DISSERTATION

METACOGNITIVE STATES AND FEELINGS OF CURIOSITY: INFORMATION-SEEKING BEHAVIORS DURING MOMENTARY RETRIEVAL-FAILURE

Submitted by

Katherine L. McNeely-White

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2022

Doctoral Committee:

Advisor: Anne M. Cleary

Carol A. Seger
Kimberly Henry
Nathaniel Blanchard

ABSTRACT

METACOGNITIVE STATES AND FEELINGS OF CURIOSITY: INFORMATION SEEKING BEHAVIORS DURING MOMENTARY RETRIEVAL-FAILURE

Curiosity during learning increases information-seeking behaviors and subsequent memory retrieval success, yet the mechanisms that drive curiosity and subsequent information-seeking behaviors are poorly understood from a theoretical perspective. Hints throughout the literature suggest that curiosity may be a metacognitive signal, encouraging the experiencer to seek out additional information that will resolve a knowledge gap. Furthermore, a recently demonstrated association between a retrieval-failure-based metacognitive state (the tip-of-the-tongue state) and increased feelings of curiosity points toward an adaptive function of these states. The current study examined the relationship between curiosity and the retrieval-failure-based metacognitive states déjà vu and déjà entendu. Participants received test lists containing novel visual environment cues (Experiment 1) or novel isolated tonal sequence cues (Experiment 2) for previously studied episodes. Across both experiments, participants gave higher curiosity ratings during target retrieval failure to cue stimuli that contained previously encountered features. Further, higher curiosity ratings were given during reported déjà vu or déjà entendu, and these states were associated with increased expenditure of limited resources to discover the answer. The full pattern suggests that déjà vu and déjà entendu may drive curiosity, serve adaptive roles in encouraging further search efforts, and that curiosity may emerge due to feature-matching familiarity-detection processes.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Anne Cleary, for all of her guidance and help on this research project, in addition to being a model example of a successful woman in STEM. All of her support and counseling throughout my time in graduate school have made me the researcher I am today. Also, thank you to my husband, Dave, for always being willing to discuss research and debate the age-old question of “What is a feature?” I could not have chosen a better partner to go through graduate school with.

TABLE OF CONTENTS

Chapter 1 – Introduction

Feelings of curiosity are foundational to learning and memory, with curiosity potentially motivating information-seeking behaviors; in turn, self-motivated information-seeking subsequently strengthens the encoding of the to-be-learned information (e.g., Gruber et al., 2014; Kang et al., 2009; Wade & Kidd, 2019). Indeed, research has demonstrated that information that originally prompted intense levels of curiosity is more likely to be recollected in a subsequent memory test even after a long delay of two weeks (Kang et al., 2009). Further, Gruber et al. have shown that, when experiencing intense levels of curiosity, participants are more likely to encode incidental information, such as a face interleaved between to-be-learned trivia information (see also Murphy et al., 2021). Findings such as this from the curiosity domain have had large implications for learning in educational settings, informing instructors on how best to encourage learners within classroom settings (e.g., Arnone & Small, 1995; Lindholm, 2018; Malone, 1981; Maw & Maw, 1966; Pluck & Johnson, 2011).

The mechanisms and origins of curiosity, though, are still not well-understood by researchers. Some theorists have proposed that curiosity is a form of metacognition, which is one's awareness of their own cognitive processes (e.g., Koriat, 2007; Nelson & Narens, 1990). For example, Litman (2009) has proposed that curiosity is a metacognitive signal that arises due to a perceived gap between one's current knowledge needs and the current accessible knowledge state, which motivates the person to resolve the gap; in particular, curiosity is thought to involve a match between the to-be-learned information and the participant's capacity or urgent need to encode or

discover it (e.g., Wade & Kidd, 2019). However, despite these compelling proposals, explicit attempts to connect the metacognition literature with the curiosity literature are scarce. Therefore, the present study sought to further integrate the two bodies of literature by investigating whether the mechanisms underpinning forms of retrieval-failure-based metacognitive states, specifically familiarity-detection, might similarly drive feelings of curiosity, as there are hints throughout the literature that feelings of curiosity and feelings of familiarity may interact.

Curiosity

Despite the relative separation of research attempting to connect the curiosity and metacognition literatures, there have been many theories put forth attempting to explain the adaptive purpose of curiosity feelings, thereby providing a potential theoretical mechanism for the metacognitive signal of curiosity. Like many cognitive phenomena, curiosity has been a topic of study since the dawn of cognitive psychology, with researchers examining not just the circumstances under which it occurs and consequences of it, but also *why* humans exhibit this information-seeking behavior (e.g., Berlyne, 1950; 1958; 1960; 1962; 1966). Curiosity has been described as a desire to know or experience new information, in which accessing that information results in a reward, whether it be an external reward, such as discovering a new piece of information that will ensure survival, or an internal reward, such as the resolution of uncertainty or merely the pleasure found in acquiring information (FitzGibbon et al., 2020; Kang et al., 2009; Litman, 2005). Indeed, the notion of merely finding pleasure in acquiring non-essential information is what has puzzled researchers for some time now, as it seems at odds with the more evolutionarily plausible mechanism of curiosity, which

is to explore one's environment and discover crucial pieces of information for survival. Experiencing curiosity for information that is paramount for survival versus non-essential seem to be at opposite ends of the spectrum, and yet humans frequently engage in both. Due to this, a number of theories have emerged in an attempt to capture the purpose, function, and phenomenology of curiosity as a cognitive construct.

