# AUTOMATICITY AS A HIDDEN COST OF EXPERTISE: SITUATIONAL AND INDIVIDUAL-DIFFERENCE FACTORS UNDERPINNING ERRORS OF AUTOMATICITY

By

Alexander P. Burgoyne

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology—Doctor of Philosophy

2019

ABSTRACT

AUTOMATICITY AS A HIDDEN COST OF EXPERTISE:
SITUATIONAL AND INDIVIDUAL-DIFFERENCE FACTORS UNDERPINNING
ERRORS OF AUTOMATICITY

By

Alexander P. Burgoyne

A traditional view of automaticity holds that the ability to perform well-practiced skills without attention is adaptive because it frees mental resources to process other information. Without denying the benefits of automaticity, I show how it can also lead to error in domains of expertise such as driving, medical diagnosis, problem solving, and reading. I begin by discussing automaticity within the context of two theoretical frameworks, classical theories of skill acquisition and dual process theory. I then examine situational and individual-difference factors that make errors of automaticity more likely to occur. Next, using proofreading as a testbed, I demonstrate how knowledge, expectations, and other top-down constraints influence reading behaviors and comprehension. I then present two experiments to investigate the *self-generation effect* in proofreading: the hypothesis that it is more difficult to detect mistakes in one's own writing than in the writing of others. The reasoning behind this hypothesis is that overfamiliarity with self-generated text increases the probability that errors are overlooked or seen but undetected. Finally, I discuss implications of the research and argue that understanding the benefits and consequences of automaticity is critical to improve decision-making outcomes across a wide range of applied contexts.

This dissertation is dedicated to my mother, Tova Shaban,
my father, Robert Burgoyne, and my brother, Brian Burgoyne.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

## INTRODUCTION

One of the hallmarks of expertise is automatic processing. As we become highly skilled in complex tasks, we perform them with increasing facility until attention seems to be withdrawn altogether. Examples range from the mundane (e.g., walking to campus while holding a conversation) to the arcane (e.g., sight-reading piano music while shadowing continuous speech; Allport, Antonis, & Reynolds, 1972). Automatic processes are cognitive operations that can be performed without attention and are thus unaffected by cognitive load (Anderson, 1982; Evans, 2008; Fitts & Posner, 1967). That is not to say that all tasks performed by experts within their domain of skill are performed automatically but, rather, that many proceduralized tasks *can* be performed automatically with sufficient training (Fitts & Posner, 1967).

The cognitive processes that support expert performance are closely linked to those that support automaticity. As a case in point, classical theories of skill acquisition characterize the final stage of learning as proceduralized, autonomous, and effortless (Anderson, 1982; Fitts & Posner, 1967). However, although automaticity is often adaptive, it may come with a cost and lead to mistakes. That is, the same cognitive mechanisms that support expert performance may also contribute to the likelihood of error. This “hidden cost of expertise” — the errors that result from automatic processing — is the focus of the present work.

In this dissertation, I investigate the potential for automatic processing to cause errors, the factors that make errors more likely to occur, and what can be done to reduce or even eliminate them. I begin by reviewing the role of automaticity in theories of skill acquisition and information processing, highlighting common themes and important distinctions across 50 years of research and theory. I then discuss automaticity as a source of expert error and the situational and individual-difference factors that increase the probability of mistakes. Next, using proofreading as a testbed, I examine the cognitive processes underlying reading and proofreading and present two experiments designed to evoke errors of automaticity in the laboratory. Finally, I discuss implications of the research and argue that understanding the benefits and consequences of automaticity is critical to improve decision-making outcomes across a wide range of applied contexts.

# AUTOMATICITY

Scientists and laypeople alike have long been interested in automatic processes, with scientific interest dating back at least to William James (1890). In fact, James (1890) quotes Henry Maudsley (1868), a British psychiatrist, who argues that much of everyday life might be characterized by automaticity:

> If an act became no easier after being done several times, if the careful direction of consciousness were necessary to its accomplishment on each occasion, it is evident that the whole activity of a lifetime might be confined to one or two deeds – that no progress could take place in development. (p. 155)

This view has remained popular among some contemporary researchers. For example, the idea that automaticity pervades daily life is echoed by the title of Bargh and Chartrand's (1999) highly cited *American Psychologist* article, "The Unbearable Automaticity of Being."

Cognitive processes that are described as "automatic" tend to meet a number of criteria. Jonides, Naveh-Benjamin, and Palmer (1985) defined automatic processes as those that are free from demands of processing capacity, difficult to alter or inhibit with voluntary control, and characterized by stereotypy, or marked similarity across repetitions. Along the same lines, Schneider and Shiffrin (1977) conceived of automatic processes as those that can operate without attention, are unaffected by capacity limitations, and can operate in parallel, while Fitts and Posner (1967) explained that automatic or autonomous processes are those that can be executed without attention.

There is general agreement that automaticity develops as a function of experience (Anderson, 1982; Fitts & Posner, 1967). More specifically, the consistent pairing of stimulus inputs and response outputs in the form of motor behaviors, stimulus discriminations, judgments, and decisions can lead to the automatic elicitation of learned responses (Stanovich, West, & Toplak, 2011). As Kahneman and Klein (2009) stated, the strengthening of input-output associations requires a "high-validity environment" (p. 519). That is, for automaticity to develop, cues from the environment must be predictive of outcomes and therefore useful to the performer. In a perfectly valid environment, the same input paired with the same response results in the same outcome every time.

**Theories of Skill Acquisition**

In classical theories of skill acquisition, the process is characterized by different stages of learning. In the early stages, performance of a skill is cognitively demanding, but with practice it becomes increasingly automatic (Fitts & Posner, 1967). Specifically, Fitts and Posner (1967) delineated three stages of skill acquisition: a cognitive stage, an associative stage, and an autonomous stage.

During the cognitive stage, the performer's primary goal is to "understand the task and what it demands" (Fitts & Posner, 1967, p. 11). That is, the performer must actively attend to instruction, events, and outcomes related to the execution of the skill. Typically, beginners' skill execution is "slow, nonfluent, and error-prone" (Gray, 2004, p. 42). Errors may arise because the performer misunderstands the skill or lacks the requisite capacity. For example, they may be unable to maintain all of the relevant instructions in working memory (Anderson, 1982). Anderson (1982) states that verbal mediation, such as recalling and reciting instructions aloud, is common during the cognitive stage because declarative knowledge about the task must be maintained in working memory to influence performance. Performance is inconsistent in this first stage of skill acquisition as learners test different strategies, retaining those that work and discarding those that do not (Schmidt & Lee, 2005).

The second stage of skill acquisition, the associative stage, represents a transition from the cognitively-demanding first stage towards the autonomous final stage. In the associative stage, knowledge about the skill is converted from a propositional network of facts into a procedure, that is, an ordered sequence of steps. When a stimulus is presented, the performer retrieves the appropriate procedure stored in long-term memory and executes the skill (Anderson, 1982). With practice, associations between the steps are formed and strengthened such that performance of a given step cues performance of the next. However, because the associations are still being learned, attention is necessary to keep one's place and ensure that each step is performed correctly. Speed of performance increases and errors are typically reduced in the associative stage (Fitts & Posner, 1967; Tenison & Anderson, 2016; Wiedenbeck, 1985).

In the final stage, the autonomous stage, performance of the skill becomes automatic. An automated skill is one that is driven by procedural knowledge, can operate without much (or any) attention, places little demand on working memory, and is less susceptible to interference due to other ongoing activities or distractions (Beilock, Wierenga, & Carr, 2002; Fitts & Posner, 1967; Schmidt & Lee, 2005). Fitts and Posner (1967) note that "there is a good deal of similarity between highly practiced skills and reflexes" (p. 15). Along the same lines, Anderson (1982) suggests that this stage reflects a transition from "stimulus-retrieval-response" to a "stimulus-response" process (Tenison & Anderson, 2016). In other words, in this stage, stimuli automatically cue a performer's response without requiring effortful search for the relevant procedure stored in long-term memory. Skills in the autonomous phase continue to increase in speed and efficiency as a function of training, but the benefits of practice are subject to diminishing returns (Fitts & Posner, 1967). Because attention is no longer required for execution of the skill, performers may be able to simultaneously attend to other tasks without loss in accuracy, as in the earlier example of sight-reading piano music while shadowing (i.e., listening to and repeating aloud) continuous speech (Allport et al., 1972).

Evidence for the automaticity of well-practiced skills is provided by experiments that manipulate experts' and novices' focus of attention during performance. In the dual-task expert/novice paradigm, people attempt to perform a motor skill while either attending to aspects of the skill's execution or to a distractor task, such as detecting a tone within an auditory stream. Three predictions follow from Fitts and Posner's (1967) theory of skill acquisition that are relevant to the role of attention in procedural task performance. First, experts' performance will be largely unaffected by a distractor task because execution of an automated motor skill requires little to no attention (Beilock et al., 2002; Gray, 2004). Second, novices' performance will deteriorate when given a distractor task because in the early stages of skill acquisition one must pay attention to declarative knowledge about the skill (e.g., instruction) in order to properly perform it (Anderson, 1982). Third, experts' performance will deteriorate when they focus on specific aspects of a skill because attending to its step-by-step execution undermines its proceduralization and automaticity (Anderson, 1982; Baumeister, 1984; Fitts & Posner, 1967).

All three predictions have received support from the dual-task expert/novice paradigm in domains ranging from hockey puck handling and soccer dribbling to golf putting and baseball batting (Beilock et al., 2002; Gray, 2004; Leavitt, 1979; Smith & Chamberlin, 1992). For example, expert baseball players' batting performance is unaffected by a distractor task, suggesting that attention is not necessary for the accurate execution of well-practiced skills (Gray, 2004). As another example, novices' soccer dribbling suffers greatly under dual-task conditions, whereas experts are hardly affected (Smith & Chamberlin, 1992). This reinforces the idea that attention is more critical when performing novel tasks than well-practiced tasks.

Finally, skilled golfers show substantial performance decrements when asked to attend to specific mechanics related to the swing of the golf club, suggesting that attention can actually disrupt the performance of automated skills (Beilock et al., 2002).

Taken together, the results of these studies provide evidence for the skill acquisition frameworks described by Fitts and Posner (1967) and Anderson (1982) and the role of automaticity in expertise. That is, performance of a skill is initially cognitively demanding and requires attention, but after sufficient practice, can become automated such that it no longer requires attention. This enables experts to focus on higher-level aspects of task performance during skill execution. For example, orchestral musicians can play their instruments while watching the conductor for cues on tempo and dynamics (Berlioz, 1915).

**Dual Process Theory**

*Dual process theory* provides a more contemporary account of the role of automaticity in information processing. Dual process theory posits that cognitive operations can be categorized into one of two modes. *Type 1 processes* are described as fast, parallel, automatic, unconscious, intuitive, and seemingly effortless, whereas *Type 2 processes* are described as slow, sequential, controlled, conscious, reflective, and effortful (Evans, 2008; Evans & Stanovich, 2013; Frankish, 2010; Norman et al., 2017). Dual process theory has been applied to domains such as judgment and decision making, learning and memory, and social judgment, to name a few (Evans, 2008).

Although numerous attributes have been ascribed to Type 1 and Type 2 processes, Evans (2008) and Stanovich and Toplak (2012) argue that only some of these attributes are defining characteristics; the rest are merely correlates. The need to distinguish between defining features and incidental correlates of Type 1 and Type 2 processes was motivated in part by criticisms of dual process theory. For example, Keren and Schul (2009) argued that many of the attributes that have been used to describe Type 1 or Type 2 processes do not always co-occur (see Kruglanski & Gigerenzer, 2011, for a similar argument). Stanovich and Toplak (2012) countered that the attributes used to describe Type 1 and Type 2 processes should be thought of as “family resemblances” rather than “necessarily co-occurring properties” (p. 5).

As Stanovich and Toplak (2012) stated, the defining characteristic of Type 1 processing is *autonomy*. That is, Type 1 processes do not require controlled attention and, as a result, place minimal demands on working memory capacity (Evans & Stanovich, 2013). Stanovich and Toplak (2012) suggest that Type 1 processes can include “innately specified processing modules” and “experiential associations that have been learned to automaticity” (p. 8). For example, stimulus discriminations that have been practiced extensively, such as determining whether a character is a letter or digit (Schneider & Shiffrin, 1977), can operate as Type 1 processes (Evans & Stanovich, 2013). Other examples of Type 1 processes include completing the phrase “bread and ______,” answering the question “2 + 2 = ___,” reading words on large billboards, detecting hostility in a voice, and recognizing the face of one’s mother (Kahneman, 2011, p. 21).

To link Type 1 processing to automaticity as described by Fitts and Posner (1967), consider the execution of an autonomous skill. Performing an autonomous skill can be considered an operation of Type 1 processing because it no longer requires attention and places minimal demands on working memory capacity (Anderson, 1982; Fitts & Posner, 1967). Indeed, fully proceduralized autonomous skills align well with the description of “experiential associations that have been learned to automaticity” (Stanovich & Toplak, 2012, p. 8) and with the broader class of Type 1 processes described by Evans and Stanovich (2013).

Returning to dual process theory, Evans (2008) argues that the defining characteristic of Type 2 processing is the engagement of attention control and working memory capacity. Examples include calculating the answer to the question "17 × 24 = ____," filling out a tax form, checking the validity of a complex logical argument, and comparing apartments in a new city (Kahneman, 2011, pp. 20-22). Hypothetical thinking is also considered a Type 2 process because it requires maintaining information in an active state and manipulating it in service of a goal. Hypothetical thinking is an example of *cognitive decoupling* – maintaining a secondary representation of a real-world state for the purposes of hypothesis testing and simulation. As an aside, the term "decoupling" appears to have first been used by Leslie (1987) to describe how children engage in "pretend play" (p. 416). Because Type 2 processes such as cognitive decoupling rely on working memory, they are typically "slow, sequential, and correlated with measures of general intelligence" (Evans & Stanovich, 2013, p. 235).

Type 2 processing can also be reconciled with Fitts and Posner's (1967) theory of skill acquisition and, in particular, to the early stages of learning. Consider the role of working memory and attention control when attempting to learn a new skill. Beginners must maintain instructional material presented to them in working memory and must attend to the execution of the skill and the resulting outcome (Anderson, 1982; Fitts & Posner, 1967). This cognitively-demanding stage of skill acquisition meets the essential criteria of Type 2 processing, requiring both attention and working memory. Performance of skills in the early stages of learning can also be described using many of the same terms as Type 2 processes: slow, sequential, controlled, conscious, reflective, and effortful. And with practice, an initially cognitively-demanding skill can become automated, just as Type 1 processes emerge when experiential associations are learned to the point of automaticity (Stanovich & Toplak, 2012).

Three lines of research provide evidence for the distinction between Type 1 and Type 2 processes: experimental manipulations designed to evoke Type 1 or Type 2 processes; neuroimaging studies that distinguish between patterns of activation presumably reflecting Type 1 or Type 2 processes; and individual-difference research showing that Type 2 processing correlates positively with cognitive ability but that Type 1 processing does not (Evans & Stanovich, 2013).

Two experimental manipulations that appear to induce Type 1 processing are time constraints (Roberts & Newton, 2001) and concurrent working memory load (Neys, 2006). Both increased the likelihood that participants provided the intuitive yet incorrect response to the Wason card selection task (Wason, 1968), which has been taken as an indication of Type 1 processing (see the section on problem solving in the next chapter). Neuroimaging work, on the other hand, has revealed distinct patterns of brain activity when participants think about immediate or delayed monetary rewards, the latter of which was presumed to evoke prospective hypothetical thinking (McClure, Laibson, Loewenstein, & Cohen, 2004). Finally, Evans and Stanovich (2013) report that cognitive ability correlates positively and significantly with correct responses on challenging problem-solving tasks, but correlates negatively or not at all with intuitive (yet incorrect) responses to the same problems (West & Stanovich, 2003). Evans and Stanovich (2013) argue that calculating the correct response to these problems requires working memory and attention control, whereas providing the intuitive response does not because intuitive responses are generated automatically by Type 1 processes. Thus, individual differences in cognitive ability should correlate with Type 2 but not with Type 1 processing performance, and indeed, this is the pattern of results that Evans and Stanovich (2013) report.

To sum up, dual process theory contrasts automatic processes (Type 1) with controlled, effortful processes (Type 2). The key distinction between these two types of cognitive operations is the role of attention and working memory (Evans, 2008; Stanovich & Toplak; 2012). Similarly, attention and working memory play a critical role in the distinction between skills that have been practiced to the point of automaticity and those that have not (Anderson, 1982; Fitts & Posner, 1967). Although automatic processing is typically thought of as adaptive within skill-acquisition frameworks, dual process theorists have identified ways in which automaticity can lead to mistakes. In the next chapter, errors of automaticity are discussed within the contexts of driving, medical diagnosis, and problem solving.

## ERRORS OF AUTOMATICITY

"The best swordsman in the world doesn't need to fear the second best swordsman in the world; no, the person for him to be afraid of is some ignorant antagonist who has never had a sword in his hand before; he doesn't do the thing he ought to do, and so the expert isn't prepared for him" (Twain, 1889, p. 344).

In this chapter, I discuss how automaticity can lead to errors in the domains of driving, medical diagnosis, and problem solving. I also review evidence suggesting that errors of automaticity are associated with situational and individual-difference factors. The main argument to be presented is that automatic processes, which are typically thought of as adaptive, can also be a source of error.

### Driving

In some respects, we are all experts at going about our daily lives and solving the problems that arise along the way. For example, if we need to get to campus, we get in our car and go, often taking the same route each time. Driving a personal vehicle along a familiar route is one of the best-practiced skills in the adult population (Charlton & Starkey, 2013), and it is often done "on autopilot" while conversing with a friend, finishing breakfast, or listening to the radio.

Indeed, some aspects of driving seem to be automated such that they no longer require attention. For instance, studies using dual-task paradigms have found that steering, driving speed, and following distance are unaffected by a distractor task in experienced drivers (Brown, Tickner, & Simmonds, 1969; Duncan, Williams, Nimmo-Smith, & Brown, 1992). By contrast, braking times show performance impairments when drivers are engaged in a secondary task (Duncan et al., 1992). Charlton and Starkey (2013) found that repeatedly driving on a familiar route led to inattention blindness for roadside features, with many participants reporting that they were “driving without thinking about it” or “zoning out,” comments indicative of “driving without awareness” (p. 131). Similarly, a frequently reported experience is arriving at a familiar destination with little to no recollection of the trip (e.g., “how did I get here?”; May & Gale, 1998). Because attention is necessary to encode episodic memory (Beilock & Carr, 2001; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996), the absence of memory for a drive can be taken as evidence that attention was not fully engaged throughout the journey. What this suggests is that many of the actions that we use to navigate a familiar route are automated.

Now consider driving on the freeway after a heavy snowfall. In hazardous conditions, a car may lose traction without warning and quickly veer towards the guardrail. When sliding on ice, many drivers reflexively slam on the brakes and wrench the steering wheel away from danger, only to discover that this exacerbates the skid and causes them to lose control of their vehicle. In normal driving conditions, these braking and steering responses might save a driver from fast-approaching danger. In icy conditions, these same reactions can have dire consequences.

In this scenario, the hasty response of applying the brakes and steering away from danger occurs automatically, but it is the wrong course of action. This is a failure of *overgeneralizing* (Taylor, 1975) – an approach that usually works is applied to a novel situation, in this case with disastrous results. Even drivers who have heard the mantra “you can’t steer out of a skid” have probably made this mistake because the braking and steering reactions occur rapidly and there is not much time to consider alternative solutions. In this case, the error might be considered a result of failing to inhibit a habitual response. However, with enough practice driving through the snow, drivers can (and do) learn to automate the correct procedure. In more general terms, then, the problem is that automaticity is brought to bear on a situation for which it is not appropriate, but the performer discovers this only after a mistake has been made.

