & Stanton, J. M. (2008). Organizational survey research. In S. G. Rogelberg (Ed.), *Handbook of Research Methods in Industrial and Organizational Psychology* (pp. 141-160). Malden, MA: Blackwell Publishing.
- Rogelberg, S. G., Fisher, G. G., Maynard, D. C., Hakel, M. D., & Horvath, M. (2001). Attitudes toward surveys: Development of a measure and its relationship to respondent behavior. *Organizational Research Methods*, 4, 3-25.
- Rossi, P. H., Wright, J. D., & Anderson, A. B. (1983). Sample surveys: History, current practice, and future prospects. In P. H. Rossi, J. D. Wright, A. B. Anderson (Eds.), *Handbook of Survey Research* (pp. 1-20). New York, NY: Academic Press.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80, 1-28.
- Sackett, P. R. (2011). Faking in personality assessments: Where do we stand? In M. Ziegler, R. D. Roberts, C. MacCann (Eds.), *New perspectives on faking in personality assessment* (pp. 330-344). Oxford, UK: Oxford University Press.

- Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology, 102*, 254-273.
- Schmitt, N., Cortina, J. M., Ingerick, M. J., & Wiechmann, D. (2003). Personnel selection and employee performance. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology, Vol. 12* (pp. 565-593). Hoboken, NJ: Wiley.
- Shah, P. P. (1998). Who are employees' social referents? Using a network perspective to determine referent others. *Academy of Management Journal, 41*, 249-268.
- Sheeran, P., Maki, A., Montanaro, E., Avishai-Yitshak, A., Bryan, A., Klein, W. M., ... & Rothman, A. J. (2016). The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: A meta-analysis. *Health Psychology, 35*, 1178-1188.
- Smith, J. R., & Louis, W. R. (2008). Do as we say and as we do: The interplay of descriptive and injunctive group norms in the attitude-behaviour relationship. *British Journal of Social Psychology, 47*, 647-666.
- Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*, 240-261.
- Stutzer, A., & Lalive, R. (2004). The role of social work norms in job searching and subjective well-being. *Journal of the European Economic Association, 2*, 696-719.
- Vinokur, A. D., Price, R. H., & Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology, 23*, 39-74.



- Vroom, V. (1964). *Work and motivation*. New York: Wiley.
- Ward, M. K., & Meade, A. W. (2018). Applying social psychology to prevent careless responding during online surveys. *Applied Psychology: An International Review*, 67, 231-263.
- Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, 76, 417-430.
- Ward, M. K., & Pond III, S. B. (2015). Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys. *Computers in Human Behavior*, 48, 554-568.
- Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, 49, 737-747.
- Weiss, D. J., Dawis, R. V., England, G. W., & Lofquist, L. H. (1967). *Manual for the Minnesota Satisfaction Questionnaire*. Industrial Relations Center, University of Minnesota.
- Wertheimer, M. E. (2017). *Identifying the types of insufficient effort responders* (Doctoral dissertation, Middle Tennessee State University).
- Wildman, J. L., Bedwell, W. L., Salas, E., & Smith-Jentsch, K. A. (2010). Performance measurement at work: A multilevel perspective. In S. Zedeck (Ed.), *Handbook of industrial and organizational psychology: Vol. 1. Building and developing the organization* (pp. 303-342). Washington, DC: American Psychological Association.

- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28, 189-194.
- Wright, K. B. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10.
- Yzer, M. (2012). Perceived behavioral control in reasoned action theory: A dual-aspect interpretation. *The Annals of the American Academy of Political and Social Science*, 640, 101-117.

## APPENDIX A: INTERVENTION MATERIALS

### Study 1 Attitudes Intervention Materials

Congratulations, you qualify to participate!

**It is very important to us that you read the questions in this survey and respond carefully.** We are hoping to learn what makes employees happy and to use the results of the study to convince companies to improve the workplace for their employees. This will only be possible if we get real information from the people who take this survey.

It is very important to us that you read the questions in this survey and respond carefully. **As long as you read each question before you answer, your work will 100% be approved.** In order for your work to be approved, you must respond carefully. We truly appreciate your time and effort. You are helping to make work a better place!

----

### Study 2 Subjective Norms Manipulation Materials

Congratulations, you qualify to participate!

Before you proceed to the study, we wanted to provide you with some information about how your fellow MTurk workers are answering survey questions on MTurk.