Curiosity-Drive Theories

One of the earliest classes of theories examining curiosity is known as *curiosity-drive theory*, sometimes referred to as *drive reduction theory* (see Litman, 2005). This class of theories proposes that curiosity can be equated with rather unpleasant experiences of uncertainty, and the reduction of those feelings of uncertainty is rewarding. The main assumption behind this class of theories is that humans strive for coherence such that there are little to no unknowns in their environment, as those could be potentially threatening for survival. Thus, whenever one encounters a novel, complex, and/or ambiguous stimulus, such as a new peculiar bug, this elicits a sense of uncomfortable uncertainty that must be resolved. The individual will seek out information, such as by inspecting or interacting with the stimulus, in order to learn about its properties and characteristics, leading to the resolution of uncertainty that could potentially threaten survival. Indeed, early research on curiosity provided support for this viewpoint, as can be seen in a series of experiments summarized by Berlyne (1966), who described curiosity as a "condition of discomfort, due to inadequacy of information, that motivates specific exploration" (p.26). Berlyne proposed that this can emerge due to both external perceptual processes, such as complex or ambiguous

visual images, or internal conceptual conflict, such as discrepancies between cognitions or beliefs.

In his experiments examining curiosity and information-seeking behaviors for visual stimuli, Berlyne (1958) presented participants with images of varying levels of ambiguity (see Figure 1 below, taken from Berlyne, 1958). The images he used were of animals, such as an elephant or bird, but were altered in a way such that some of the animal's features, such as the legs, were incongruent. For example, the elephant might have had the legs and torso of a cat, which is highly incongruent with one's pre-existing representation of an elephant. In his experiment, Berlyne presented participants with two images side-by-side, with one being the image of a congruent animal (e.g., a tiger) and the other being the altered, incongruent version of that animal (e.g., a tiger with the body of a camel). In monitoring eye movements and fixations, Berlyne found that participants tended to spend much more time fixating on the ambiguous, strange animal containing incongruent body parts.

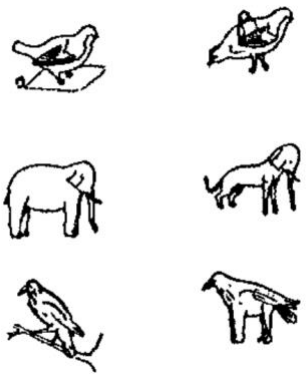

Figure 1. Stimuli used by Berlyne (1958) to examine how visual ambiguity and incongruity affect information-seeking and curiosity behaviors.

From these patterns of results, Berlyne (1966) proposed that, when presented with surprising, complex, or ambiguous stimuli, people will experience perceptual curiosity, which is interpreted as subjective uncertainty. As this is an aversive state,

resolving any feelings of curiosity is desirable. Beyond perceptual curiosity, though, Berlyne also proposed that individuals can experience conceptual conflicts that lead to curiosity-resolving behaviors, such as when internal thought processes conflict, which he coined “epistemic curiosity.” This form of curiosity is thought to reflect a more intense desire to acquire new information, and thus motivates exploratory behaviors that result in knowledge acquisition (Litman, 2005). Of specific interest to the current study is the finding that participants tend to be most curious about information that they feel is most familiar to them (Berlyne, 1954).

In his 1954 study examining epistemic curiosity, Berlyne first presented participants with a list of questions concerning invertebrate animals with the task being that they were to choose the correct answer from two potential answers. Participants were also prompted to indicate to which questions they would most like to know the answer. Upon completing this list of questions, participants reviewed a list of statements about invertebrate animals, which contained the correct answers to the previously presented list. Their only task was to read through each statement. Finally, participants were given the first list of questions again, but in the form of a free recall task. Berlyne’s results showed that participants were more likely to correctly produce the answer in the final free recall phase of the experiment for questions that prompted a desire to learn the answer (interpreted as curiosity) during the initial forced-choice recognition phase. That is, participants exhibited a memory boost for information which initially prompted feelings of curiosity, suggesting that curiosity serves as an adaptive cognitive function for future memory access. Additionally, of great interest to the current study was Berlyne’s finding that, during the initial forced-choice recognition phase, participants

were most curious about questions for which they indicated felt most familiar. When participants felt a sense of familiarity with the concept, in this case the invertebrate animal, this led to the greatest level of conflict, as it signaled that they were missing information for something that was stored within memory. Collectively, these results provide an early link between familiarity-detection, a form of metacognition, and feelings of curiosity.