Automaticity in driving can lead to other errors, too. For instance, if skilled drivers believe that they can operate their vehicle without attention, they may be more likely to engage in distracted driving. This may be why cell phone use while driving correlates with overestimates of driving competence and drivers' illusion of control (Schlehofer et al., 2010). Drivers who text message while driving also report lower levels of feeling distracted while doing so (Struckman-Johnson, Gaster, Struckman-Johnson, Johnson, & May-Shinagle, 2015). Taken together, these findings suggest that people may drive while distracted because they think they can get away with it. Research, however, shows that talking on the phone or text messaging while driving substantially impairs performance, leading to an increased incidence of car crashes (Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009; Strayer & Johnston, 2001). As a case in point, distracted driving was cited as the cause of approximately 1 in 6 fatal vehicle collisions in the United States in 2008 (National Highway Traffic Safety Administration, 2011). Thus, automatic processes while driving could lead to errors because they allow drivers to direct their attention elsewhere when it is not prudent to do so.

**Medical Diagnosis**

One of the most consequential mistakes a physician can make is a misdiagnosis (Croskerry, 2009). Patients given the wrong diagnosis may be given the wrong treatment or fail to receive essential medicine. In the United States, diagnostic errors cost between 40,000 to 80,000 lives each year (Hayward, 2002). Graber, Franklin, and Gordon (2005) report that 74% of diagnostic failures can be attributed to errors in the physician's thinking, as opposed to, for example, lack of knowledge. A portion of these errors is caused by automatic processing of symptomology, such as using heuristics like pattern recognition to make diagnoses (Pham et al., 2012). These errors may be avoidable. However, there are also contextual pressures that increase the likelihood that physicians rely on automatic processes to make diagnoses.

One way physicians make diagnoses is via pattern recognition (Croskerry, 2009). For example, if a child presents a familiar set of symptoms, such as fever, malaise, and red bumps on the skin, a pediatrician may diagnose chickenpox. This diagnosis might be based on prior experience with patients presenting similar symptoms and the pediatrician's medical training. For common ailments, the pattern recognition process may be so well-practiced that the physician seems to know immediately what is wrong with the patient (e.g., "seen this many times before," Croskerry, Singhal, & Mamede, 2012, p. 58). In such cases, the physician may be able to rely on Type 1 processing to make a diagnosis without loss in accuracy. Type 1 processing in this circumstance saves the physician time and effort (Croskerry et al., 2012).

However, a major challenge for medical diagnosis is that there is always some degree of "irreducible uncertainty" (Hammond, 1996). That is, there is no guarantee that a set of symptoms indicates one particular disease and not another, although for chickenpox, the likelihood of accurate diagnosis is probably quite high. Nevertheless, if 0.01% of patients diagnosed with chickenpox actually have infective endocarditis, an infection of the heart valves that also results in red bumps, a misdiagnosis could be fatal (Mayo Clinic Staff, 2018). In this case, the physician has to override Type 1 processing and spend more time with the patient to make an accurate diagnosis, asking follow-up questions and conducting additional tests. This Type 2 approach to diagnosis is more cognitively demanding and resource intensive, but it is less likely to result in error (Pham et al., 2012).

Although studying failures of automatic processing in real-world medical contexts has proven difficult, numerous studies in the laboratory have shown that physicians are susceptible to errors arising from Type 1 processing. For example, Mamede et al. (2010) had residents in internal medicine make diagnoses based on case reports in different decision-making conditions. On trials designed to tap Type 1 processing, the residents were given no time for deliberation; they were asked to write down the first diagnosis that came to mind after reading the case report. On trials designed to tap Type 2 processing, they were asked to think carefully about each report, spending up to 8 minutes following instructions that induced an elaborate analysis of each case. The residents made significantly more misdiagnoses in the immediate-decision condition than in the careful-deliberation condition, suggesting that relying on Type 1 processing led to an increase in errors.

These results indicate that physicians should spend more time during diagnosis to allow Type 2 analytical reasoning to unfold and to mitigate Type 1 errors. In many medical settings, however, additional time and cognitive resources may not always be available. Medical emergencies demand rapid decision-making, and physicians may need to see many patients in a short time (Klein, 2005). For this reason, emergency medicine has been referred to as a “natural laboratory of error” (Bogner, 1994). Monitoring Type 1 processes becomes difficult under conditions of *cognitive overload*, a term used to describe the competing attentional demands placed on anyone needing to make complex decisions while multitasking, particularly under time constraints (Gilbert, Pelham, & Krull, 2003; Kirsch, 2000). Thus, physicians may struggle to override fast intuitive responses when making decisions about patients’ health during peak hours.

Physicians are also notoriously overworked and often sleep-deprived. As a case in point, residents regularly work on-call shifts for over 24 hours at a time (Landrigan et al., 2004).

Croskerry (2009) suggests that sleep deprivation might compromise the effectiveness of deliberative Type 2 processes, which may cause physicians to rely on automatic Type 1 processes more often. Indeed, sleep deprivation has been shown to impair working memory capacity and sustained attention (see Frenda & Fenn, 2016, for a review), both of which support Type 2 processing (Evans, 2008).

Support for Croskerry's (2009) point is also provided by a landmark study by Landrigan et al. (2004), showing that reducing residents' work hours substantially reduced the incidence of medical errors. Residents who regularly worked traditional 30-hour shifts made 35.9% more serious medical errors than residents who were assigned to a staggered schedule with no shifts exceeding 16 hours. The errors ranged from misdiagnoses to ordering drug overdoses. This evidence is consistent with the idea that sleep deprivation increases the likelihood of expert error, perhaps because physicians are more likely to rely on Type 1 processes when Type 2 processes are compromised. However, an alternative explanation is that Type 2 processes are more likely to fail when decision makers are sleep-deprived (Harrison & Horne, 2000), which could also result in error. The combination of both factors could be especially damaging, but more work is needed to disentangle these possibilities.

## Problem Solving

Although Type 1 processes provide quick and intuitive responses, these automated responses are not always optimal (Toplak, West, & Stanovich, 2014). Thus, one of the primary roles of Type 2 processing is to override Type 1 processing (Toplak et al., 2014). However, as demonstrated by studies of judgment and decision making, people are often *cognitive misers* when problem solving (Fiske & Taylor, 1984). That is, they rely on Type 1 processing to provide quick, intuitive responses rather than deliberating carefully about the problem to be solved.

Cognitive miserliness is a dispositional trait reflecting a preference for Type 1 processing irrespective of one's ability or potential to engage in careful, reflective thought.

The typical explanation for why people engage in miserly cognitive processing, or Type 1 processing, is that Type 2 processing is cognitively demanding and perceived as aversive (Toplak et al., 2014). If most of the everyday problems we encounter can be solved quickly and effortlessly with Type 1 processing, then this approach becomes an attractive default mode even if it results in occasional errors. A potential benefit of relying on Type 1 processing in everyday circumstances is that time and effort are spared for situations that clearly demand careful, reflective thinking.

Perhaps the most compelling evidence that people rely on Type 1 processing during problem solving is provided by Frederick's (2005) 3-item cognitive reflection test. The questions are designed to prime an intuitive yet incorrect Type 1 response but are simple enough that anyone with basic mathematics ability can calculate the correct answer through Type 2 processing. For example, consider this question: "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?" The intuitive response is 10 cents. People often report that this answer immediately springs to mind as if it were automatically primed by the question (Frederick, 2005). The rationale is that participants substitute the question being asked for a simpler one, namely, "what is the difference between the two values?" The correct answer, however, is not 10 cents, but 5 cents ($0.05 + $1.05 = $1.10).

Most people do not perform well on the cognitive reflection test. At Michigan State University, students answered, on average, only 0.79 problems correctly out of 3 (Frederick, 2005). The argument that poor performance indicates an overreliance on Type 1 processing is supported by three converging lines of evidence. First, when people answer incorrectly, they almost always provide the intuitive responses primed by the questions (Frederick, 2005). That is, errors do not occur at random. This rules out the possibility that people make haphazard guesses when confronted with these problems; there is a systematic operating principle underlying their incorrect responses. Second, when people answer correctly, they often report thinking of the intuitive responses first before calculating the correct solutions (Frederick, 2005). This is consistent with the idea that Type 1 processes operate quickly, whereas Type 2 processes operate in a slower, more controlled manner. Third, when participants are asked to rate the difficulty of the items, people who answer them incorrectly rate them as easier than people who answer them correctly (Frederick, 2005). This finding supports the argument that to perform well on the cognitive reflection test, one must override Type 1 responses primed by the questions and instead engage in more effortful thinking. Those who fail to override Type 1 processing find the problems easier, despite getting the answers wrong, because they do not engage in effortful problem solving to calculate the correct solution. Taken together, this work suggests that miserly cognitive processing may be the result of automatic processes.

Thus, although often adaptive, automaticity can lead to errors in many aspects of daily life. In the next chapter, I review the role of automaticity in reading, another well-practiced skill in the adult population. Although the cognitive processes that support reading may go unnoticed most of the time, I provide examples showing that what we read is not always what is presented to us. That is, knowledge and expectations are brought to bear on the reading process, which can result in mistakes, particularly when expectations are defied.

# COGNITIVE PROCESSES IN PROOFREADING

Clear and accurate writing is essential in the professional world. Whether it is an attorney's contract, a journalist's report, or an academic's article, many professions depend on the ability to communicate effectively through writing. Proofreading is an important step in the process. A misspelled word or misplaced comma can betray a lack of scrupulousness and even change the meaning of a sentence. Despite our best efforts, the occasional typographical error may go undetected in our writing, finding its way into a colleague's inbox or a reviewer's workflow. These mistakes can be costly.

In this chapter, I synthesize the literature on proofreading through the lens of cognitive science. In particular, I examine proofreading performance as it relates to familiarity with the material being read, top-down and bottom-up processing, and eye movements during reading. I present evidence suggesting that proofreading failures when reading our own work are caused by the automatic influence of memory of the text being proofread. That is, overfamiliarity with the text leads to an expectancy-driven style of processing that reduces the type of textual analysis needed for successful proofreading.

## Proofreading and Familiarity

In one of the first studies of proofreading, Levy and colleagues tested whether familiarity with a passage of text affected the ability to detect spelling errors (Levy, 1983; Levy & Begin, 1984). Levy (1983) had participants read short passages and mark any errors that they noticed. There were two experimental conditions. In the *unfamiliar* condition, participants simply proofread error-filled versions of the passages. In the *familiar* condition, they first read error-free versions multiple times and then proofread error-filled versions of the same passages. Levy

(1983) found that participants in the familiar condition detected approximately 10% more typographical errors than participants in the unfamiliar condition. They also were faster.

One explanation for why familiarity facilitated proofreading is based on the observation that proofreading is a complex cognitive activity requiring multiple levels of processing, including word recognition, text integration, and comprehension (Levy, 1983). If participants in the familiar condition understood the text prior to proofreading it, they could allocate greater attention to word recognition and text integration while proofreading. Greater attention to the visual aspects of the text might have facilitated error detection, especially because the errors were all *non-words*: for example, the word *about* was changed to *ahout.* These types of errors would be particularly salient if one were attending primarily to lower-level word recognition processes. By contrast, participants in the unfamiliar condition would have to simultaneously engage in both lower-level word recognition and higher-level comprehension during the proofreading task, possibly taking attentional resources away from a more thorough visual analysis. This could explain why participants in the unfamiliar condition performed worse than participants in the familiar condition.

However, there are concerns with Levy's (1983) findings. First, modern word processors have "spell check" to flag non-words. If familiarity only facilitates the detection of non-words during proofreading, then this benefit is largely inconsequential given present-day technology (handwritten messages would be an exception). Second, because participants in the familiar condition proofread not only more accurately but also more rapidly, they may not have been reading the text at all. Instead, they may have been visually scanning it for non-word letter combinations and eschewing higher order comprehension altogether.

To test this hypothesis, Levy, Newell, Snyder, and Timmins (1986) investigated whether familiarity also facilitated the detection of word errors. A *word error* occurs when a word is changed to a different, contextually-inappropriate word. For example, Levy et al. (1986) exchanged the word *sale* for *sail*, a homophone. Other errors were prefix substitutions (e.g., *loved* was swapped for *unloved*) and single-letter substitutions that altered a word's pronunciation (e.g., *major* was swapped for *mayor*). Levy et al. (1986) manipulated familiarity in the same way as Levy (1983): in one condition, participants proofread passages; in the other condition they familiarized themselves with error-free versions of the passages before proofreading. If participants in the familiar condition were scanning the text for salient non-words without analyzing the passages' meaning, word errors might go undetected.

Instead, Levy et al. (1986) found that familiarity with the text facilitated the detection of both non-word and word errors. The advantage of familiarity extended to all types of word errors, including homophones, prefix substitutions, and single-letter alterations affecting pronunciation. This suggests that participants in the familiar condition were processing higher-level syntactic-semantic aspects of the text while proofreading, ruling out a "visual scanning" account of the results. Participants in the familiar condition processed the meaning of the text while proofreading and may have detected more word errors because they could compare their understanding of the text as they read it with their prior understanding of it.

A straightforward implication of these results is that error detection should be facilitated when we proofread our own work. Typically, we are highly familiar with our own writing, especially after having spent weeks or months working on a project. If Levy et al.'s (1986) hypothesis is correct, then having prior understanding of our work should support the detection of word errors because we can compare what we wrote to what we meant to write. Furthermore, if familiarity facilitates reading comprehension, then attentional resources may be spared for more thorough visual processing of the text. This, in turn, might lead to greater detection of non-word errors. Based on Levy and colleagues' hypothesis, it follows that we should be excellent proofreaders of our own writing.

**The Self-Generation Effect**

However, there is reason to think that we may be *worse* at proofreading our own writing than the writing of others. Daneman and Stainton (1993) had participants spend 30 minutes writing a short essay on student life. They then assigned them to one of three conditions: in the *self-generated* condition, they proofread the essay they had just written; in the *unfamiliar other-generated* condition, they proofread a different student's essay; and in the *familiar other-generated* condition, they familiarized themselves with a different student's essay by reading it three times prior to proofreading an error-filled version. All participants were told that errors had been added to the essays before they began proofreading.

Daneman and Stainton (1993) found that students who proofread their own essays detected 20% fewer errors than students who proofread a familiar essay written by someone else. Furthermore, replicating the results of Levy et al. (1986), students who proofread a familiar essay written by someone else detected 7% more errors than students who proofread an unfamiliar essay. This suggests that familiarity enhances proofreading only up to a point — once an essay is *overfamiliar*, performance decreases substantially, as shown in Figure 1.

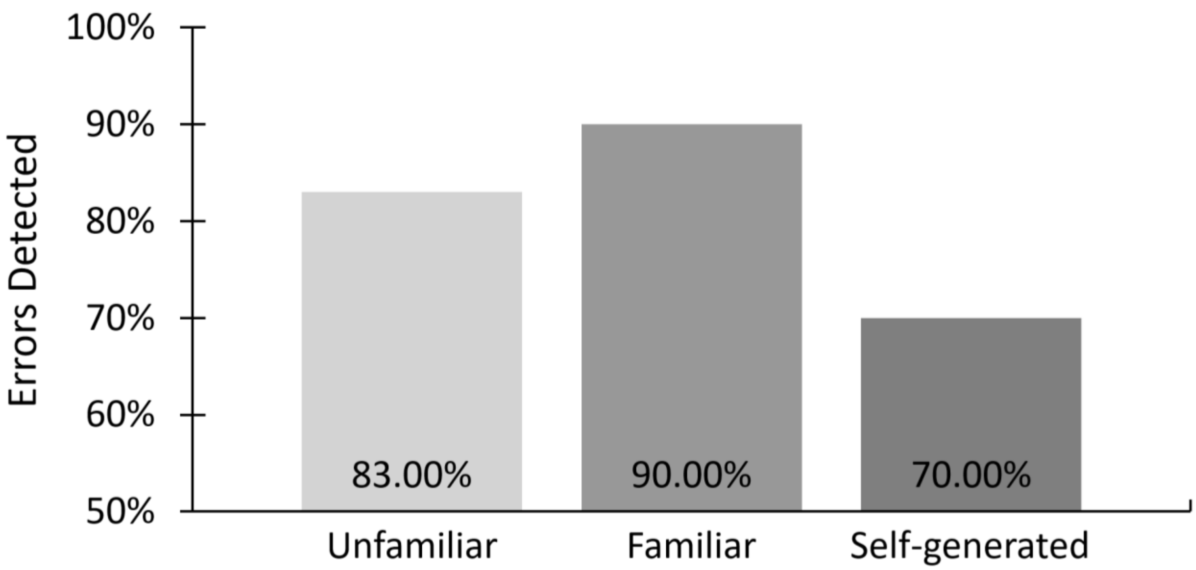


**Figure 1.** Relationship between proofreading performance and familiarity with the essay being proofread. Means are presented at the base of each bar. Data from Experiment 1 of Daneman and Stainton (1993).

Supporting this interpretation of the results, in a second experiment Daneman and Stainton (1993) had participants proofread essays two weeks after writing them. They hypothesized that if overfamiliarity led to poor proofreading performance, then after two weeks participants' memory for their own essays would decay and overfamiliarity would no longer be a problem. Indeed, two weeks later, participants were able to proofread their own essays just as accurately as a familiar essay written by someone else.

Daneman and Stainton's (1993) results suggest that *overfamiliarity*, from having just written an essay and being "intimately acquainted with [its] semantic and syntactic features," hinders proofreading performance (p. 306). Specifically, they argued that overfamiliarity automatically influences reading behaviors, causing participants to engage in a top-down, "expectancy-driven" (p. 299) style of processing that reduces attention to the visual and semantic-syntactic aspects of the text. This worsens proofreading performance, they argued, because a thorough visual and semantic-syntactic analysis is necessary to detect errors.

Today, nearly three decades later, Daneman and Stainton's (1993) hypothesis about the mechanism underlying the self-generation effect has still not been rigorously tested. However, eye tracking can shed light on the focus of attention during reading and, in turn, can illuminate cognitive processes supporting proofreading performance. In particular, eye tracking could be used to measure indicators of top-down or "expectancy-driven" processing during proofreading. This would allow us to test Daneman and Stainton's (1993) hypothesis that a reliance on top-down processing drives the relationship between overfamiliarity and poor proofreading performance.

**Top-Down and Bottom-Up Processing**

Before reviewing the literature on eye movements during reading, it is worth noting the distinction between *top-down* and *bottom-up* processes and how it might relate to reading comprehension. Typically, top-down processing refers to user-driven, or internally guided, allocation of attention, whereas bottom-up processing refers to stimulus-driven, or externally guided, allocation of attention (Wolfe, 2014).

Kintsch (2005) argued that both top-down and bottom-up processes interact to produce text comprehension. In particular, Kintsch suggested that bottom-up processing during reading — a data-driven process based on analysis of the perceptual details of the text — results in knowledge activation. In turn, activated knowledge serves as a top-down constraint or backdrop against which further information is processed.

In one example of this interaction between bottom-up and top-down processing, Kintsch (2005) tells a story about a trip to the grocery store, but the story is filled with ambiguities because the context is never stated explicitly. However, once the reader has established a "grocery shopping" schema, either through induction or instruction, ambiguous sentences are readily understood within this top-down constraint. For instance, given the sentence "She was surprised by how much she had to pay" (p. 127), the reader can infer that a woman was at a checkout counter and paying for her groceries. None of this information is provided in the vignette. This example of an interaction between top-down and bottom-up processing reveals that textual information is not processed in a vacuum but, rather, is readily influenced by our expectations about it. As Kintsch (2005) stated, "What we see is in part determined by what we expect to see" (p. 127). A demonstration of this effect is found in some of the popular illusions shown in Figure 2, in which readers tend to see what they expect instead of what is presented.