- **Research has found that over 90% of MTurk workers carefully read and respond to all survey questions.**<sup>1</sup>
- When asked about why it is so important to respond to surveys carefully, MTurk workers respond with two important reasons. First, it is important to respond carefully because many requestors check for careful responding before approving work. Second, careful responding is important because it helps researchers answer important scientific questions about how people think and behave.<sup>2</sup>

It is important to us that you take the time to answer this survey as carefully and accurately as other MTurk workers do. Please take the time to read all survey instructions, questions, and response choices before selecting the most accurate response. As long as you take the time to read questions carefully and provide accurate responses, your work will be approved.

---

<sup>1</sup> Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PloS One*, 11(6).

<sup>2</sup> Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, 43, 800.

### Study 3 Perceived Behavioral Control Manipulation Materials

Congratulations, you qualify to participate!

**This survey is designed so that it will be as easy as possible for you to provide accurate responses.** We have kept the survey as short as possible, made sure the survey instructions are clear, and written straightforward questions.

You can also take other steps to make the survey as easy as possible for yourself:

- Remove distractions from your environment. It may help to go to a quiet area, turn off any TV or music, or remove any other distractions that you believe may get in the way of you providing careful responses.
- Take a break from the survey if you are interrupted or if you find yourself becoming distracted. You will not be penalized for taking breaks or taking a long time to respond to the survey.

As long as you take the time to read questions carefully and provide accurate responses, your work will 100% be approved. Once again, thank you so much for participating in this study. Please select the forward arrow to proceed to the survey.

## APPENDIX B: SURVEY ITEMS

All items answered on a 7-point Likert-type scale unless otherwise indicated.

### Inclusion Criteria

1. What is your age? \_\_\_\_\_
2. Are you currently employed outside of MTurk?
  - a. Yes, I am employed full time (30+ hours a week)
  - b. Yes, I am employed part time (less than 30 hours a week)
  - c. No, but I am looking for a job
  - d. No, I am not employed outside of MTurk

### Theory of Planned Behavior Variables

*Before you start the survey, we want to ask you a few questions about how you feel about taking surveys. This is for research purposes only, so we want you to be completely honest. Nothing you say could cause us to reject your work. Please let us know what you really think !*

1. I think it is important to read and respond to survey questions carefully.
2. Most MTurk workers respond carefully to surveys.
3. It will be easy to respond carefully to this survey.
4. I plan to respond carefully to this survey

### Personality: Big Five Inventory

*Please select the response option that best describes the extent to which you agree or disagree with the statement.*

*I am someone who...*

1. Is talkative
2. Tends to find fault with others
3. Does a thorough job
4. Is depressed, blue
5. Is original, comes up with new ideas
6. Is reserved
7. Is helpful and unselfish with others
8. Can be somewhat careless
9. Is relaxed, handles stress well
10. Is curious about many different things
11. Is full of energy
12. Starts quarrels with others

13. Is a reliable worker
14. Can be tense
15. Is ingenious, a deep thinker
16. Generates a lot of enthusiasm
17. Has a forgiving nature
18. Tends to be disorganized
19. Worries a lot
20. Has an active imagination
21. Tends to be quiet
22. Is generally trusting
23. Tends to be lazy
24. Is emotionally stable, not easily upset
25. Is inventive
26. Has an assertive personality
27. Can be cold and aloof
28. Perseveres until the task is finished
29. Can be moody
30. Values artistic, aesthetic experiences
31. Is sometimes shy, inhibited
32. Is considerate and kind
33. Does things efficiently
34. Remains calm in tense situations
35. Prefers work that is routine
36. Is outgoing, sociable
37. Is sometimes rude to others
38. Makes plans and follows through
39. Gets nervous easily
40. Likes to reflect, play with ideas
41. Has few artistic interests
42. Likes to cooperate with others
43. Is easily distracted
44. Is sophisticated in art, music, or literature

### **Job Satisfaction: Minnesota Satisfaction Questionnaire**

*Please rate how satisfied you are with each component of your primary job (not your work done on MTurk).*

1. Being able to keep busy all the time.
2. The chance to work alone on the job.
3. The chance to do different things from time to time.
4. The chance to be “somebody” in the community.
5. The way my boss handles his/her workers.
6. The competence of my supervisor in making decisions.
7. Being able to do things that don’t go against my conscience.
8. The way my job provides for steady employment.