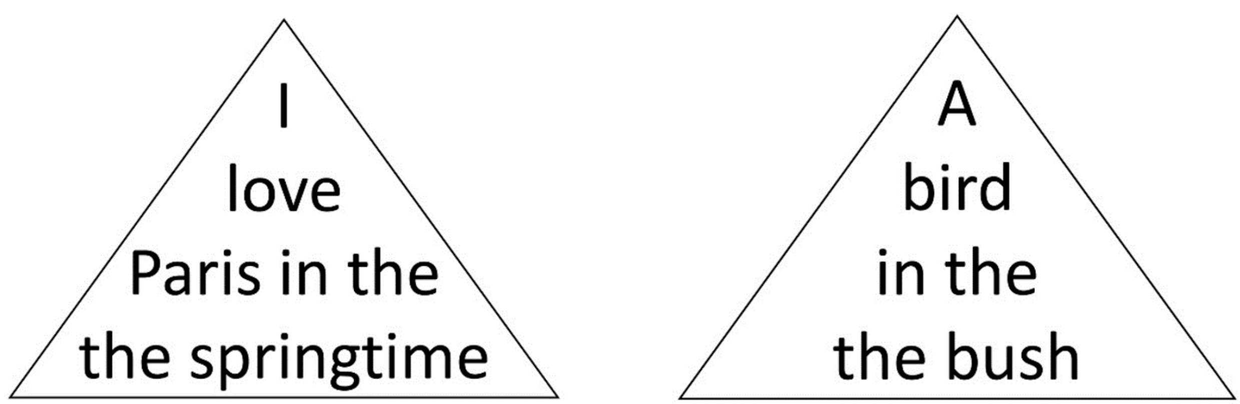


**Figure 2.** Readers often see the statements "I love Paris in the springtime" and "A bird in the bush" in the examples above, failing to detect the repetition of the word "the" in each case. One explanation for this phenomenon is an overreliance on top-down processing.

In the context of proofreading, Daneman and Stainton (1993) characterize top-down processing as "expectancy-driven processing" (p. 299), by which they mean the guidance of attention by a mental representation of the material being read (i.e., a *knowledge-driven* process;

Kintsch, 1988). Mental representations can take many forms, such as general schemas — as was the case in the “grocery shopping” example — or a more specific memory of a passage of text, perhaps because one has seen the same passage before.

A familiar idiom can serve as the basis for expectancy-driven processing. For example, a key word can be missing from the following phrase, but most readers will not have trouble understanding its meaning: “*Fool me once, shame on you. Fool me twice, shame on* ____.” In this example, expectancy-driven processing readily fills in the blank with the word “*me*.” As another example, our use of expectancy-driven processing becomes particularly salient when expectations are defied in noticeable ways, as in the now-classic quote from George W. Bush, “Fool me once, shame on you. Fool me—you can’t get fooled again” (“Top 10 Bushisms,” 2009).

Daneman and Stainton (1993) argued that an overreliance on expectancy-driven processing during proofreading resulted in a failure to detect word and non-word errors. Their hypothesis was that participants did not thoroughly process most of the words in their essays and instead used the occasional word as a retrieval cue, filling in the blanks with a mental representation of the essay they wrote.

An important question, however, is why participants who were asked to carefully proofread an essay relied on a style of processing ill-suited to spot mistakes. Pilotti and Chodorow (2012) suggest that a shift to relying on top-down processing might be an automatic and unintentional consequence of overfamiliarity. Specifically, they cite evidence from priming experiments to argue that prior exposure to a stimulus affects the likelihood that memory influences processing of that stimulus (see, e.g., Crabb & Dark, 1999, 2003).

Indeed, there are instances when experts appear to reflexively engage in top-down processing of material within their domain of skill, sometimes at the expense of task performance. For instance, chess experts are reliably faster than novices at detecting whether a King is in check (i.e., under attack) (Reingold, Charness, Pomplun, & Stampe, 2001). However, a cost of expertise, similar to a Stroop-like interference effect, can be observed when players are told to indicate if the King is in check by a cued piece. On trials in which only a non-cued piece attacks the King, chess experts are slower to respond than when no cue is used at all (Reingold, Charness, Schultetus, & Stampe, 2001). This suggests that chess experts automatically process the whole position despite instructions to focus on only a particular piece, and their performance is negatively affected as a result. Thus, it seems plausible that students who were overly familiar with their own writing might have had similar trouble disengaging from an automatic, top-down processing style during the proofreading task.

**Eye Movements**

One way to ascertain whether people rely on top-down processing when proofreading their own work is to examine their eye movements. The use of eye tracking to study reading behaviors has a rich history, spanning well over a century (Javal, 1878; Tinker, 1927). This is because the movement of the eyes during reading can be used to gain insight into the cognitive processes supporting comprehension. As Just and Carpenter (1980) state:

> Unlike a listener, a reader can skip over portions of the text, reread sections, or pause on a particular word. A reader can take in information at a pace that matches the internal comprehension processes. By examining where a reader pauses, it is possible to learn about the comprehension processes themselves. (p. 329)

Similarly, Rayner (1998) states that “eye movements are intimately related to the moment-to-moment cognitive processing activities of readers” (p. 390). This suggests that eye tracking could be used to better understand the cognitive processes involved in proofreading.

What do typical reading behaviors look like through the lens of the eye tracker? Readers make rapid eye movements, called *saccades*, which are interleaved by periods of *fixation*, during which the eyes are relatively still (Rayner, 1978). The average fixation during silent reading lasts around 225 ms (Rayner, 1998). The average saccade spans approximately 8 letters and is somewhat invariant to the size of the typeface (Rayner, 1998). Readers make saccades in order to align the *fovea*, the center of the field of view, with the portion of the text that needs to be perceived accurately. This is because the fovea has better visual acuity than the parafoveal and peripheral visual fields (Latham & Whitaker, 1996; Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981).

English is read from left to right, but readers occasionally make *regressions*, moving their eyes from right to left or back to a previous line. Approximately 10-15% of saccades are regressions, which tend to occur if the reader has made too far of a forward saccade or is having difficulty understanding the material just read (Frazier & Rayner, 1982; Rayner, 1998). Readers can also fixate the same word multiple times; these fixations are often grouped together prior to data analysis so that the total time spent fixating the word, the *dwell time,* can be used as an index of processing time (Gidlöf, Wallin, Dewhurst, & Holmqvist, 2013).

Not all words are fixated during reading. Just and Carpenter (1980) found that people averaged 1.2 words per fixation when reading unfamiliar text. Providing a slightly different estimate, Brysbaert and Vitu (1998) report that proficient English readers can skip up to one third of the words in a passage. Furthermore, some words are more likely to be fixated than others.

Just and Carpenter (1980) state that readers are more likely to fixate longer words than shorter words. Short function words such as "the" are often skipped by skilled readers (recall "A bird in the the bush" from Figure 2) (O'Regan, 1975, as cited in Rayner, 1998).

For example, Healy (1976) asked participants to read short passages of text and circle instances of the letter "t." Participants were more likely to overlook the letter "t" when it was embedded in the word "the" than when it was embedded in the word "thy." Healy (1976) attributed this effect to the high frequency of the word "the" (see also Minkoff & Raney, 2000). Healy's (1976) work reinforces Kintsch's (1988) point that reading is not driven exclusively by bottom-up processing. If it were, participants would have no more trouble spotting the letter "t" in the word "the" than in the word "thy." Rather, skilled readers are akin to experts, who at times have difficulty inhibiting automatic processing of material within their domain of skill.

One debate in the literature concerns the frequency with which content and function words are fixated during reading. *Function words*, such as "and," "or," "the," and "a," primarily serve to indicate grammatical relations between the other words in a sentence (Schmauder, Morris, & Poynor, 2000). As such, function words include articles, pronouns, particles, and conjunctions. By contrast, *content words*, such as "school," "notes," "test," and "essay," are laden with compositional meaning (Schmauder et al., 2000). Content words include nouns, verbs, and adjectives.

Carpenter and Just (1983) found that readers fixated 83% of content words but only 38% of function words in an essay reading task. This would seem to suggest that content words are fixated more often than function words. However, frequency and word length were confounding variables in Carpenter and Just's (1983) experiment. That is, the function words were shorter in length and higher in frequency than the content words, and both factors have been found to influence the likelihood of fixation (Rayner, 1998). To disentangle these confounds, Schmauder et al. (2000) matched function and content words for length and frequency and embedded them in short passages of text. They found that participants were just as likely to skip function words as content words: approximately 17% of each were skipped.

**Indicators of Top-Down and Bottom-Up Processing**

Given the tight coupling of the focus of attention and the movement of the eyes during reading, indicators of bottom-up or top-down processing should be detectable in the eye movements of readers. To reiterate, top-down processing is knowledge-driven. By contrast, bottom-up processing is stimulus-driven; during reading, it is based on the perceptual details of the text (Kintsch, 2005). Thus, if a reader engaged primarily in bottom-up processing, one might expect to observe fixations on many of the words in a passage. Some support for this prediction is provided by Just and Carpenter's (1980) observation that participants average 1.2 words per fixation when reading unfamiliar text. Although participants were not prevented from engaging in top-down processing, the use of an unfamiliar passage meant that they had less knowledge to bring to bear on the reading process than if a familiar passage had been used.

If readers fixate most words when engaging in bottom-up processing, then the average length of their saccades must be short. This is because fixating every word in a sequence necessitates short saccades between words. Readers engaging in more bottom-up processing might also have longer dwell times because a thorough visual analysis of the text requires more processing time than a less thorough analysis (Rayner, 1998). In sum, readers engaging in more bottom-up processing might make more fixations, shorter saccades, and have longer dwell times than readers engaging in less bottom-up processing.

Conversely, a top-down or “expectancy-driven” reading style might result in a different pattern of eye movements. Readers who have a robust mental representation of a passage may not fixate every word during reading (Pilotti & Chodorow, 2009). Instead, they may make few fixations with large saccades in between. Furthermore, if readers do not engage in a thorough perceptual analysis of the text, they may not spend as much time dwelling per fixation (Rayner, 1998). Thus, readers engaging in more top-down processing may make fewer fixations, longer saccades, and have shorter dwell times than readers engaging in less top-down processing.

One way to test the influence of top-down or knowledge-driven processes on reading is to observe how eye movements change after reading the same passage multiple times. On a second reading, knowledge of the text can be brought to bear on the reading process. Therefore, the second reading is more conducive to top-down processing than the first. Hyönä and Niemi (1990) asked participants to read the same essay twice and instructed them to read for comprehension because there would be a memory test later. Relative to the first reading, on the second reading participants made fewer fixations per sentence (9.3 vs. 9.8), fewer regressions per sentence (1.0 vs. 1.2), and spent less time dwelling per fixation (243 ms vs. 254 ms). This result indicates that top-down influences affect the movement of the eyes during reading and that this effect can be detected using an eye tracker.

In a related study, Raney, Therriault, and Minkoff (2000) found that the effect of knowledge on eye movements during reading extended to paraphrased text. First, they had participants read an essay once for comprehension. Next, they had them read a paraphrased version of the same essay. If participants had developed an understanding of the first essay, then this knowledge might influence how they read the paraphrased version. When they compared participants’ eye movements while they read the first essay and the paraphrased version, they found that relative to the first essay, participants spent less time reading paraphrased passages (28.4 s vs. 33.5 s), made fewer forward fixations (74.9 vs. 87.6), and spent less time dwelling per forward fixation (245 ms vs. 255 ms). Thus, it appears that there are reliable effects of knowledge on reading behaviors, as shown in Table 1.

**Table 1.** Eye movements during reading indicative of top-down or bottom-up processing

| Processing Style | Number of Fixations | Saccade Distance | Dwell Time |
|---|---|---|---|
| Top-Down | Less | Longer | Shorter |
| Bottom-Up | More | Shorter | Longer |

# AUTOMATIC PROCESSING DURING PROOFREADING

In this chapter, I present two experiments designed to test Daneman and Stainton's (1993) hypothesis that robust memory for self-generated text leads to an overreliance on top-down processing, which in turn leads to reduced error detection. Although the goal of this work is to better understand the self-generation effect in proofreading, the broader purpose is to reveal a hidden cost of expertise: the sometimes detrimental effect of automatic processing of familiar material.

## Experiment 1

Experiment 1 was conducted in two sessions and was designed as a near replication of Daneman and Stainton (1993), incorporating the use of a computerized proofreading task and eye tracking during proofreading. The predictions are based on Daneman and Stainton's (1993) results and hypotheses. The first prediction is that during Session 1, students who proofread their own essay shortly after writing it (the *self-generated* condition) will detect fewer errors than students who proofread someone else's essay (the *other-generated* condition). The second prediction is that students in the self-generated condition will make fewer fixations, longer saccades, have shorter dwell times, and spend less time proofreading than students in the other-generated condition. The third prediction is that during Session 2, which occurs one week after Session 1, students who proofread the essay they wrote one week prior will detect more errors than students who proofread an unfamiliar essay written by someone else.

### Participants

The participants were 64 undergraduate students (53 females, $M_{\text{age}}$ = 19.28, $SD$ = 1.29) recruited from introductory psychology courses at Michigan State University. They earned course credit for participating. All reported normal or corrected-to-normal vision and stated that

English was their first language. All participants provided informed consent. Of the 64 participants who attended Session 1, 50 returned to the laboratory one week later for Session 2. One participant failed to follow instructions and was excluded from proofreading analyses for both sessions. Another participant was dismissed from Session 2 because the eye tracker calibration failed. This left a useable sample of 63 for Session 1 and 48 for Session 2 for the proofreading analyses.

## Procedure

Participants were tested individually in the laboratory. Each session lasted approximately 1.5 hours. During Session 1, participants typed a short essay on student life, describing their coursework, the food on campus, and recreational activities. Next, they completed a reading comprehension test. While they were completing the reading comprehension test, any errors in their essays were corrected by the experimenter and 20 errors were added to the essays. Next, participants proofread either their own essay or another participant's essay. Finally, they completed a test of fluid intelligence (Raven's Advanced Progressive Matrices) and a test of perceptual speed (Letter/Number Comparison).

Session 2 occurred exactly one week later. First, participants completed another reading comprehension test. Next, they proofread either the essay they had written one week prior or another participant's essay. Finally, participants completed a test of working memory (Symmetry Span), a test of perceptual speed (Pattern Comparison), another test of working memory (Visual Arrays), and a test of fluid intelligence (Letter Sets).

## Materials

**Essay Writing Task.** Participants were asked to type an essay on student life using a computer with "spell check" disabled. They were given the following instructions, adapted from

Daneman and Stainton (1993): "*In this task we would like you to write a short essay about college life. Try to write as quickly as possible because you will only be given 20 minutes. We would like you to write about three topics related to student life: your classes and coursework, food, and things students do for fun. Don't be concerned about the literary quality of your work or your typing accuracy. Your essay may be used to determine what events are most typical in a student's life.*" Participants were asked to continue writing until a large textbox presented on the screen was filled with text; essays were approximately 500 words long.

**Adding Errors to Essays.** Prior to adding errors to participants' essays, an experimenter read each essay and corrected typographical or grammatical mistakes. Next, a computer program[1] added 20 errors to each essay: six function word errors (e.g., *or* changed to *of*), six function non-word errors (e.g., *are* changed to *ane*), four content word errors (e.g., *life* changed to *like*), and four content non-word errors (e.g., *notes* changed to *nofes*).[2] Thus, 10 word errors and 10 non-word errors were added to the essays; 12 were function errors and 8 were content errors. The errors added to the essays were randomly drawn from a list of target words. The list was developed during pilot testing and consisted of words that frequently occurred in participants' essays and the 24 error words provided in an example essay by Daneman and Stainton (1993, p. 303). The use of the list ensured that similar types of errors were embedded in different participants' essays.

Prior to saving the essays with errors added to them, an experimenter read each essay to ensure that the word errors were indeed erroneous in context. That is, they were checked to make

[1] The error-adding program was developed by a friend and colleague, Sari Sadiya. I am grateful to Sari for his help with these experiments.
[2] During pilot testing, this appeared to be the maximum number of errors that could be added to participants' essays while consistently preserving the intended meaning of the sentences in which they were embedded. For comparison, Daneman and Stainton (1993) added 24 errors to their essays: 12 were word errors and 12 were non-word errors. Ten of the errors altered content words and 14 altered function words.

sure that they were nonsensical in the sentences in which they were embedded. Occasionally, a sentence made sense with a word error embedded. When this occurred, the error-adding computer program was run again, and the resulting essay was checked again.

**Reading Comprehension Tests.** Participants completed computerized versions of the Nelson Denny Reading Test (Brown, Fishco, & Hanna, 1993), the same test used by Daneman and Stainton (1993). Participants completed different versions of the test during Sessions 1 and 2. Each version included seven reading passages and 38 questions with five answer choices. The time limit was 20 minutes, and the measure was the number of correctly answered questions. An overall reading comprehension score was computed by averaging performance on the two tests.

**Proofreading Task.** Participants performed the proofreading task using a specially-designed computer program.[3] Participants read the following instructions: "*In this task you will read an essay. A number of errors have been inserted into the essay. Your task is to read the essay and highlight the errors. You can highlight errors in the essay by clicking and dragging with the left mouse button. To unhighlight something, click and drag with the right mouse button. You should try to proofread as quickly and accurately as possible.*" Then, an essay with errors embedded in it was presented, and they proofread the essay. Participants were instructed to press the "escape" key on the keyboard when they reached the end of the essay. That is, participants were instructed not to re-read the essay.

Eye movements during proofreading were recorded using an EyeLink 1000. Left eye gaze was tracked at a temporal resolution of 1000 Hz. Stimuli were presented on a monitor positioned 692 mm from the participant, with screen dimensions of 433 × 271 mm and a resolution of 1680 × 1050 px. Drift correction was administered prior to displaying the essay to

[3] I am grateful to David MacFarlane for his help developing the proofreading program.

be proofread. The EyeLink 1000 on-line parser was used to detect saccade and blink events. Additional filtering removed saccades larger than half the display size to omit eye movements between new lines.

Participants were randomly assigned to one of two conditions, as shown in Table 2. During Session 1, participants in the *self-generated* condition proofread the essay that they had written approximately 20 minutes prior, whereas participants in the *other-generated* condition proofread an essay written by another participant. Participants were tested individually, but their condition assignments were "yoked" in pairs, such that the first participant in each pair was assigned to the self-generated condition and the second participant was assigned to the other-generated condition and proofread the essay written by the first participant in their pair. During Session 2, participants in the self-generated condition proofread the essay written by the other participant in their pair; this will be referred to as the "other-generated" group in Session 2. Conversely, participants in the other-generated condition proofread the essay they had written one week prior; this will be referred to as the "self-generated one week ago" group in Session 2.

**Table 2.** Condition assignments for the proofreading task in Experiment 1

| Condition | Session 1 | Session 2 |
|---|---|---|
| **Self-generated** | Proofread own essay ($N$ = 33) | Proofread another participant's essay ($N$ = 25) |
| **Other-generated** | Proofread another participant's essay ($N$ = 30) | Proofread own essay written one week ago ($N$ = 23) |

**Raven's Advanced Progressive Matrices.** In this test of fluid intelligence, participants were shown a set of patterns with the lower-right part missing. Participants selected from a list of options the pattern that logically completed the set. They were given 10 minutes to complete the

18 odd-numbered items from the test (Raven & Court, 1998). The measure was the number of correct responses.

**Letter/Number Comparison.** In this test of perceptual speed, participants determined whether two strings of letters or numbers were the same or different. They were given 30 s for each set of 72 items (Salthouse & Babcock, 1991); there were two sets of letter items and two sets of number items. The measure was the number of correct responses.

**Symmetry Span.** In this test of working memory, participants made judgments about the symmetry of images while memorizing the position of squares appearing after each judgment (Oswald, McAbee, Redick, & Hambrick, 2015). The measure was the number of correctly recalled square positions.

**Pattern Comparison.** In this test of perceptual speed, participants determined whether two symbols were the same or different. They were given 30 s for each set of 30 items; there were two sets of items (Salthouse & Babcock, 1991). The measure was the number correct minus two times the number incorrect.

**Visual Arrays.** In this test of working memory, participants were sequentially shown a memory array of two to eight colored squares, a blank display, and a test array, which was either identical to the memory array or different. They determined whether the arrays were the same or different. There were 80 trials (Burgoyne, Hambrick, & Altmann, 2019). The measure was the number of correct responses.

**Letter Sets.** In this test of fluid intelligence, participants were shown five sets of four letters (e.g., DEFG) and chose the set that did not follow the same pattern as the other four. Participants were given five minutes to complete 20 items (Harmon, Ekstrom, French, & Dermen, 1976). The measure was the number of correct responses.

## Results of Experiment 1

### Session 1 Results

Examination of participants' proofreading performance suggests that Daneman and Stainton's (1993) results were not replicated (see Figure 3). That is, whereas the self-generated group was expected to detect *fewer* errors than the other-generated group during Session 1, the opposite pattern of results emerged: the self-generated group ($M$ = 88.64%) detected 5.31% *more* errors than the other-generated group ($M$ = 83.33%), but this difference was not statistically significant, $t(61) = 1.92$, $d = .49$, $p = .059$.

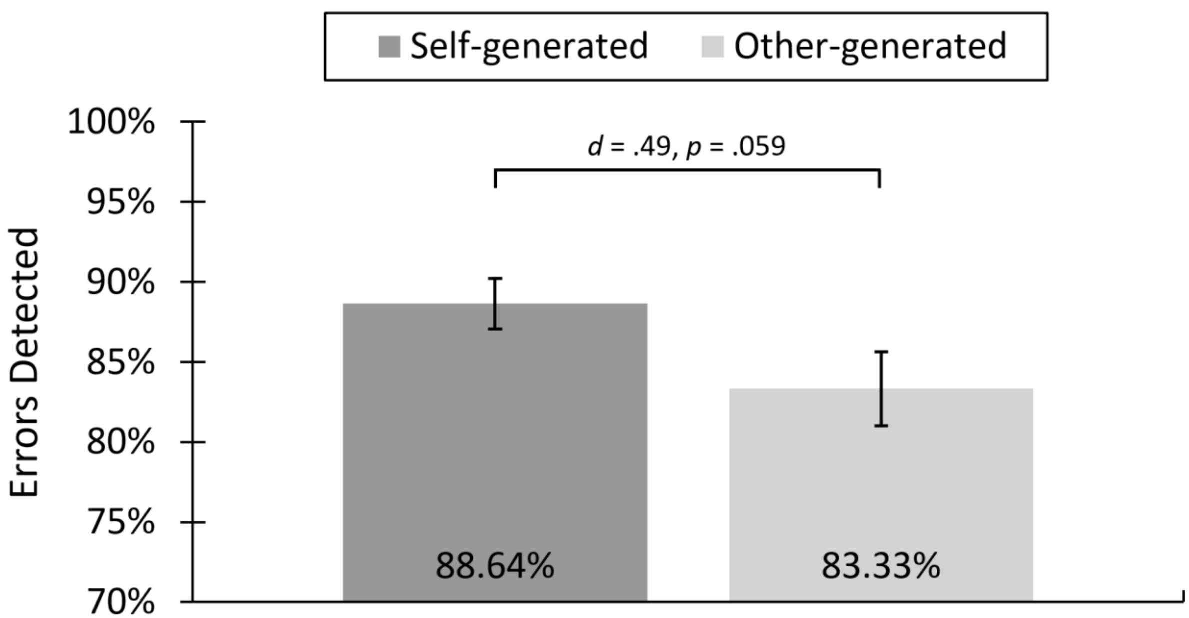


**Figure 3.** Proofreading performance in Session 1 of Experiment 1. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Word and Non-Word Errors.** A paired-samples *t*-test revealed that participants did not detect significantly more non-word errors (*M* = 87.30%, *SD* = 12.85%) than word errors (*M* = 84.92%, *SD* = 15.95%), *t*(62) = 1.02, *d* = .16, *p* = .310.

Word error and non-word error detection rates are plotted in Figure 4. The self-generated group detected significantly more word errors (*M* = 89.39%) than the other-generated group (*M* = 80.00%), *t*(61) = 2.43, *d* = .61, *p* = .018. That is, familiarity from having generated the text oneself facilitated the detection of word errors. By comparison, the self-generated group did not detect significantly more non-word errors (*M* = 87.88%) than the other-generated group (*M* = 86.67%), *t*(61) = 0.37, *d* = .09, *p* = .712.

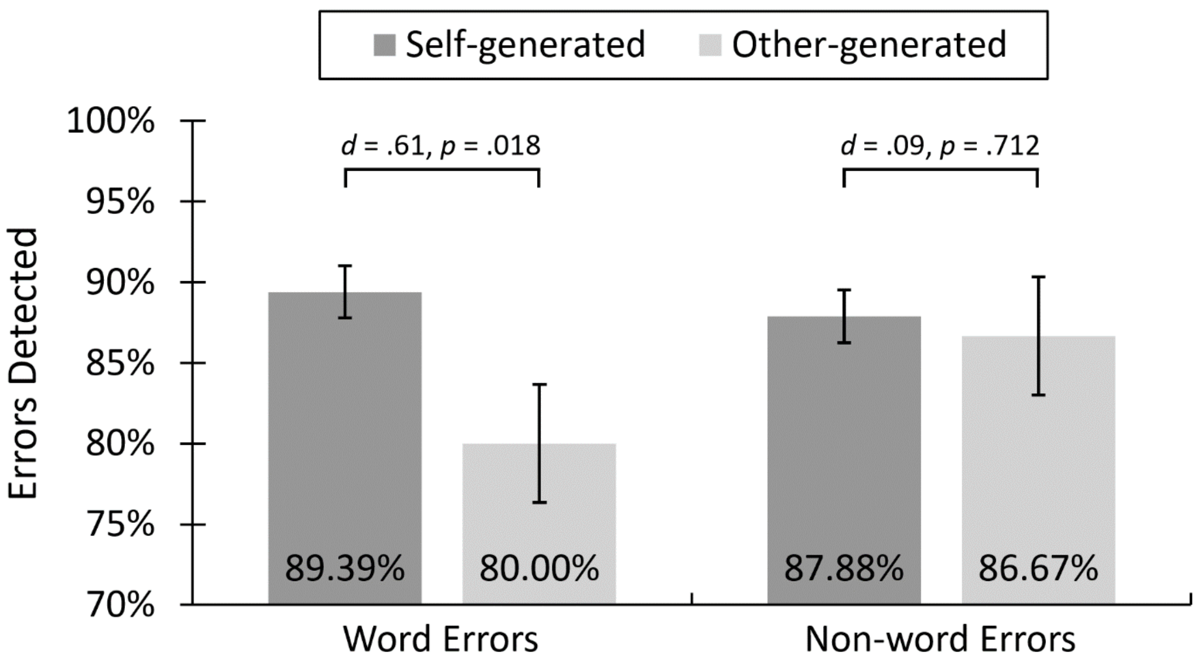


**Figure 4.** Word error and non-word error detection rates in Session 1 of Experiment 1. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Content and Function Errors.** A paired-samples *t*-test revealed that participants detected significantly more content errors (*M* = 89.09%, *SD* = 12.80%) than function errors (*M* = 84.13%, *SD* = 13.61%), *t*(62) = 2.67, *d* = .38, *p* = .010.

Content error and function error detection rates are plotted in Figure 5. The self-generated group detected 6.22% more content errors (*M* = 92.05%) than the other-generated group (*M* = 85.83%), but this difference was not significant, *t*(61) = 1.97, *d* = .50, *p* = .054. The self-generated group did not detect significantly more function errors (*M* = 86.36%) than the other-generated group (*M* = 81.67%), *t*(61) = 1.38, *d* = .35, *p* = .173.

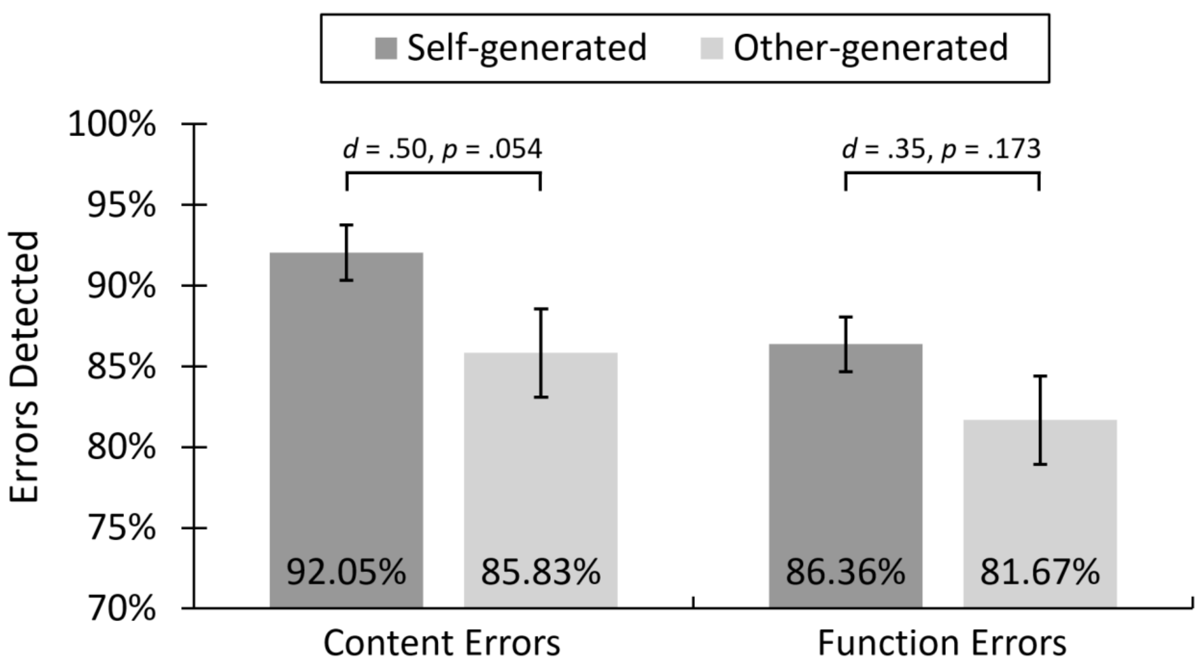


**Figure 5.** Content error and function error detection rates in Session 1 of Experiment 1. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Eye Tracking.** Eye tracking data from the proofreading task suggest that the self-generated group engaged in less thorough visual analysis than the other-generated group during

Session 1 (see Table 3). The self-generated group made fewer fixations ($M$ = 660.9) than the other-generated group ($M$ = 776.7), $t(61) = 2.30$, $d = .58$, $p = .025$. The self-generated group also made larger saccades ($M$ = 44.5 px) than the other-generated group ($M$ = 38.4 px), $t(61) = 2.21$, $d = .56$, $p = .031$. Although not significantly different, the self-generated group spent less time proofreading ($M$ = 187.8 s) than the other-generated group ($M$ = 209.4 s), $t(61) = 1.75$, $d = .44$, $p = .085$. The only result to trend opposite the predicted direction was for dwell time; the self-generated group spent slightly more time dwelling per fixation ($M$ = 284.7 ms) than the other-generated group ($M$ = 275.2 ms), $t(61) = 1.10$, $d = .28$, $p = .274$.

**Table 3.** Descriptive statistics for the proofreading task in Session 1 of Experiment 1

| **Measure** | **Self-Generated** | **Other-Generated** | **Independent Samples Test of Difference Between Groups** |
|---|---|---|---|
| **Session 1 Proofreading** | | | |
| Total errors detected (%) | 88.6 (09.0) | 83.3 (12.7) | $t(61) = 1.92$, $d = .49$, $p = .059$ |
| Word errors (%) | 89.4 (09.3) | 80.0 (20.0) | $t(61) = 2.43$, $d = .61$, $p = .018$ |
| Non-word errors (%) | 87.9 (14.3) | 86.7 (11.2) | $t(61) = 0.37$, $d = .09$, $p = .712$ |
| Content errors (%) | 92.1 (09.8) | 85.8 (15.0) | $t(61) = 1.97$, $d = .50$, $p = .054$ |
| Function errors (%) | 86.4 (11.0) | 81.7 (15.8) | $t(61) = 1.38$, $d = .35$, $p = .173$ |
| **Session 1 Eye Tracking** | | | |
| Number of fixations | 660.9 (128.0) | 776.7 (257.0) | $t(61) = 2.30$, $d = .58$, $p = .025$ |
| Saccade distance (px) | 44.5 (10.5) | 38.4 (11.4) | $t(61) = 2.21$, $d = .56$, $p = .031$ |
| Dwell time (ms) | 284.7 (30.8) | 275.2 (37.3) | $t(61) = 1.10$, $d = .28$, $p = .274$ |
| Time spent (s) | 187.8 (38.4) | 209.4 (58.7) | $t(61) = 1.75$, $d = .44$, $p = .085$ |

*Note.* Means are presented with standard deviations in parentheses.

Two eye tracking measures correlated significantly with overall proofreading performance, as shown in Table 4: number of fixations ($r = .27$, $p = .035$) and time spent proofreading ($r = .35$, $p = .005$). Participants who made more fixations or spent more time proofreading detected more errors than participants who made fewer fixations or spent less time proofreading. Three eye tracking measures correlated significantly with non-word error detection: number of fixations (*r* = .29, *p* = .023), saccade distance (*r* = -.26, *p* = .039), and time spent proofreading (*r* = .32, *p* = .011). The negative correlation with saccade distance indicates that participants who made shorter saccades detected more non-word errors.

**Table 4.** Correlations between proofreading performance and eye movements in Session 1 of Experiment 1

| Measure | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|---|
| 1. Proofreading performance | --- | | | | | | | |
| 2. Word errors detected | **.83** | --- | | | | | | |
| 3. Non-word errors detected | **.71** | .19 | --- | | | | | |
| 4. Content errors detected | **.74** | **.56** | **.58** | --- | | | | |
| 5. Function errors detected | **.91** | **.77** | **.61** | **.38** | --- | | | |
| 6. Number of fixations | **.27** | .14 | **.29** | .16 | **.27** | --- | | |
| 7. Saccade distance | -.17 | -.03 | **-.26** | -.12 | -.16 | **-.71** | --- | |
| 8. Dwell time | .17 | .19 | .06 | .16 | .13 | **-.37** | .20 | --- |
| 9. Time spent proofreading | **.35** | .23 | **.32** | .23 | **.33** | **.89** | **-.68** | .08 |

*Note.* Bolded correlations are statistically significant at $p < .05$. $N = 63$.

**Cognitive Ability.** Descriptive statistics for the cognitive ability measures are presented in Table 5. None of the cognitive ability measures correlated significantly with overall proofreading performance, word error detection, non-word error detection, or content error detection during Session 1, as shown in Table 6. However, function error detection correlated significantly with performance on the two tests of fluid intelligence, Raven’s Matrices (*r* = .26, *p* = .042) and Letter Sets (*r* = .29, *p* = .043).

**Table 5.** Descriptive statistics for cognitive ability measures in Experiment 1

| **Measure** | ***N*** | ***M*** | ***SD*** | **Range** | **Reliability** |
|---|---|---|---|---|---|
| Reading comprehension | 64 | 30.00 | 4.49 | 15.00 – 35.50 | .72 |
| Raven's matrices | 61 | 9.34 | 3.38 | 3 – 17 | .76 |
| Letter sets | 49 | 10.63 | 2.91 | 4 – 15 | .76 |
| Symmetry span | 49 | 28.37 | 7.69 | 12 – 42 | .75 |
| Visual arrays | 49 | 58.53 | 13.68 | 10 – 75 | .93 |
| Letter/number comparison | 61 | 21.11 | 3.50 | 12.50 – 29.00 | .78 |
| Pattern comparison | 49 | 31.37 | 7.15 | 8 – 48 | .68 |

*Note.* Reliability was calculated using Cronbach's alpha (α) based on item-level performance for all measures except letter/number comparison and pattern comparison, for which split-half coefficients were computed. Raven's matrices and letter/number comparison were not completed by three participants during Session 1 due to time constraints.

**Table 6.** Correlations between proofreading performance and cognitive ability in Experiment 1

| **Measure** | **1.** | **2.** | **3.** | **4.** | **5.** | **6.** | **7.** | **8.** | **9.** | **10.** | **11.** | **12.** | **13.** | **14.** | **15.** | **16.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Proofreading performance (S1) | --- | | | | | | | | | | | | | | | |
| 2. Word errors detected (S1) | **.83** | --- | | | | | | | | | | | | | | |
| 3. Non-word errors detected (S1) | **.71** | .19 | --- | | | | | | | | | | | | | |
| 4. Content errors detected (S1) | **.74** | **.56** | **.58** | --- | | | | | | | | | | | | |
| 5. Function errors detected (S1) | **.91** | **.77** | **.61** | **.38** | --- | | | | | | | | | | | |
| 6. Proofreading performance (S2) | **.37** | **.45** | .08 | .18 | **.39** | --- | | | | | | | | | | |
| 7. Word errors detected (S2) | **.33** | **.43** | .03 | .15 | **.35** | **.86** | --- | | | | | | | | | |
| 8. Non-word errors detected (S2) | .24 | .24 | .11 | .13 | .24 | **.65** | .19 | --- | | | | | | | | |
| 9. Content errors detected (S2) | **.31** | **.34** | .10 | .14 | **.33** | **.63** | **.48** | **.51** | --- | | | | | | | |
| 10. Function errors detected (S2) | .28 | **.36** | .04 | .14 | **.29** | **.88** | **.80** | **.52** | .18 | --- | | | | | | |
| 11. Reading comprehension | .01 | .08 | -.09 | .02 | .00 | .20 | .10 | .23 | .18 | .14 | --- | | | | | |
| 12. Raven's matrices | .21 | .18 | .14 | .04 | **.26** | .22 | .06 | **.32** | .25 | .12 | **.34** | --- | | | | |
| 13. Letter sets | .23 | .13 | .23 | .03 | **.29** | **.45** | **.35** | **.35** | **.35** | **.36** | **.35** | **.51** | --- | | | |
| 14. Symmetry span | .18 | .02 | .28 | .19 | .13 | .16 | .11 | .14 | .00 | .20 | .02 | .22 | **.37** | --- | | |
| 15. Visual arrays | -.12 | -.10 | -.07 | -.11 | -.09 | -.10 | -.01 | -.20 | -.18 | -.02 | .20 | .09 | .18 | **.32** | --- | |
| 16. Letter/number comparison | -.07 | -.19 | .11 | -.12 | -.02 | -.18 | -.08 | -.22 | -.14 | -.14 | .03 | -.04 | .04 | .23 | -.05 | --- |
| 17. Pattern comparison | .13 | .22 | -.05 | .24 | .03 | -.01 | -.03 | .03 | -.10 | .05 | .05 | .20 | .08 | .16 | .07 | -.11 |

*Note*. Bolded correlations are statistically significant at $p < .05$. *Ns* range from 45 to 63. S1 = Session 1; S2 = Session 2.

**Session 2 Results**

To reiterate, the names of the conditions change from Session 1 to Session 2. The “self-generated” group from Session 1 will be referred to as the “other-generated” group in Session 2 because they proofread an essay written by someone else during Session 2. The “other-generated” group from Session 1 will be referred to as the “self-generated one week ago” group in Session 2 because they proofread the essay they wrote one week prior during Session 2.

Examination of proofreading performance during Session 2 revealed further lack of support for Daneman and Stainton’s (1993) findings (see Figure 6 and Table 7). Contrary to prediction, the other-generated group detected 3.60% more errors ($M$ = 87.69%) than the self-generated one week ago group ($M$ = 84.09%) during Session 2, but this difference was not significant, $t(46) = 1.44$, $d = .42$, $p = .158$.