9. The chance to do things for other people.
10. The chance to tell people what to do.
11. The chance to do something that makes use of my abilities.
12. The way company policies are put into practice.
13. My pay for the amount of work I do.
14. The chances for advancement on this job.
15. The freedom to use my own judgment.
16. The chance to try my own methods of doing the job.
17. The working conditions.
18. The way my co-workers get along with each other.
19. The praise I get for doing a good job.
20. The feeling of accomplishment I get from the job.

### **Organizational Commitment: Allen & Meyer 1990**

*Please select the response option that best describes the extent to which you agree or disagree with the statement about your current employer (this should be your primary employer - not MTurk).*

1. I would be very happy to spend the rest of my career with this organization
2. I enjoy discussing my organization with people outside it
3. I really feel as if this organization's problems are my own
4. I think that I could easily become as attached to another organization as I am to this one – R
5. I do not feel like 'part of the family' at my organization – R
6. I do not feel 'emotionally attached' to this organization – R
7. This organization has a great deal of personal meaning for me
8. I do not feel a strong sense of belonging to my organization – R
9. I am not afraid of what might happen if I quit my job without having another one lined up – R
10. It would be very hard for me to leave my organization right now, even if I wanted to
11. Too much in my life would be disrupted if I decided I wanted to leave my organization now
12. It wouldn't be too costly for me to leave my organization now – R
13. Right now, staying with my organization is a matter of necessity as much as desire
14. I feel that I have too few options to consider leaving this organization
15. One of the few serious consequences of leaving this organization would be the scarcity of available alternatives
16. One of the major reasons I continue to work for this organization is that leaving would require considerable personal sacrifice — another organization may not match the overall benefits I have here
17. I think that people these days move from company to company too often
18. I do not believe that a person must always be loyal to his or her organization – R
19. Jumping from organization to organization does not seem unethical to me – R
20. One of the major reasons I continue to work for this organization is that I believe that loyalty is important and therefore feel a sense of moral obligation to remain

21. If I got another offer for a better job elsewhere I would not feel it was right to leave my organization
22. I was taught to believe in the value of remaining loyal to one organization
23. Things were better in the days when people stayed with one organization for most of their careers
24. I do not think that remaining at one company for most of your career is sensible anymore  
– R

### Openness to Workplace Health Programs: Cave & Daigle 2017

1. If a health promotion program was offered by my employer, I would be supportive of it.
2. I like the idea of my employer offering health promotion programs
3. My employer should offer health promotion programs
4. Work settings are an appropriate place for health promotion programs
5. I would appreciate my employer expressing interest in health promotion programs
6. I would participate in workplace health promotion programs if they were offered
7. I **do not** see the reason for my employer to offer health promotion programs – R
8. Health promotion programs **should not** take place at work – R
9. Health promotion programs at work are an invasion of privacy – R
10. It is **inappropriate** for my employer to offer health promotion programs at work – R
11. I would like to participate in health promotion programs if my workplace offered them
12. I would encourage my coworkers to participate in health promotion programs if my workplace offered them
13. I support workplace health promotion programs
14. Employers should only promote external health promotion programs (e.g. discounted gym memberships), not internal ones (e.g. weight loss or fitness challenges in the office)  
– R

### Attitudes towards Workplace Health Programs

1. I would benefit from participating in workplace health promotion programs
2. Health promotion programs would benefit others in my organization
3. Health promotion programs would benefit my organization
4. I believe health promotion programs benefit employees
5. I believe health promotion programs lead to positive outcomes
6. It is beneficial for employees to participate in health promotion programs at work
7. I **do not** see much value in workplace health promotion programs (R)
8. I **do not** see the value in having health promotion programs at work (R)
9. I believe workplace health promotion programs are **ineffective** (R)
10. Participating in health promotion programs would **not** bring any benefit to people in my organization (R)
11. I would **not** benefit from participating in health promotion programs (R)
12. It would **not** benefit my organization to offer health promotion programs to employees (R)
13. The benefits of offering health promotion programs would outweigh the costs