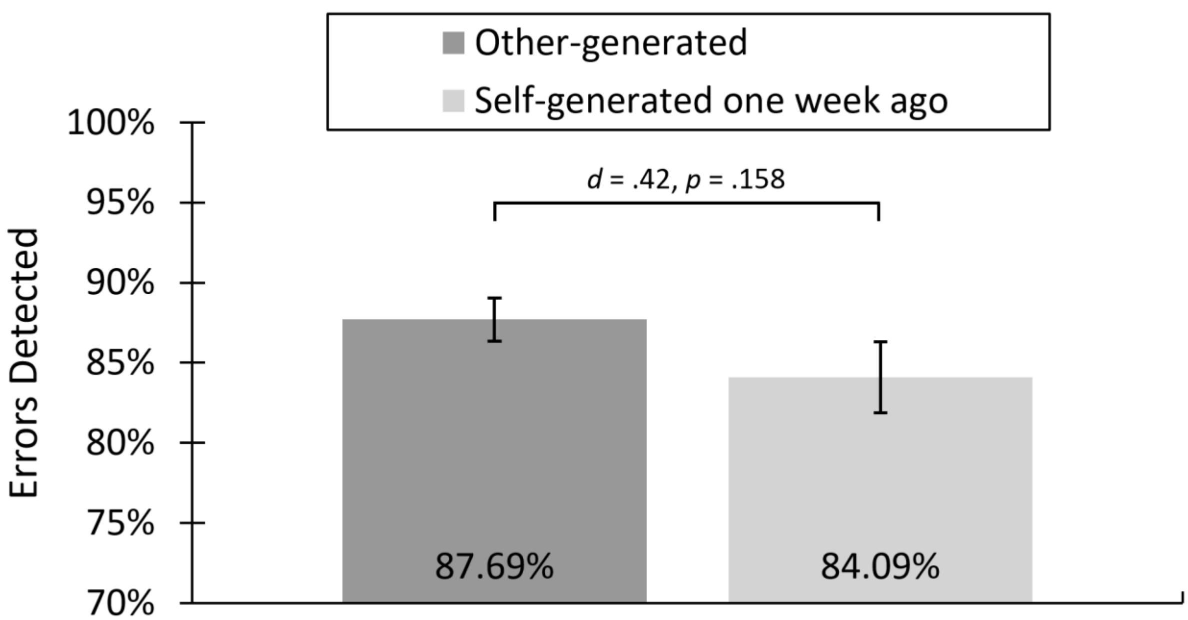


**Figure 6.** Proofreading performance in Session 2 of Experiment 1. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Table 7.** Descriptive statistics for the proofreading task in Session 2 of Experiment 1

| Measure | Other-Generated | Self-Generated One Week Ago | Independent Samples Test of Difference Between Groups |
|---|---|---|---|
| **Session 2 Proofreading** | | | |
| Total errors detected (%) | 87.7 (07.0) | 84.1 (10.3) | $t(46) = 1.44$, $d = .42$, $p = .158$ |
| Word errors (%) | 81.5 (11.9) | 80.9 (15.4) | $t(46) = 0.16$, $d = .05$, $p = .874$ |
| Non-word errors (%) | 93.8 (05.7) | 87.3 (10.8) | $t(46) = 2.70$, $d = .78$, $p = .010$ |
| Content errors (%) | 91.8 (08.6) | 88.1 (12.5) | $t(46) = 1.24$, $d = .36$, $p = .223$ |
| Function errors (%) | 84.9 (10.0) | 81.4 (13.1) | $t(46) = 1.04$, $d = .30$, $p = .302$ |
| **Session 2 Eye Tracking** | | | |
| Number of fixations | 658.1 (183.6) | 710.3 (212.4) | $t(46) = 0.91$, $d = .26$, $p = .366$ |
| Saccade distance (px) | 44.9 (12.7) | 41.6 (11.2) | $t(46) = 0.93$, $d = .27$, $p = .356$ |
| Dwell time (ms) | 265.9 (25.3) | 277.2 (46.6) | $t(46) = 1.06$, $d = .31$, $p = .293$ |
| Time spent (s) | 174.4 (46.5) | 194.5 (56.6) | $t(46) = 1.35$, $d = .39$, $p = .182$ |

*Note.* Means are presented with standard deviations in parentheses.

**Word and Non-Word Errors.** A paired-samples *t*-test revealed that participants detected significantly more non-word errors ($M$ = 90.83%, $SD$ = 8.95%) than word errors ($M$ = 81.25%, $SD$ = 13.47%) during Session 2, $t(47) = 4.51$, $d = .83$, $p < .001$. This result is consistent with Session 1, but only in Session 2 was the difference significant.

Word error and non-word error detection rates are plotted in Figure 7. The other-generated group did not detect significantly more word errors ($M$ = 81.54%) than the self-generated one week ago group ($M$ = 80.91%), $t(46) = 0.16$, $d = .05$, $p = .874$. However, the other-generated group detected significantly more non-word errors ($M$ = 93.85%) than the self-generated one week ago group ($M$ = 87.27%), $t(46) = 2.70$, $d = .78$, $p = .010$. It is unclear what accounts for this unexpected finding. Perhaps participants in the other-generated group were more motivated during Session 2 or knew which errors to look for, because they had proofread their own essay during Session 1 and recognized similar errors embedded in their writing.

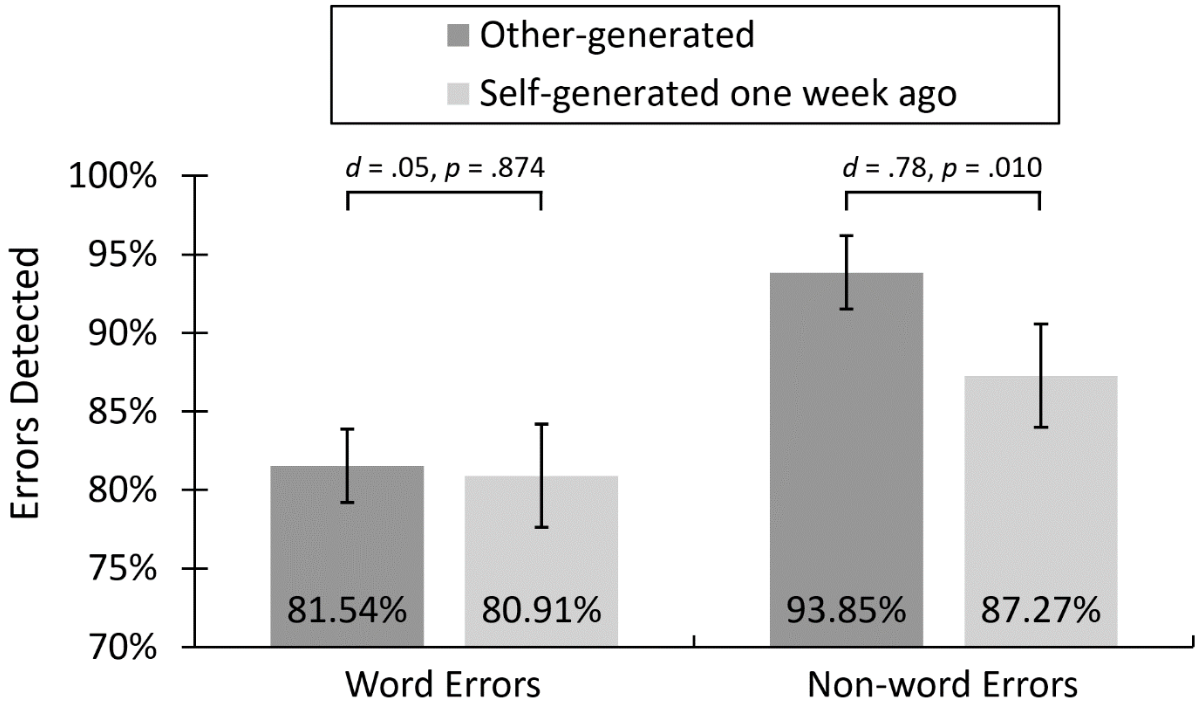


**Figure 7.** Word error and non-word error detection rates in Session 2 of Experiment 1. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Content and Function Errors.** A paired-samples *t*-test revealed that participants detected significantly more content errors ($M$ = 90.11%, $SD$ = 10.62%) than function errors ($M$ = 83.33%, $SD$ = 11.53%) during Session 2, $t(47) = 3.32$, $d = .61$, $p = .002$. This finding is consistent with the results from Session 1.

Content error and function error detection rates are plotted in Figure 8. The other-generated group did not detect significantly more content errors ($M$ = 91.85%) than the self-generated one week ago group ($M$ = 88.07%), $t(46) = 1.24$, $d = .36$, $p = .223$. The other-generated group did not detect significantly more function errors ($M$ = 84.92%) than the self-generated one week ago group ($M$ = 81.44%), $t(46) = 1.04$, $d = .30$, $p = .302$.

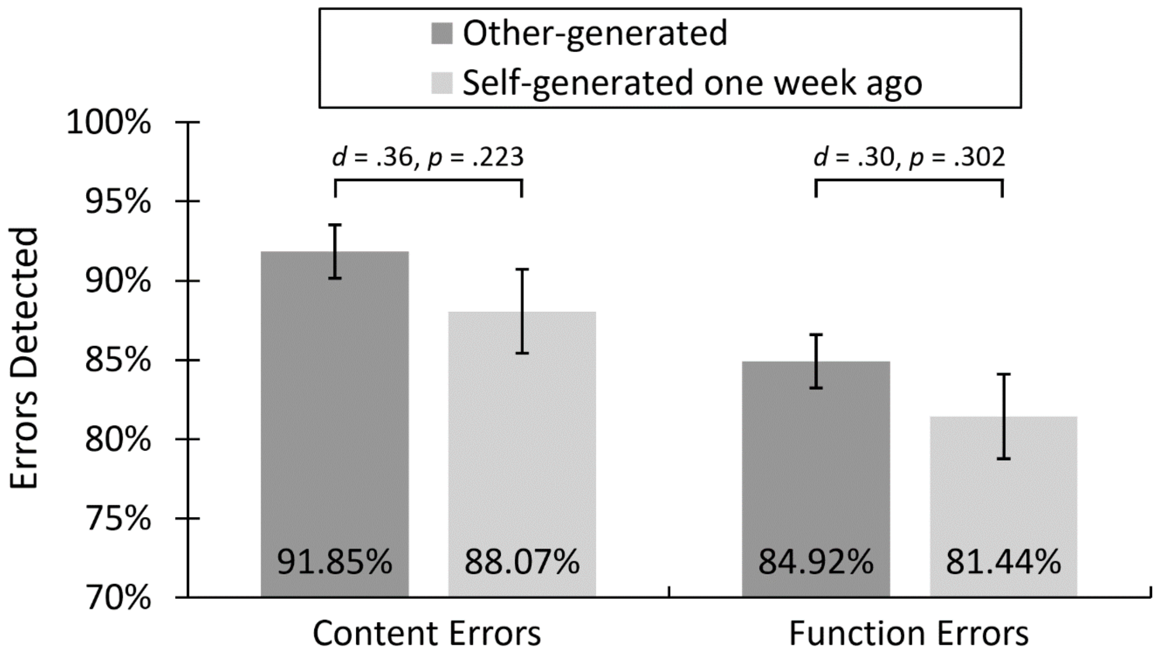


**Figure 8.** Content error and function error detection rates in Session 2 of Experiment 1. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Cognitive Ability.** Next, I examined whether performance on the cognitive ability tests correlated with proofreading performance during Session 2. Performance on Letter Sets, one of the two tests of fluid intelligence, correlated positively with proofreading performance ($r$ = .45, $p$ = .001; see the left panel of Figure 9), word error detection ($r$ = .35, $p$ = .016), non-word error detection ($r$ = .35, $p$ = .014), content error detection ($r$ = .35, $p$ = .015), and function error detection ($r$ = .36, $p$ = .013). Performance on Raven's Matrices, the other test of fluid intelligence, correlated positively with non-word error detection ($r$ = .32, $p$ = .030; see the right panel of Figure 9). These significant positive relationships indicate that participants with higher fluid intelligence detected more errors than participants with lower fluid intelligence.

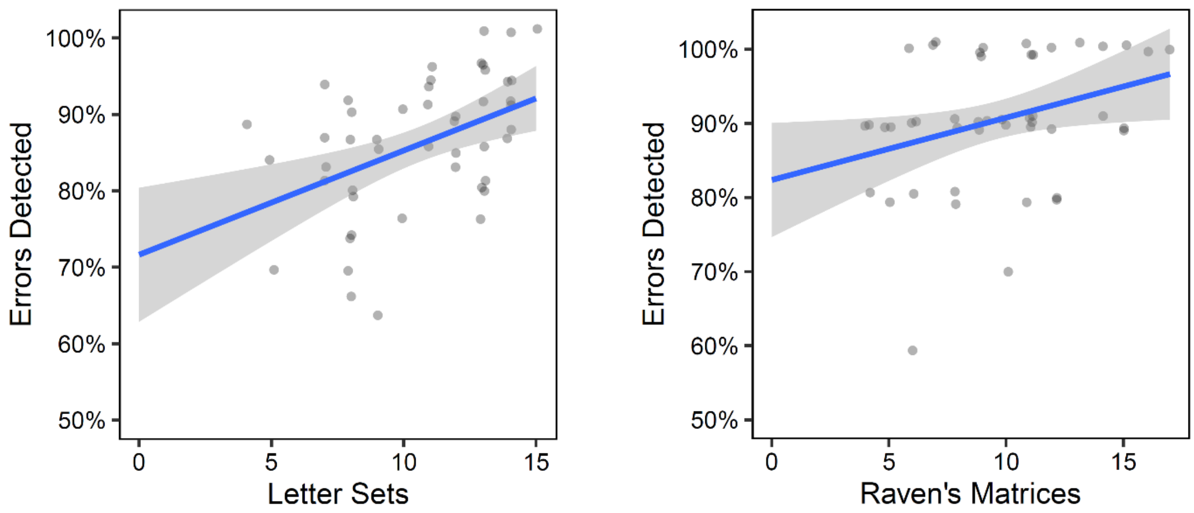


**Figure 9.** Left: scatterplot depicting correlation between Letter Sets and overall proofreading performance ($r$ = .45, $p$ = .001) during Session 2 of Experiment 1. Right: scatterplot depicting correlation between Raven's Matrices and non-word error detection rates ($r$ = .32, $p$ = .030) during Session 2 of Experiment 1. Data points have been made semi-transparent and jittered slightly for readability. Confidence bands represent 95% confidence intervals.

**Eye Tracking.** Data from the eye tracker during the proofreading task in Session 2 did not reveal any significant differences between groups (all *p*s > .18). The numerical difference between groups, however, trended opposite the predicted direction. That is, the self-generated one week ago group made slightly more fixations, shorter saccades, longer dwells, and spent more time proofreading than the other-generated group, as shown in Table 7.

As in Session 1, time spent proofreading correlated significantly with overall proofreading performance ($r$ = .30, $p$ = .042; see Table 8). However, unlike in Session 1, the correlation between number of fixations and overall proofreading performance was non-significant ($r$ = .27, $p$ = .064).

**Table 8.** Correlations between proofreading performance and eye movements in Session 2 of Experiment 1

| **Measure** | **1.** | **2.** | **3.** | **4.** | **5.** | **6.** | **7.** | **8.** |
|---|---|---|---|---|---|---|---|---|
| 1. Proofreading performance | --- | | | | | | | |
| 2. Word errors detected | **.86** | --- | | | | | | |
| 3. Non-word errors detected | **.65** | .19 | --- | | | | | |
| 4. Content errors detected | **.63** | **.48** | **.51** | --- | | | | |
| 5. Function errors detected | **.88** | **.80** | **.52** | .18 | --- | | | |
| 6. Number of fixations | .27 | .26 | .13 | .06 | **.30** | --- | | |
| 7. Saccade distance | -.24 | -.24 | -.11 | -.04 | -.28 | **-.80** | --- | |
| 8. Dwell time | .12 | .12 | .06 | .07 | .11 | -.21 | .00 | --- |
| 9. Time spent proofreading | **.30** | **.29** | .14 | .07 | **.33** | **.87** | **-.78** | .28 |

*Note.* Bolded correlations are statistically significant at $p < .05$. $N = 48$.

## Interim Discussion

Experiment 1 failed to replicate Daneman and Stainton's (1993) results, as shown in Figure 10. Comparing the current results to theirs, although the unfamiliar groups did not perform significantly differently from one another, $t(38) = 0.07$, $d = .03$, $p = .944$, the self-generated groups had a significant mean difference of 18.6%, $t(41) = 5.71$, $d = 2.06$, $p < .001$. Daneman and Stainton (1993) did not report standard deviations; estimates from Session 1 of Experiment 1 were used for these calculations. However, it is possible that participants in Session 1's self-generated group were merely *familiar* with their essays, not *overfamiliar* with them. This seems plausible: following Daneman and Stainton (1993), participants were instructed to write their essays quickly and to not worry about their literary quality. This may have discouraged participants from committing much of what they had written to memory.

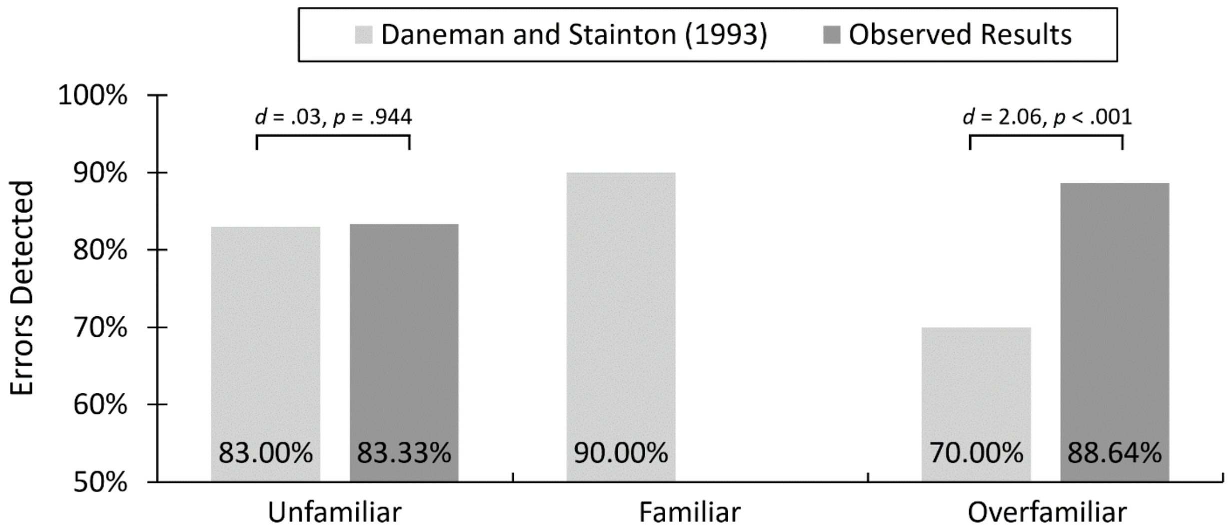


**Figure 10.** Comparison of observed results from Session 1 of Experiment 1 to Daneman and Stainton's (1993) results from Experiment 1. "Unfamiliar" refers to the other-generated condition; "Familiar" refers to participants who proofread another participant's essay after previewing it (not tested in Experiment 1); "Overfamiliar" refers to the self-generated condition. Means are presented at the base of each bar.

The results of Session 1 of Experiment 1 are consistent with this possibility, as shown in Figure 11. That is, if participants were merely "familiar" with the essays they had written, the results are similar to Daneman and Stainton (1993): the self-generated group from Session 1 and the "familiar" group from Daneman and Stainton (1993) did not perform significantly differently from one another, $t(41) = 0.42$, $d = .15$, $p = .679$. From this standpoint, familiarity facilitated proofreading and was associated with less thorough visual processing of the text, consistent with the literature (Hyönä & Niemi, 1990). Although the results of Session 1 trended in this direction, most comparisons between groups were not significant due to insufficient statistical power, a point addressed below.

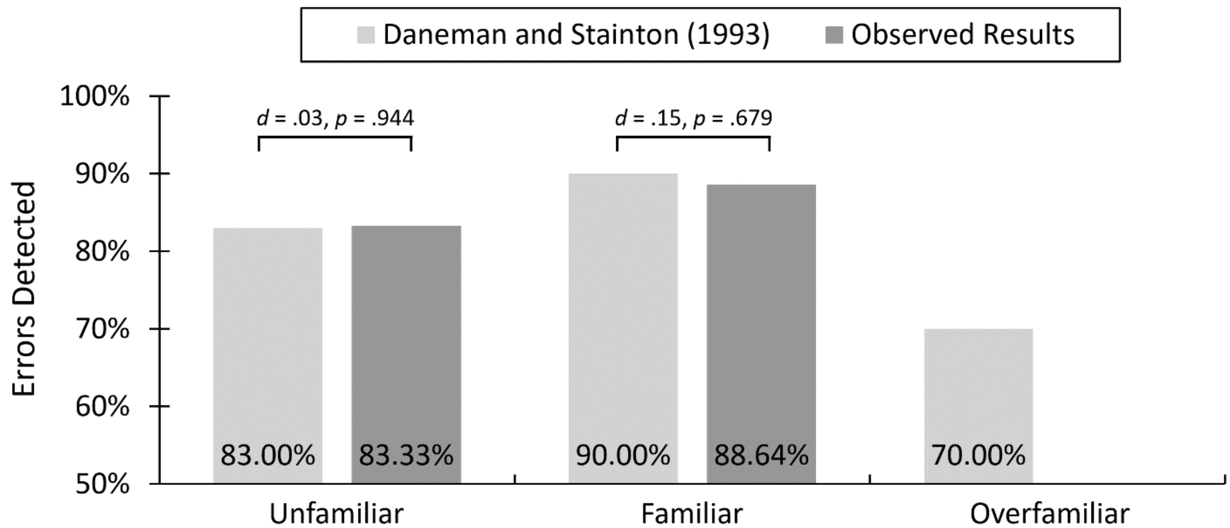


**Figure 11.** Re-plotting observed results, assuming that participants in the self-generated condition were familiar with their essays but not overfamiliar with them. Means are presented at the base of each bar.

## Experiment 2

The purpose of Experiment 2 was to attempt to induce the self-generation effect in proofreading by encouraging automated processing of self-generated text. The key manipulation was that some participants were asked to study an essay and commit it to memory prior to proofreading. Specifically, participants were assigned to a self-generated condition or an other-generated condition, as in Experiment 1, but they were also assigned to either study an essay prior to proofreading or not. Thus, there were four conditions: “self-study,” “self-no study,” “other-study,” and “other-no study.”

In addition to the study manipulation, Experiment 2 included a recognition-based memory test as a manipulation check to examine the effect of studying on essay memory. Participants also completed the cognitive reflection test to assess whether miserly cognitive processing correlated negatively with proofreading, and the need for cognition questionnaire to test whether need for cognition correlated positively with proofreading performance.

**Power Analyses**

In most cases, Experiment 1 lacked the statistical power needed to detect significant differences between groups. For example, although the difference between the self-generated and other-generated conditions' proofreading performance during Session 1 was of moderate size ($d$ = .49), it was not significant ($p$ = .059). A post-hoc power analysis revealed that at an alpha level of .05, Experiment 1 had an estimated power of .59 to detect this effect.

Nevertheless, the sample size, with *n*s of 33 and 30, exceeded Daneman and Stainton's (1993) *n*s of 10 per group in Experiment 1 and *n*s of 20 per group in Experiment 2. Daneman and Stainton (1993) found a large effect ($d \approx .99$) when comparing proofreading in the self-generated and other-generated conditions. Given the sample size, Experiment 1 would have had an estimated power of .98 to detect an effect of that magnitude.

With this in mind, in Experiment 2 I planned to recruit more participants (approximately $N$ = 100) so that the total sample could be divided into four conditions with adequate power. A sample of this size, according to G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), would yield an estimated power of .96 to detect a difference of $d$ = .99 between two groups and an estimated power of .54 to detect a difference of $d$ = .50.

**Predictions**

The first prediction for Experiment 2 concerns performance on the memory test. Participants in the self-study condition are predicted to outperform participants in the self-no study condition because the former will have typed and studied their own essays. The prediction therefore is that the studying manipulation will induce greater familiarity with the essay to be proofread. The other pairwise comparisons are exploratory; it is unclear, for example, whether memory performance will be better for the self-no study group or the other-study group.

The second prediction concerns proofreading performance. The major prediction is that studying will negatively affect proofreading in the self-generated condition but facilitate proofreading in the other-generated condition. Participants in the self-study condition are predicted to proofread worse than participants in the self-no study condition. The rationale is that participants in the self-study condition will be “overfamiliar” with their essays, such that they fail to detect errors in their own writing. By contrast, participants in the self-no study condition will be merely “familiar” with their essays, facilitating proofreading. Participants in the other-study condition are predicted to proofread better than participants in the other-no study condition. This would be consistent with the literature showing that familiarity with an essay written by someone else facilitates proofreading.

The third prediction concerns eye movements during proofreading. The prediction is that as essay familiarity increases, indications of visual processing will decrease. Therefore, participants in the self-study condition (the “most familiar” group) are predicted to make the fewest fixations, largest saccades, shortest dwells, and spend the least time proofreading. By contrast, participants in the other-no study condition (the “least familiar” group) are predicted to make the most fixations, shortest saccades, longest dwells, and spend the most time proofreading. It is unclear whether the self-no study and other-study conditions will differ, but these groups’ eye tracking results are predicted to fall between the self-study and other-no study conditions.

**Participants**

The participants were 100 undergraduate students (74 females, $M_{\text{age}} = 18.79$, $SD = 0.87$) recruited from introductory psychology courses at Michigan State University. Participants earned course credit for participating. All reported normal or corrected-to-normal vision and stated that English was their first language. All participants provided informed consent.

### Procedure

Participants were tested individually in the laboratory. They were first asked to type a short essay on student life. Next, they completed a reading comprehension test. During this time, any errors in their essays were corrected by the experimenter and 20 errors were added to the essays. Next, participants assigned to the "study" conditions spent five minutes reading an essay and committing it to memory (see "Study Time" below). Afterward, all participants completed a memory test in which they indicated whether they had seen a series of sentences before. Next, they proofread either their own essay or another participant's essay. Finally, they completed a demographic questionnaire, the need for cognition questionnaire, and the cognitive reflection test.

### Materials

**Essay Writing Task.** The essay writing task was the same as in Experiment 1.

**Adding Errors to Essays.** Twenty errors were added to the essays using the same procedure as in Experiment 1.

**Reading Comprehension Test.** Participants completed Form G of the Nelson Denny Reading Test, used in Session 1 of Experiment 1.

**Study Time.** Participants were randomly assigned to either study an essay or not. There were two study conditions: the self-study condition studied their own essay; the other-study condition studied another participant's essay. In both cases, the essay they studied was the essay they would later proofread. The essays had been corrected by the experimenter but were without errors so as not to spoil the proofreading task. The instructions were as follows: "*Please read the* *following essay three times and commit as much of it to memory as possible. There will be a memory test on this essay later on. You will have 5 minutes, starting now.*"

**Memory Test.** All participants completed a memory test in which they were shown 10 sentences, one after another, and were asked whether they had seen each sentence before. They answered using the keyboard, pressing "Y" for yes and "N" for no. Five of the sentences were foils taken from essays in Experiment 1 (e.g., "Classes are going well this semester"). The remaining five sentences were previously seen by the participant.

Participants in the study conditions were shown sentences from the essay they had studied and would later proofread. In the self-no study condition, they were shown sentences from the essay they had written and would later proofread. Importantly, participants in the other-no study condition were shown sentences they had written, *not* sentences from the essay they would later proofread. This is because in the other-no study condition, they needed to proofread an unfamiliar essay, so using sentences from the essay they would later proofread could spoil the proofreading task. These participants completed the memory test simply to match task demands across conditions.

The purpose of the memory test was to determine whether studying improved participants' familiarity with the essay they would proofread. The comparison of interest is between the self-study and self-no study conditions. It is assumed that participants in the other-study condition would be more familiar with the essay they proofread than participants in the other-no study condition, because participants in the latter condition were asked to proofread an essay they had never seen before.

**Proofreading Task.** The proofreading task was the same as in Experiment 1. Participants in the "self" conditions proofread their own essay and participants in the "other" conditions proofread another participant's essay.

**Demographic Questionnaire.** Participants reported their age, sex, GPA, and ACT score, if applicable.

**Need for Cognition.** Participants completed the 18-item need for cognition questionnaire (Cacioppo, Petty, & Feng Kao, 1984), rating their agreement with each item on a 5-point Likert scale with response options ranging from strongly agree to strongly disagree. Two example items are "I only think as hard as I have to" (reverse scored) and "I find satisfaction in deliberating hard and for long hours." The measure was the mean response to the items.

**Cognitive Reflection Test.** Participants completed the 3-item cognitive reflection test (Frederick, 2005) as a measure of their ability or disposition to consider a question and inhibit reporting the first response that springs to mind. The three items are: 1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? (answer: 5 cents). 2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (answer: 5 minutes). 3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (answer: 47 days). The measure was the number of correct responses.

## Results of Experiment 2

Overall proofreading performance is depicted in Figure 12; descriptive statistics are presented in Table 9. Examination of Figure 12 shows that, consistent with predictions, the self-study group performed worst on the proofreading task, whereas the other-study group performed best. A two-way ANOVA on overall proofreading performance was conducted, with essay condition (self vs. other) and study condition (study vs. no study) as between-subjects factors. The main effect of essay condition was not significant, $F(1, 96) = 1.96$, $\eta_p^2 = .020$, $p = .165$, nor was the main effect of study condition, $F(1, 96) = 0.02$, $\eta_p^2 = .000$, $p = .893$. The Essay Condition × Study Condition interaction trended in the predicted direction but was not statistically significant, $F(1, 96) = 1.71$, $\eta_p^2 = .018$, $p = .194$.

None of the pairwise comparisons of overall proofreading performance were significant (all $p$s $\geq .051$). Trends are reported below. The self-study group ($M$ = 82.80%) detected 6.00% fewer errors than the other-study group ($M$ = 88.80%), $t(48) = 2.01$, $d = .57$, $p = .051$. The self-study group ($M$ = 82.80%) detected 2.60% fewer errors than the self-no study group ($M$ = 85.40%), $t(48) = 0.68$, $d = .19$, $p = .498$. Conversely, the other-study group ($M$ = 88.80%) detected 3.20% more errors than the other-no study group ($M$ = 85.60%), $t(48) = 1.41$, $d = .40$, $p = .164$. Finally, the self-no study group ($M$ = 85.40%) detected roughly the same number of errors as the other-no study group ($M$ = 85.60%), $t(48) = 0.06$, $d = .02$, $p = .951$.

Broadly speaking, these results are consistent with predictions: studying facilitated error detection when proofreading someone else's essay but made error detection more difficult when proofreading one's own essay. Studying effects were small to moderate in size ($d$s of .19 and .40); the comparison between the self-study and other-study groups was moderate in size ($d$ = .57). The lack of significant differences between groups is due to insufficient power: although studying seemed to harm or help proofreading performance depending on whether participants proofread their own essay or someone else's, the effects were not strong enough to reach statistical significance.

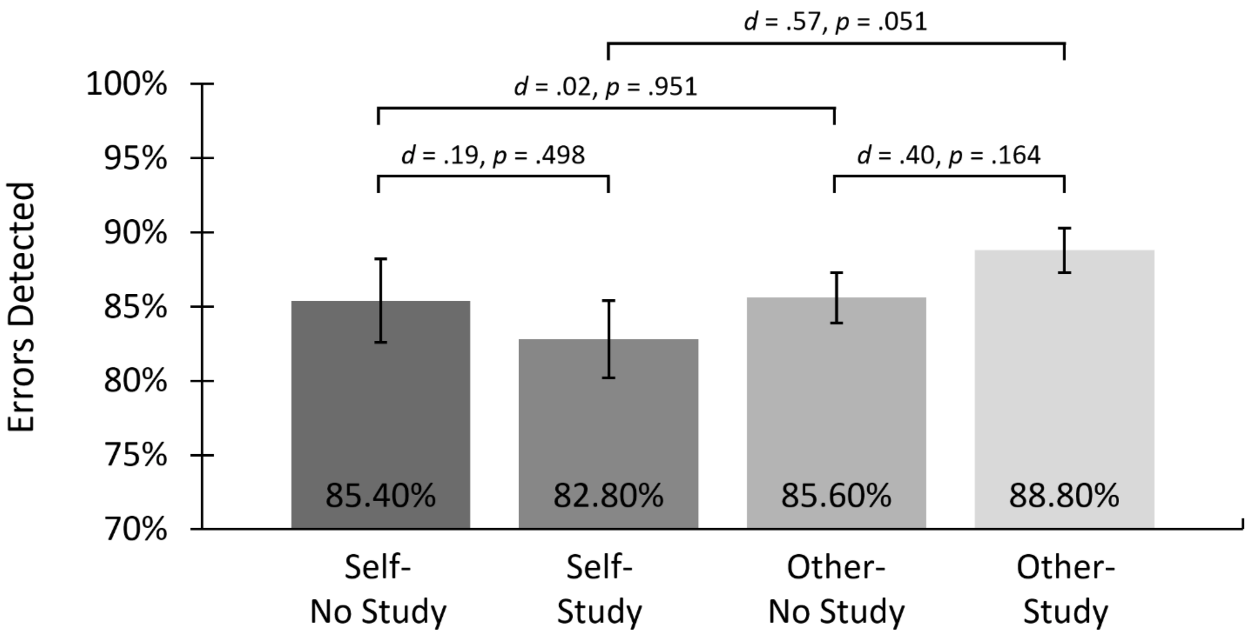


**Figure 12.** Proofreading performance in Experiment 2. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Table 9.** Descriptive statistics for proofreading, eye tracking, and memory test in Experiment 2

| Measure | Self-No Study | Self-Study | Other-No Study | Other-Study |
|---|---|---|---|---|
| **Proofreading** | | | | |
| Total errors detected (%) | 85.4 (14.0) | 82.8 (12.9) | 85.6 (08.5) | 88.8 (07.5) |
| Word errors (%) | 80.4 (18.8) | 81.2 (16.7) | 82.4 (11.3) | 88.8 (11.3) |
| Non-word errors (%) | 90.4 (11.4) | 84.8 (13.6) | 88.8 (10.5) | 88.8 (09.7) |
| Content errors (%) | 92.0 (13.9) | 90.0 (12.0) | 88.5 (10.8) | 94.0 (07.3) |
| Function errors (%) | 81.0 (16.8) | 78.0 (17.3) | 83.3 (11.0) | 85.7 (09.8) |
| | | | | |
| **Eye Tracking** | | | | |
| Number of fixations | 808.6 (314.8) | 707.0 (231.7) | 804.7 (260.9) | 857.2 (280.0) |
| Saccade distance (px) | 44.8 (14.1) | 45.5 (14.2) | 42.5 (13.1) | 40.3 (12.2) |
| Dwell time (ms) | 227.9 (28.3) | 231.2 (27.0) | 228.9 (21.6) | 234.1 (32.9) |
| Time spent (s) | 181.1 (65.4) | 162.7 (53.0) | 183.0 (57.4) | 196.7 (56.2) |
| | | | | |
| **Memory Test** | 9.6 (0.6) | 9.8 (0.5) | 9.6 (0.9) | 9.2 (0.7) |

*Note.* Means are presented with standard deviations in parentheses.

**Word and Non-Word Errors.** A paired-samples *t*-test revealed that participants detected significantly more non-word errors (*M* = 88.20%, *SD* = 11.40) than word errors (*M* = 83.20%, *SD* = 15.03%), *t*(99) = 3.30, *d* = .37, *p* = .001. This result is consistent with the results from Experiment 1.

Word error detection rates are depicted in Figure 13. Examination of Figure 13 shows that the self-no study group detected the fewest word errors, whereas the other-study group detected the most. A two-way ANOVA with word error detection as the dependent measure and essay condition and study condition as between-subjects factors was conducted. The main effect of essay condition was not significant, *F*(1, 96) = 2.60, $\eta_p^2$ = .026, *p* = .110. Also non-significant were the main effect of study condition, *F*(1, 96) = 1.46, $\eta_p^2$ = .015, *p* = .230, and the Essay Condition × Study Condition interaction, *F*(1, 96) = 0.88, $\eta_p^2$ = .009, *p* = .349.

None of the pairwise comparisons of word error detection were significant (all *p*s ≥ .051). Trends are reported below. The self-study group (*M* = 81.20%) detected 7.60% fewer word errors than the other-study group (*M* = 88.80%), *t*(48) = 1.89, *d* = .54, *p* = .065. The self-study group (*M* = 81.20%) detected roughly the same number of word errors as the self-no study group (*M* = 80.40%), *t*(48) = 0.16, *d* = .05, *p* = .874. The other-study group (*M* = 88.80%) detected 6.40% more word errors than the other-no study group (*M* = 82.40%), *t*(48) = 2.00, *d* = .57, *p* = .051. Finally, the self-no study group (*M* = 80.40%) detected 2.00% fewer word errors than the other-no study group (*M* = 82.40%), *t*(48) = 0.46, *d* = .13, *p* = .651.

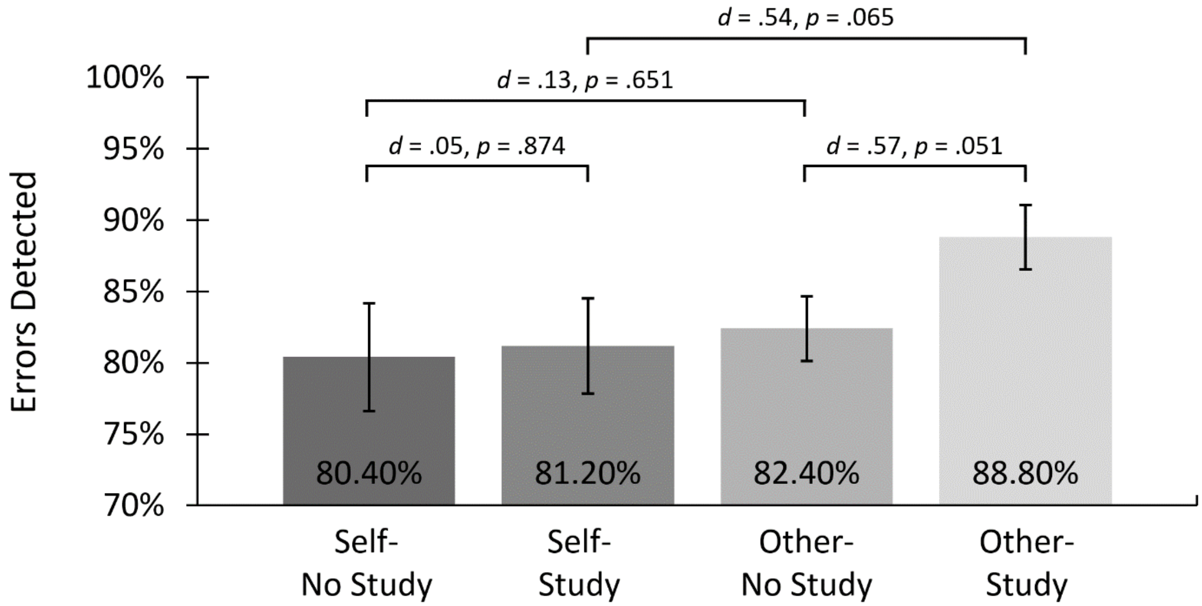


**Figure 13.** Word error detection rates in Experiment 2. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

Non-word error detection rates are depicted in Figure 14. Examination of Figure 14 shows that the self-study group detected the fewest non-word errors, whereas the self-no study group detected the most. A two-way ANOVA with non-word error detection as the dependent measure and essay condition and study condition as between-subjects factors was conducted. The main effect of essay condition was not significant, $F(1, 96) = 0.28$, $\eta_p^2 = .003$, $p = .599$, as were the main effect of study condition, $F(1, 96) = 1.51$, $\eta_p^2 = .016$, $p = .222$, and the Essay Condition × Study Condition interaction, $F(1, 96) = 1.51$, $\eta_p^2 = .016$, $p = .222$.

None of the pairwise comparisons of non-word error detection were significant (all *p*s ≥ .120). Trends are reported below. The self-study group ($M$ = 84.80%) detected 4.00% fewer non-word errors than the other-study group ($M$ = 88.80%), $t(48) = 1.20$, $d = .34$, $p = .237$. The self-study group ($M$ = 84.80%) detected 5.60% fewer non-word errors than the self-no study group ($M$ = 90.40%), $t(48) = 1.58$, $d = .45$, $p = .120$. The other-study group ($M$ = 88.80%) detected the same number of non-word errors as the other-no study group (*M* = 88.80%), *t*(48) = 0.00, *d* = .00, *p* = 1.00. Finally, the self-no study group (*M* = 90.40%) detected 1.60% more non-word errors than the other-no study group (*M* = 88.80%), *t*(48) = 0.52, *d* = .15, *p* = .608.