### **Openness to Organizational Change: Miller et al., 1994**

1. I would consider myself to be “open” to changes brought to my work role
2. Right now, I am somewhat resistant to changes at work (R)
3. I look forward to changes in my work role
4. I am quite reluctant to consider changing the way I now do my work (R)
5. I think that the implementation of changes at work can have a positive effect on how I accomplish my work
6. From my perspective, changes at work tend to be for the better
7. Changes at work tend to be for the worse in terms of the way I get my work done (R)
8. I think that changes at work tend to have a negative effect on how I perform my role in the organization (R)

### **Trust in Management: Mayer & Gavin, 2005**

1. If I had my way, I wouldn't let top management have any influence over issues that are important to me (R)
2. I would be willing to let top management have complete control over my future in this company
3. I really wish I had a good way to keep an eye on top management (R)
4. I would be comfortable giving top management a task or problem which was critical to me, even if I could not monitor their actions
5. I would tell top management about mistakes I've made on the job, even if they could damage my reputation
6. I would share my opinion about sensitive issues with top management even if my opinion was unpopular
7. I am afraid of what top management might do to me at work (R)
8. If top management asks why a problem happened, I would speak freely even if I were partly to blame
9. If someone questioned top management's motives, I would give top management the benefit of the doubt
10. If top management asked me for something, I respond without thinking about whether it might be held against me

### **Careless Responding Behavior: Self-Reported Carefulness**

Before you finish this study, we have a few more questions to ask you.

These questions are for research purposes only - nothing you say for the rest of the survey could cause us to reject your work! Please answer as honestly as possible. We want to know what you really think and feel!

1. Remember, we want you to be honest! Even if you respond with a 1, your work will be approved.

In your honest opinion, how carefully do you think you responded to this survey?

- a. 1: I responded as quickly as I possibly could, without reading the instructions or questions
- b. 2
- c. 3
- d. 4: I was careful some of the time, but other times I sped through questions or answered while I was distracted
- e. 5
- f. 6
- g. 7: I took my time on every single item, and I responded as carefully as I possibly could

### **Manipulation Checks**

1. At the beginning of this survey, were you shown a page of information designed to convince you that you should read and respond carefully to the survey questions?
  - a. Yes
  - b. Not sure
  - c. No

[IF PARTICIPANT ANSWERS YES TO #1, then #2, #3, and #4 ARE DISPLAYED; IF PARTICIPANT ANSWERS NO TO #1, then #5 IS DISPLAYED]

2. Do you think the information you were asked to read was primarily designed to convince you that...
  - a. Careful responding will lead to payment
  - b. Other MTurk workers respond carefully so you should too
  - c. It would be easy to provide careful responses
3. Do you think the information you read influenced you to respond more carefully than you otherwise would have?
  - a. 1: Definitely not
  - b. 2: Probably not
  - c. 3: Maybe not
  - d. 4: Not sure
  - e. 5: Maybe yes
  - f. 6: Probably yes
  - g. 7: Definitely yes
4. After reading that information, do you think you planned to respond carefully?
  - a. 1: Definitely not
  - b. 2: Probably not
  - c. 3: Maybe not
  - d. 4: Not sure
  - e. 5: Maybe yes
  - f. 6: Probably yes
  - g. 7: Definitely yes

5. When you first began this survey, do you think you planned to respond carefully?
- a. 1: Definitely not
  - b. 2: Probably not
  - c. 3: Maybe not
  - d. 4: Not sure
  - e. 5: Maybe yes
  - f. 6: Probably yes
  - g. 7: Definitely yes

## Demographics

1. Approximately how long have you been working as an MTurk worker?
- a. Less than a week
  - b. 1-4 weeks
  - c. 1-12 months
  - d. 1-3 years
  - e. 3+ years
2. Approximately how many MTurk HITs have you completed? If you cannot make a reasonable guess, you may skip this question.
- \_\_\_\_\_
3. Approximately what percentage of the MTurk HITS you've completed have been approved? If you are not sure, you may skip this question.
- \_\_\_\_\_
4. What gender do you identify with?
- a. Male
  - b. Female
  - c. Other
5. What racial or ethnic groups do you identify with?
- a. Asian/Pacific Islander
  - b. Black/African-American
  - c. Hispanic or Latino
  - d. Native American
  - e. White/Caucasian
  - f. Multiracial or Multiethnic
  - g. Other
6. What industry do you work in? \_\_\_\_\_
7. What is your job title? \_\_\_\_\_
8. Do you supervise others as part of your job?

- a. Yes
- b. No