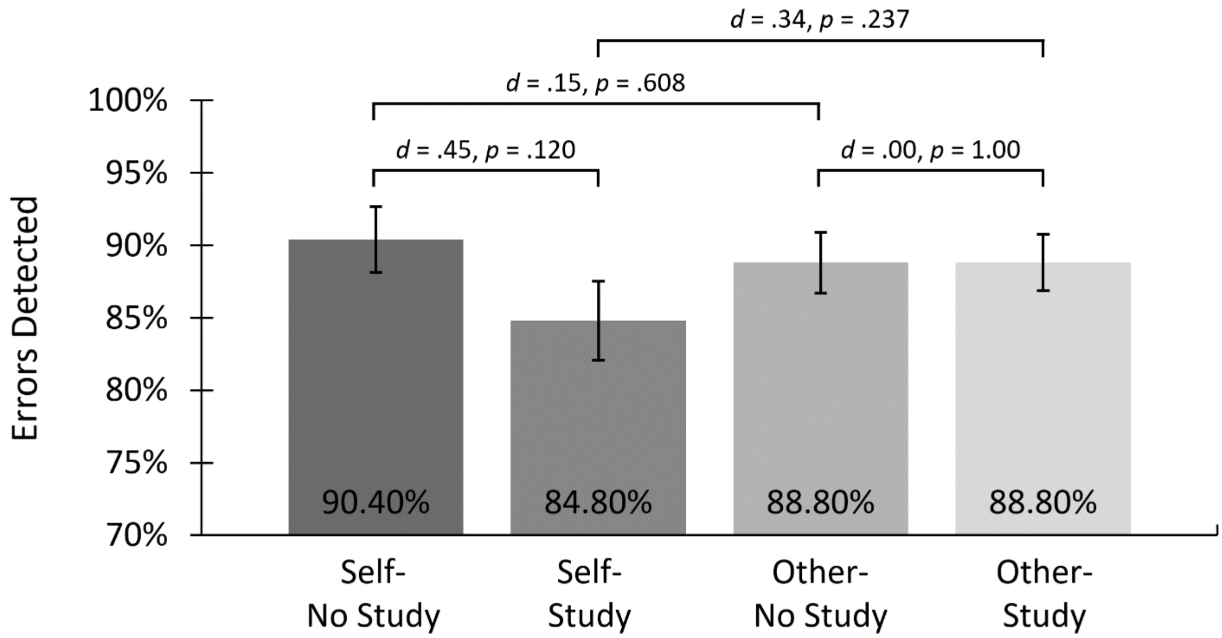


**Figure 14.** Non-word error detection rates in Experiment 2. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Content and Function Errors.** Next, I examined participants' detection of content and function errors. A paired-samples *t*-test revealed that participants detected more content errors (*M* = 91.13%, *SD* = 11.28%) than function errors (*M* = 82.00%, *SD* = 14.20%), *t*(99) = 6.40, *d* = .71, *p* < .001. This result is consistent with Experiment 1.

Content error detection rates are depicted in Figure 15. Examination of Figure 15 shows that the other-no study group detected the fewest content errors, whereas the other-study group detected the most. A two-way ANOVA with content error detection as the dependent measure and essay condition and study condition as between-subjects factors was conducted. The main effect of essay condition was not significant, $F(1, 96) = 0.01$, $\eta_p^2 = .000$, $p = .912$, as were the main effect of study condition, $F(1, 96) = 0.60$, $\eta_p^2 = .006$, $p = .439$, and the Essay Condition × Study Condition interaction, $F(1, 96) = 2.78$, $\eta_p^2 = .028$, $p = .099$.

Pairwise comparisons revealed that the self-study group ($M$ = 90.00%) detected 4.00% fewer content errors than the other-study group ($M$ = 94.00%), but this difference was not significant, $t(48) = 1.43$, $d = .40$, $p = .161$. The self-study group ($M$ = 90.00%) did not detect significantly fewer content errors than the self-no study group ($M$ = 92.00%), $t(48) = 0.55$, $d = .16$, $p = .588$. The other-study group ($M$ = 94.00%) detected significantly more content errors than the other-no study group ($M$ = 88.50%), $t(48) = 2.11$, $d = .60$, $p = .040$. Finally, the self-no study group ($M$ = 92.00%) did not detect significantly more content errors than the other-no study group ($M$ = 88.50%), $t(48) = 0.99$, $d = .28$, $p = .325$.

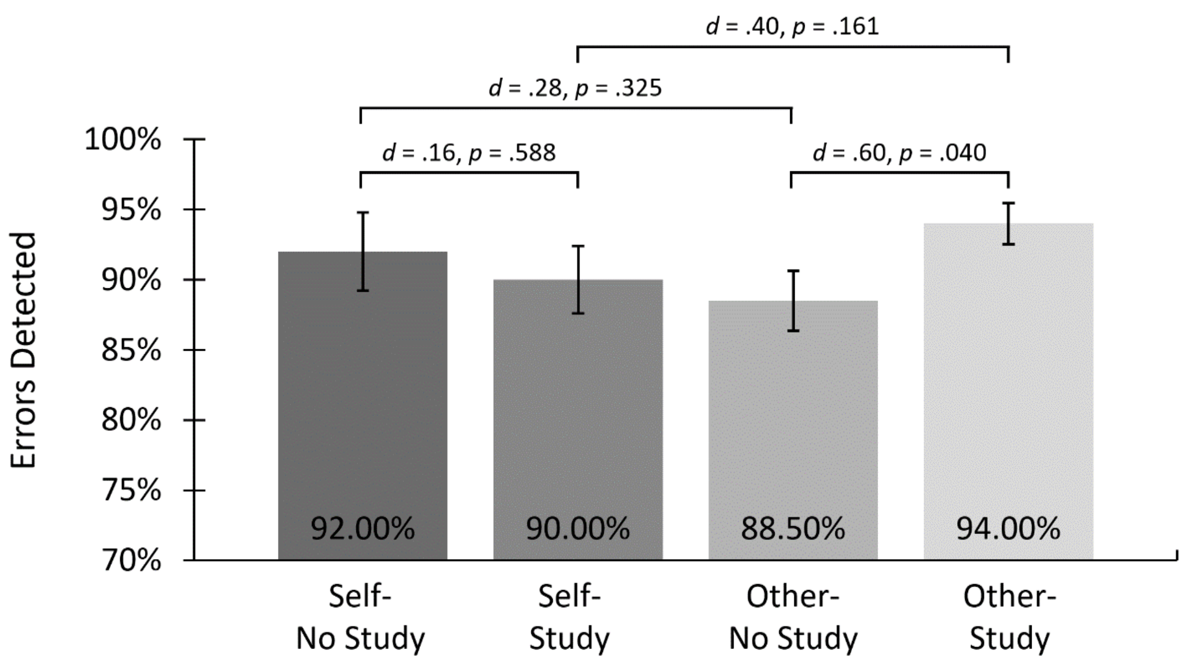


**Figure 15.** Content error detection rates in Experiment 2. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

Function error detection rates are depicted in Figure 16. Examination of Figure 16 shows that the self-study group detected the fewest function errors, whereas the other-study group detected the most. A two-way ANOVA with function error detection as the dependent measure and essay condition and study condition as between-subjects factors was conducted. The main effect of essay condition was not significant, $F(1, 96) = 3.13$, $\eta_p^2 = .032$, $p = .080$, nor was the main effect of study condition, $F(1, 96) = 0.01$, $\eta_p^2 = .000$, $p = .906$, nor was the Essay Condition × Study Condition interaction, $F(1, 96) = 0.89$, $\eta_p^2 = .009$, $p = .348$.

None of the pairwise comparisons of function error detection were significant (all $p$s $\geq$ .060). Trends are reported below. The self-study group ($M$ = 78.00%) detected 7.67% fewer function errors than the other-study group ($M$ = 85.67%), $t(48) = 1.93$, $d = .55$, $p = .060$. The self-study group ($M$ = 78.00%) detected 3.00% fewer function errors than the self-no study group ($M$ = 81.00%), $t(48) = 0.62$, $d = .18$, $p = .537$. The other-study group ($M$ = 85.67%) detected 2.34% more function errors than the other-no study group ($M$ = 83.33%), $t(48) = 0.79$, $d = .22$, $p = .432$. Finally, the self-no study group ($M$ = 81.00%) detected 2.33% fewer function errors than the other-no study group ($M$ = 83.33%), $t(48) = 0.58$, $d = .16$, $p = .564$.

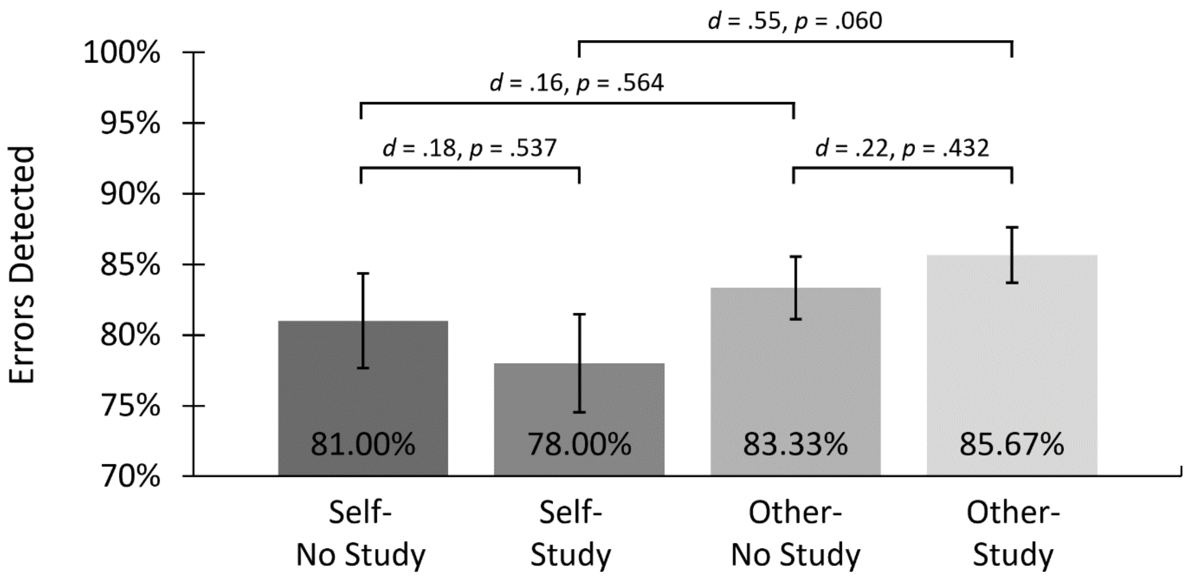


**Figure 16.** Function error detection rates in Experiment 2. Means are presented at the base of each bar. Error bars represent ± 1 standard error around the mean.

**Eye Tracking.** Data from the eye tracker during the proofreading task revealed no significant main effects of essay condition, study condition, or their interaction (all *p*s > .12). All comparisons between the self-study and self-no study groups were non-significant (all *p*s ≥ .20). Similarly, comparisons between the other-study and other-no study groups were non-significant (all *p*s ≥ .40), as were comparisons between the self-no study and other-no study groups (all *p*s > .56). However, consistent with predictions, the self-study group made significantly fewer fixations ($M = 707$) than the other-study group ($M = 857$), $t(48) = 2.07$, $d = .59$, $p = .044$. Furthermore, the self-study group spent significantly less time proofreading ($M = 163$ s) than the other-study group ($M = 197$ s), $t(48) = 2.20$, $d = .62$, $p = .032$.

As in Session 1 of Experiment 1, number of fixations correlated significantly with proofreading performance ($r = .22$, $p = .029$), as did saccade distance ($r = -.32$, $p = .001$; see Table 10). Participants who made more fixations and shorter saccades detected more errors. The correlation between time spent proofreading and proofreading performance was non-significant ($r = .19$, $p = .058$).

**Table 10.** Correlations between proofreading performance and eye movements in Experiment 2

| Measure | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|---|
| 1. Proofreading performance | --- | | | | | | | |
| 2. Word errors detected | **.88** | --- | | | | | | |
| 3. Non-word errors detected | **.77** | **.37** | --- | | | | | |
| 4. Content errors detected | **.71** | **.65** | **.51** | --- | | | | |
| 5. Function errors detected | **.93** | **.81** | **.72** | **.39** | --- | | | |
| 6. Number of fixations | **.22** | .13 | **.25** | .14 | **.20** | --- | | |
| 7. Saccade distance | **-.32** | **-.22** | **-.32** | **-.27** | **-.26** | **-.87** | --- | |
| 8. Dwell time | -.05 | -.02 | -.07 | -.01 | -.06 | **-.30** | **.28** | --- |
| 9. Time spent proofreading | .19 | .11 | **.22** | .12 | .18 | **.91** | **-.79** | .09 |

*Note.* Bolded correlations are statistically significant at $p < .05$. $N = 100$.

**Memory Test.** Next, I examined participants' performance on the memory test. Pairwise comparisons revealed that the self-study group performed slightly better on the memory test ($M = 9.76$) than the self-no study group ($M = 9.63$), but this difference was not statistically significant, $t(47) = 0.86$, $d = .25$, $p = .394$. The recognition-based memory test may not have been sensitive enough to detect the effect of studying on essay familiarity in the "self" groups. Memory performance was nearly at ceiling for both, but insufficient power also may have played a role; despite an effect of $d = .25$ the difference between groups was not significant.

The other-study group performed significantly worse on the memory test ($M = 9.20$) than the self-no study group ($M = 9.63$), $t(47) = 2.30$, $d = .66$, $p = .026$, and the self-study group ($M = 9.76$), $t(48) = 3.18$, $d = .90$, $p = .003$. Thus, memory for one's own essay was stronger than memory for someone else's essay. Recall that the other-no study group completed the memory test to match demands across conditions; their memory test sentences were taken from an essay they did not proofread, rendering their memory test data uninformative with respect to the proofreading task.

**Cognitive Ability and Personality.** Descriptive statistics for the cognitive ability and personality measures are presented in Table 11; correlations are presented in Table 12.

Performance on the reading comprehension test did not correlate significantly with proofreading performance ($r$ = .08, $p$ = .427), nor did performance on the cognitive reflection test ($r$ = .01, $p$ = .909). Need for cognition did not correlate significantly with proofreading performance ($r$ = -.08, $p$ = .463). The correlation between cognitive reflection test performance and need for cognition was significant ($r$ = .41, $p$ < .001).

**Table 11.** Descriptive statistics for cognitive ability and personality measures in Experiment 2

| Measure | *N* | *M* | *SD* | Range | Reliability |
|---|---|---|---|---|---|
| Reading comprehension | 98 | 27.10 | 5.87 | 11 – 38 | .83 |
| Cognitive reflection test | 93 | 0.59 | 0.99 | 0 – 3 | .78 |
| Need for cognition | 93 | 3.49 | 0.65 | 1.56 – 4.83 | .89 |
| ACT | 58 | 27.22 | 3.88 | 16 – 35 | - |
| GPA | 67 | 3.62 | 0.44 | 2.20 – 4.00 | - |

*Note.* Reliability was calculated using Cronbach's alpha (α). Two participants did not complete the reading comprehension test due to technical difficulties. Seven participants did not complete the non-reading comprehension measures due to time constraints. Some participants did not take the ACT; some were first year freshmen and did not yet have a college GPA.

**Table 12.** Correlations between proofreading, cognitive ability, and personality in Experiment 2

| Measure | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Proofreading performance | --- | | | | | | | | | |
| 2. Word errors detected | **.88** | --- | | | | | | | | |
| 3. Non-word errors detected | **.77** | **.37** | --- | | | | | | | |
| 4. Content errors detected | **.71** | **.65** | **.51** | --- | | | | | | |
| 5. Function errors detected | **.93** | **.81** | **.72** | **.39** | --- | | | | | |
| 6. Memory test | -.15 | -.17 | -.07 | -.13 | -.14 | --- | | | | |
| 7. Reading comprehension | .08 | .16 | -.05 | .16 | .03 | .11 | --- | | | |
| 8. Cognitive reflection test | .01 | -.03 | .06 | .02 | .00 | .10 | **.36** | --- | | |
| 9. Need for cognition | -.08 | .07 | **-.23** | .07 | -.14 | .13 | **.40** | **.41** | --- | |
| 10. ACT | .16 | .12 | .14 | .05 | .18 | -.09 | **.62** | **.48** | **.39** | --- |
| 11. GPA | .14 | .15 | .06 | .01 | .16 | .05 | **.32** | .05 | .14 | **.36** |

*Note.* Bolded correlations are statistically significant at $p$ < .05. *Ns* range from 45 to 100.

**Regression Analysis.** Next, I conducted a regression analysis to determine how much of the variance in overall proofreading performance could be accounted for by the predictors. The dependent measure was proofreading performance and the predictors were study condition, essay condition, their interaction, number of fixations, dwell time, saccade distance, reading comprehension, memory check performance, need for cognition, the cognitive reflection test, self-reported ACT, and self-reported GPA. The results are presented in Table 13.

The model accounted for 51.5% of the variance in proofreading performance, $F(12, 32) = 2.83$, MSE = .021, $p = .009$. The significant predictors were saccade distance ($\beta = -.72$, $p = .012$) and self-reported GPA ($\beta = .33$, $p = .045$). Memory test performance approached statistical significance ($\beta = -.34$, $p = .073$). Participants who made shorter saccades and had higher GPAs tended to detect more errors when proofreading.

**Table 13.** Regression analysis predicting overall proofreading performance in Experiment 2

| **Measure** | **β** | $r^2_{\text{semi-partial}}$ | ***t* statistic** | ***p* value** |
|---|---|---|---|---|
| Study condition | .35 | .033 | 1.48 | .150 |
| Essay condition | .03 | .000 | 0.12 | .908 |
| Study condition × essay condition | -.24 | .012 | -0.91 | .372 |
| Number of fixations | -.14 | .003 | -0.47 | .639 |
| Dwell time | -.19 | .024 | -1.26 | .216 |
| Saccade distance | -.72 | .108 | -2.67 | .012 |
| Reading comprehension | .06 | .002 | 0.34 | .736 |
| Memory test | -.34 | .052 | -1.85 | .073 |
| Need for cognition | -.22 | .023 | -1.24 | .225 |
| Cognitive reflection test | .01 | .000 | 0.05 | .958 |
| ACT | .01 | .000 | 0.05 | .962 |
| GPA | .33 | .066 | 2.09 | .045 |

*Note.* Overall model $R^2 = .515$. β = standardized regression coefficient. $r^2_{\text{semi-partial}}$ reflects the unique proportion of variance in proofreading performance accounted for by each predictor.

**Meta-Analysis.** Finally, I conducted a meta-analysis of the results of Experiments 1 and 2 to calculate the average self-generation effect in the proofreading task. Four standardized mean differences were entered into the analysis: self-generated vs. other-generated from Session 1 of Experiment 1; self-generated one week ago vs. other-generated from Session 2 of Experiment 1;

self-no study vs. other-no study from Experiment 2; and self-study vs. other study from Experiment 2. The standardized mean differences were treated as independent and random effects modeling was used for the analysis.

The results are depicted in Figure 17, where it can be seen that the meta-analytic average effect of proofreading one's own essay relative to someone else's was $d$ = -0.11, $CI_{95\%}$ [ -.58, .35], $p$ = .636. That is, people were only slightly worse at proofreading their own writing; this effect was non-significant. A post-hoc power analysis revealed that a sample of 4,298 participants split evenly among two groups would be needed to have 95% power to detect this effect.

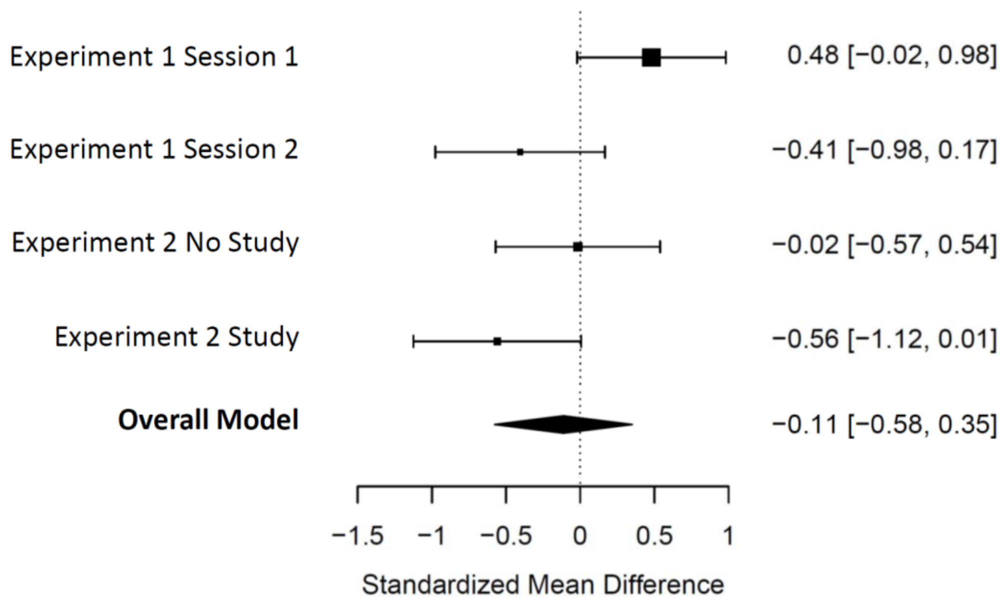


**Figure 17.** Meta-analysis of the self-generation effect in Experiments 1 and 2. Plotted values are comparisons between self-generated and other-generated conditions; positive standardized mean differences indicate that people proofread their own work more accurately than someone else's, whereas negative standardized mean differences indicate that people proofread their own work less accurately than someone else's. Error bars represent ± 1 standard error.

## Discussion

Taken together, the two experiments provide limited support for Daneman and Stainton's (1993) hypothesis that memory for self-generated text leads to an expectancy-driven processing style that makes proofreading more difficult. Experiment 1 was a near replication of Daneman and Stainton (1993), but the results trended opposite the predicted direction; students who proofread their own essays detected slightly *more* errors than students who proofread someone else's essay. One explanation for the lack of a self-generation effect is that students were not *overfamiliar* with their own essays. Experiment 2 used a studying manipulation to test this hypothesis. The prediction was that studying would facilitate proofreading another student's essay but make proofreading one's own essay more difficult. Indeed, students who studied and proofread their own essay detected the fewest errors, whereas students who studied and proofread someone else's essay detected the most errors. Moreover, students who studied and proofread their own essay appeared to engage in more expectancy-driven processing; they made significantly fewer fixations to the text and spent less time proofreading than students who studied and proofread someone else's essay.

The experiments also assessed individual differences in proofreading performance. In Experiment 1, the correlation between proofreading performance across sessions was positive and statistically significant ($r$ = .37, $p$ = .009), suggesting that individual-difference factors reliably predicted performance. Measures of fluid intelligence correlated with proofreading (*r*s ranged from .21 to .45), indicating that students with greater problem solving ability proofread more effectively. Performance on the reading comprehension tests did not significantly correlate with proofreading (*r*s ranged from .01 to .20), nor did need for cognition ($r$ = -.08, $p$ = .463) or performance on the cognitive reflection test ($r = .01$, $p = .909$). The latter two results suggest that associates of miserly cognitive processing were not predictive of proofreading performance.

There were a few points of consistency across the experiments with respect to the detection of word errors, non-word errors, function errors, and content errors. In general, non-word errors (i.e., misspellings) were more readily detected than word errors (i.e., wrong words). This finding has practical significance because word errors are also less likely to be detected by "spell check" than non-word errors. Thus, word errors appear to be especially pernicious when proofreading. I also found that content errors (e.g., *school*, *notes*, *test*) were more readily detected than function errors (e.g., *and*, *of*, *or*). However, function errors were also shorter ($M_{\text{characters}} = 3.8$, $SD = 1.2$) than content errors ($M_{\text{characters}} = 4.7$, $SD = 1.3$), $t(106) = 4.11$, $d = .79$, $p < .001$. This could drive detection rates because shorter words are less likely to be fixated (Rayner, 1998). By comparison, the length of word errors ($M_{\text{characters}} = 4.2$, $SD = 1.3$) and non-word errors ($M_{\text{characters}} = 4.2$, $SD = 1.3$) did not differ, $t(106) = 0.00$, $d = .00$, $p = 1.00$.

Examination of the eye tracking data revealed that eye movements correlated with proofreading performance. Across experiments, number of fixations was positively correlated with proofreading (*r*s ranged from .22 to .27), indicating that students who made more fixations detected more errors. Time spent proofreading also correlated with performance (*r*s ranged from .19 to .35), indicating that students who spent more time proofreading detected more errors.

**Limitations**

Three limitations of the experiments should be addressed. First, although adding errors to students' essays afforded experimental control, it also threatened external validity. That is, it is unusual to be asked to proofread an essay after being told that errors have been added to it. However, if I left students' errors in their essays instead of adding my own, the number and kind of mistakes would differ across essays, and I would not know whether students were aware of their mistakes without asking them. For example, many misspelled the word "sophomore"; it is unclear whether these students would recognize "sophmore" as a non-word error. To avoid this issue, I added errors based on relatively short, frequently-used words. I also avoided swapping words such as their/there, whether/weather, and where/were, as these are often confused. Despite the threat to external validity, a real-world analogue of this approach might be receiving copyedits from an editorial office; sometimes, a well-intentioned editor makes unwanted changes to a manuscript and the author must correct these "mistakes" embedded in their own writing.

The second limitation is that I did not investigate whether errors that went undetected were overlooked or fixated but not recognized as errors. The two cases might reflect different psychological processes. Seeing an error but failing to detect it could be the result of not thoroughly processing visual or semantic-syntactic information, perhaps due to strong top-down influences overriding visual input, whereas overlooking an error – saccading past one without fixating it – captures a skimming behavior which might also be driven by overfamiliarity. Although eye tracking data could shed light on the prevalence of these cases, it was not feasible to conduct these analyses due to time constraints. The analyses require coding interest areas around each error in each essay (e.g., 1000 interest areas in Experiment 2; 20 errors × 50 unique essays), coding whether fixations occurred within the interest areas, and coding whether the errors within those interest areas were detected. Automating this procedure will facilitate these analyses, and this is a worthwhile pursuit for a follow-up project.

The third limitation is insufficient statistical power. Although the samples in both experiments were sufficiently large to detect effects of the magnitude reported by Daneman and Stainton (1993), the effects were weaker than theirs. Nevertheless, most effects were not small.

For instance, in Experiment 2, the difference in proofreading performance between the self-study and other-study groups was $d = .57$ (larger than one half of one standard deviation), yet the associated $p$ value was $p = .051$. Greater statistical power might have also revealed significant differences between the “self” groups in the memory test of Experiment 2. Relatedly, a free recall memory test might have yielded larger differences in memory performance than the recognition test I used (Raajimakers & Shiffrin, 1992). In general, a larger sample and greater statistical power would provide more precise point estimates of observed effects. Given that my samples were larger than Daneman and Stainton’s (1993), it might have been prudent to put less confidence in their priors, which were based on small samples.

**Summary**

Two experiments were conducted to investigate automated processing of self-generated text during proofreading. The first experiment yielded results that trended opposite the predicted direction, whereas the second experiment yielded results consistent with predictions, but most group differences were not statistically significant due to insufficient power. One-hundred sixty-three participants were tested, which may be the largest sample used to investigate the self-generation effect in proofreading with an eye tracker to date. Although the results were inconclusive, the experiments demonstrate that automatic processes are difficult to study in the laboratory, perhaps more so under contrived circumstances. Future research on the self-generation effect will benefit from larger samples, a more sensitive memory test, and possibly, a more naturalistic proofreading task.

# CONCLUSION

In this dissertation I reviewed evidence indicating that automaticity poses a hidden cost of expertise. A traditional view of automaticity holds that the ability to perform well-practiced skills without attention is adaptive because it frees mental resources to process other information. Without denying the benefits of automaticity, I showed how it can also lead to expert error in domains such as driving, medical diagnosis, problem solving, and reading. In some cases, these errors arise because automatic responses are brought to bear on situations for which they are not appropriate, but the mismatch between stimulus and response is only detected after a mistake has been made. Practice and experience may reduce the prevalence of these mistakes. In other cases, knowledge and expectations automatically influence how information is processed, creating a disjunction between what is perceived and what was presented. These types of mistakes might be reduced by spending more time engaged in scrutiny and deliberation.

However, overriding automatic processes is not trivial, nor necessarily advisable in all cases. For example, in fast-paced problem-solving environments like emergency rooms, there is limited time to deliberate and consider alternative courses of action. Professionals may be overworked or sleep deprived, too, such that analytical reasoning is either compromised or too resource-intensive to be applied universally. The benefits of automaticity – that it provides fast, effortless, and often accurate responding to situations characterized by stereotypy – may outweigh the costs in low-stakes situations. But when stakes are high, as is the case when a physician makes a life-altering diagnosis, a scientist publishes the results of their research, or a driver navigates an icy highway, errors of automaticity can be costly, even dire, and deliberative processing becomes worth the time and effort.

In a pair of experiments, I attempted to induce automated processing of self-generated text by having students proofread error-filled versions of essays they had written or essays written by other students. The goal of these experiments was to shed light on the *self-generation effect*, the hypothesis that it is more difficult to spot mistakes in one's own writing than in the writing of others. The rationale is that robust memory for self-generated text exerts an automatic influence on reading behaviors, increasing the likelihood that errors are overlooked or seen but undetected. The results, however, provided only limited support for this hypothesis, and at times were at odds with the existing, albeit sparse, research on the subject. Given that the samples were larger than those in previous studies in this area, the self-generation effect in proofreading may be more tenuous than previously thought.

Ultimately, my goal was to show how automaticity causes experts to err. Understanding the situational and individual-difference factors that make errors of automaticity more likely to occur will help professionals avoid costly mistakes and could lead to the development of organizational procedures that safeguard against these types of errors. In applied contexts, the challenge is for practitioners to conduct cost-benefit analyses to determine the optimal amount of time and reflective thought required to mitigate serious mistakes without severely compromising the productivity of day-to-day procedures. These analyses will likely differ across situations and individuals. More work is needed to understand the benefits and consequences of automaticity, but this program of research stands to improve decision-making outcomes across a wide range of applied contexts.

REFERENCES

REFERENCES


Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, *24*, 225-235.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369-406.

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

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*, 610-620.

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, *130*, 701-725.

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied*, *8*, 6-16.

Beilock, S. L., Wierenga, S. A., & Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *The Quarterly Journal of Experimental Psychology Section A*, *55*, 1211-1240.

Berlioz, H. (1915). *The Orchestral Conductor: Theory of His Art*. New York, NY: Carl Fischer.

Bogner, M. S. (Ed.) (1994). *Human Error in Medicine*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, J. I., Fishco, V. V., Hanna, G. (1993). *Nelson-Denny Reading Test Forms G&H.* Austin, TX: Houghton Mifflin Harcourt.

Brown, I. D., Tickner, A. H. and Simmonds, D. C. V. (1969). Interference between concurrent tasks of driving and telephoning. *Journal of Applied Psychology, 53,* 419-424.

Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In Underwood, G. (Ed.), *Eye guidance in reading and scene perception* (pp. 125-147). Oxford, UK: Oxford University Press.

Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (2019). Is working memory capacity a causal factor in fluid intelligence? *Psychonomic Bulletin & Review*, *26*, 1333-1339.

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*, 306-307.

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275-307). New York, NY: Academic Press.

Charlton, S. G., & Starkey, N. J. (2013). Driving on familiar roads: Automaticity and inattention blindness. *Transportation research part F: Traffic psychology and behaviour*, *19*, 121-133.

Crabb, B. T., & Dark, V. J. (1999). Perceptual implicit memory requires attentional encoding. *Memory & Cognition*, *27*, 267-275.

Crabb, B. T., & Dark, V. J. (2003). Perceptual implicit memory relies on intentional, load-sensitive processing at encoding. *Memory & Cognition*, *31*, 997-1008.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159-180.

Croskerry, P. (2009). Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in Health Sciences Education*, *14*, 27-35.

Croskerry, P., Singhal, G., & Mamede, S. (2012). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Quality & Safety*, *22*(Suppl 2), ii58-ii64.

Daneman, M., & Stainton, M. (1993). The generation effect in reading and proofreading. *Reading and Writing*, *5*, 297-313.

Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human Factors*, *51*, 762-770.

Duncan, J., Williams, P., Nimmo-Smith, I., and Brown, I. D. (1992). The control of skilled behaviour: Learning, intelligence and distraction. In D. E. Meyer and S. Kornblum, (Eds.) *Attention and Performance XIV* (pp. 323-342). Cambridge, MA: MIT Press.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223-241.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition: From brains to culture*. Los Angeles, CA: Sage Publications.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks/Cole Publishing Company.

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, *5*, 914-926.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25-42.

Frenda, S. J., & Fenn, K. M. (2016). Sleep less, think worse: the effect of sleep deprivation on working memory. *Journal of Applied Research in Memory and Cognition*, *5*, 463-469.

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*, 42-54.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (2003). The psychology of good ideas. *Psychological Inquiry, 14*, 258–260.

Gidlöf, K., Wallin, A., Dewhurst, R., & Holmqvist, K. (2013). Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment. *Journal of Eye Movement Research*, *6,* 1-14.

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine*, *165*, 1493-1499.

Hayward R. A. (2002). Counting deaths due to medical errors. *JAMA*, *288*, 2404-2405.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York, NY: Oxford University Press on Demand.

Harmon, H., Ekstrom, R., French, J., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: a review. *Journal of Experimental Psychology: Applied*, *6*, 236-249.

Healy, A. F. (1976). Detection errors on the word the: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 235-242.

Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, *73*, 259-280.

James, W. (1890). *The principles of psychology* (Vol. 1, No. 2). London, UK: Macmillan.

Jonides, J., Naveh-Benjamin, M., & Palmer, J. (1985). Assessing automaticity. *Acta Psychologica*, *60*, 157-171.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329-354.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American Psychologist*, *64*, 515-526.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533-550.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163-182.

Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes*, *39*, 125-128.

Kirsch, D. (2000). A few thoughts on cognitive overload. *Intellectica*, *30*, 19-51.

Klein, J. G. (2005). Five pitfalls in decisions about diagnosis and prescribing. *BMJ*, *330*, 781-783.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*, 97-109.

Landrigan, C. P., Rothschild, J. M., Cronin, J. W., Kaushal, R., Burdick, E., Katz, J. T., Lilly, C. M., Stone, P. H., Lockley, S. W., & Czeisler, C. A. (2004). Effect of reducing interns' work hours on serious medical errors in intensive care units. *New England Journal of Medicine*, *351*, 1838-1848.

Latham, K., & Whitaker, D. (1996). A comparison of word recognition and reading performance in foveal and peripheral vision. *Vision Research, 36*, 2665-2674.

Leavitt, J. L. (1979). Cognitive demands of skating and stickhandling in ice hockey. *Canadian Journal of Applied Sport Sciences*, *4*, 46-55.

Leslie, A. M. (1987). Pretense and representation: The origins of theory of mind. *Psychological Review*, *94*, 412-426.

Levy, B. A. (1983). Proofreading familiar text: Constraints on visual processing. *Memory & Cognition*, *11*, 1-12.

Levy, B. A., & Begin, J. (1984). Proofreading familiar text: Allocating resources to perceptual and conceptual processes. *Memory & Cognition*, *12*, 621-632.

Levy, B. A., Newell, S., Snyder, J., & Timmins, K. (1986). Processing changes across reading encounters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 467-478.

Mamede, S., Schmidt, H. G., Rikers, R. M., Custers, E. J., Splinter, T. A., & van Saase, J. L. (2010). Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychological Research*, *74*, 586-592.

May, J. L., & Gale, A. G. (1998). How did I get here? Driving without attention mode. *Contemporary Ergonomics*, *50,* 456-460.

Mayo Clinic Staff. (2018, March 6). Endocarditis. Retrieved from https://www.mayoclinic.org/diseases-conditions/endocarditis/symptoms-causes/syc-20352576

Maudsley, H. (1868). *The Physiology and Pathology of Mind*. London, UK: Macmillan and Co.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503-507.

Minkoff, S. R., & Raney, G. E. (2000). detection errors in the word the: Word frequency versus syntactic structure. *Scientific Studies of Reading*, *4*, 55-76.

National Highway Traffic Safety Administration (2011). Traffic Safety Facts Research Note: Driver Electronic Device Use in 2010. DOT HS 811 517. NHTSA, Washington, DC (2011).

Neys, W. D. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428-433.

Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2017). The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine*, *92*, 23-30.

O'Regan, J. K. (1975). Structural and contextual constraints on eye movements in reading. Unpublished doctoral dissertation, University of Cambridge, Cambridge, England.

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*, 1343-1355.

Pham, J. C., Aswani, M. S., Rosen, M., Lee, H., Huddle, M., Weeks, K., & Pronovost, P. J. (2012). Reducing medical errors and adverse events. *Annual Review of Medicine*, *63*, 447-463.

Pilotti, M., & Chodorow, M. (2009). Error detection/correction in collaborative writing. *Reading and Writing*, *22*, 245-260.

Pilotti, M., & Chodorow, M. (2012). Does familiarity with text breed complacency or vigilance? *Journal of Research in Reading*, *35*, 204-214.

Raaijmakers, J. G., & Shiffrin, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology*, *43*, 205-234.

Raney, G. E., Therriault, D. J., & Minkoff, S. R. (2000). Repetition effects from paraphrased text: Evidence for an integrated representation model of text representation. *Discourse Processes*, *29*, 61-81.

Raven, J., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, *85*, 618-660.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.

Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 167-179.

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, *12*, 48-55.

Reingold, E. M., Charness, N., Schultetus, R. S., & Stampe, D. M. (2001). Perceptual automaticity in expert chess players: Parallel encoding of chess relations. *Psychonomic Bulletin & Review, 8*, 504-510.

Roberts, M. J., & Newton, E. J. (2001). Inspection times, the change task, and the rapid-response selection task. *The Quarterly Journal of Experimental Psychology: Section A*, *54*, 1031-1048.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763-776.

Schlehofer, M. M., Thompson, S. C., Ting, S., Ostermann, S., Nierman, A., & Skenderian, J. (2010). Psychological predictors of college students' cell phone use while driving. *Accident Analysis & Prevention, 42*, 1107-1112.

Schmauder, A. R., Morris, R. K., & Poynor, D. V. (2000). Lexical processing and text integration of function and content words: Evidence from priming and eye fixations. *Memory & Cognition*, *28*, 1098-1108.

Schmidt, R. A., & Lee, T. D. (2005). *Motor learning and control: A behavioral emphasis.* Champaign, IL: Human Kinetics.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66.

Smith, M. D., & Chamberlin, C. J. (1992). Effect of adding cognitively demanding tasks on soccer skill performance. *Perceptual and Motor Skills*, *75*, 955-961.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*, *11*, 3-13.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review*, *31*, 103-118.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*, 462-466.

Struckman-Johnson, C., Gaster, S., Struckman-Johnson, D., Johnson, M., & May-Shinagle, G. (2015). Gender differences in psychosocial predictors of texting while driving. *Accident Analysis & Prevention, 74,* 218-228.

Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 749-767.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20,* 147-168.

Taylor, B. P. (1975). The use of overgeneralization and transfer learning strategies by elementary and intermediate students of ESL. *Language Learning*, *25*, 73-107.

Tinker, M. A. (1927). Legibility and eye movement in reading. *Psychological Bulletin*, *24*, 621-639.

Top 10 Bushisms. (2009, January 11). Retrieved from http://content.time.com/time/specials/packages/article/0,28804,1870938_1870943_1870944,00.html

Twain, M. (1889). *A Connecticut Yankee in King Arthur's Court.* New York, NY: Harper & Brothers Publishers.

Javal, E. (1878). Essai sur la physiologie de la lecture. *Annales d'Ocilistique*, *80*, 61-73.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273-281.

West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition*, *31*, 243-251.

Wiedenbeck, S. (1985). Novice/expert differences in programming skills. *International Journal of Man-Machine Studies*, *23*, 383-390.

Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided search. In A. C. Nobre & S. Kastner (Eds.), *The Oxford Handbook of Attention* (pp. 35-44). Oxford, UK: Oxford University Press.