The research emerging from curiosity-drive theories provides results that are informative of both the evolutionary purpose of curiosity and, of primary interest to the current study, the potential relationship between feelings of familiarity, which are a form of metacognition, and curiosity behaviors. It is clear from Berlyne's research (e.g., Berlyne, 1960) that curiosity may serve a highly adaptive function in resolving uncertainty for information that might have survival-based consequences, such as whether a new, ambiguous animal might be a threat. Resolving this curiosity by seeking out information and learning about the stimulus results in a sense of reward, as the nature of the unknown stimulus is now better understood. Further, Berlyne (1950) proposed that not all novel stimuli prompt a sense of curiosity, but rather the ones that contain familiar *and* novel elements are the ones most likely to elicit curiosity. Indeed, Berlyne (1960) proposed that instances of *relative novelty*, in which a novel stimulus contains familiar patterns of elements, might prompt greater feelings of curiosity, as it signals conflict to the individual that their knowledge structures are not complete in terms of the current concept. This might encourage further information-seeking behaviors in order to resolve the conflict and learn new, important information that might later be useful.

However, while curiosity-drive theories offered a compelling explanation of the function of curiosity, specifically that it motivates the individual to engage in information-seeking behaviors, there is extant research that offers a challenge to curiosity-drive theories. Specifically, a common, universally observed behavior in both animals and humans is that organisms tend to seek out new or novel situations (Litman, 2005). Whenever an individual is in an environment that lacks novel or complex stimuli, they are motivated to seek out stimuli that offer new, ambiguous, or complex information. In other words, organisms seek out situations offering uncertainty. This frequently observed behavior is a challenge for curiosity-drive theorists, whose central tenant is that individuals seek to *reduce* uncertainty, and thus resolve curiosity. If individuals are only engaging in information-seeking behaviors to find certainty, then why would they also engage in information-seeking behaviors to find situations that present uncertainty?

Optimal-Arousal Theories

The shortcomings of curiosity-drive theories inspired an alternative account to why animals and humans demonstrate information-seeking behavior even when the information is not relevant for survival and there are currently no elements prompting a sense of uncertainty. This new class of theories was called *optimal-arousal theories*, which proposed that organisms aim to have an optimal level of arousal in which they are not over- or under-aroused, as both situations are unpleasant (Litman, 2005; Litman & Jimerson, 2004; Silvia, 2012). The optimal-arousal theories are similar to curiosity-drive theories in that they encompass the behaviors the organisms exhibit when faced with uncertainty (e.g., a highly ambiguous stimulus that results in over-stimulation), but with the added component of boredom. When individuals lack arousal and feel bored, they

tend to seek out new experiences or stimuli that generate arousal and positive feelings of curiosity. As opposed to the curiosity-drive theorists, optimal-arousal theorists propose that the experience of curiosity itself is rewarding and involves feelings of interest and pleasure rather than negative feelings of uncertainty, although those can still occur. Individuals are motivated to seek out and engage in the learning of new information, as this is considered to be the optimal level of arousal.

However, while optimal-arousal theories can well explain the common behavior of seeking out new information due to both over- and under-arousal, they fall short when trying to explain why organisms will seek out information if this just eventually leads to boredom again (Silvia, 2012). If the end goal is to be in a constant state of curiosity, as this is the optimal level of arousal, then why would organisms seek to resolve that curiosity, such as finding out the answer to an unknown question? This would just result in a state of under-arousal or boredom, which is contradictory to the basic premise of optimal-arousal theories of curiosity, as it proposes that organisms strive to be in a state of optimal arousal. If that were the case, then organisms would not seek to *resolve* the curiosity by learning new information that fills the gap in knowledge. Indeed, as discussed by Silvia (2012), learning and exploring allow organisms to reduce feelings of uncertainty or unpleasantness, and that curiosity is like “scratching a mental itch or filling a mental hole” (p. 177). Therefore, by not resolving the curiosity and staying in an optimal state of arousal, one would never experience the satisfaction of filling the gap in knowledge, which, based on behavioral patterns, is what organisms strive for. Thus, while researchers appreciated the efforts put forth by optimal-arousal theorists, the glaring shortcoming (i.e., being unable to explain why organisms do exhibit behaviors of

curiosity resolution) dissuaded many from conducting experiments on the theory. However, one major contribution of the optimal-arousal theory, specifically its proposal that curiosity can emerge due to boredom, not just due to survival-threatening uncertainty, inspired a new way of thinking about curiosity, specifically that it can occur due to both a drive-reduction state and induction (or optimal arousal) state (Litman, 2005). This new perspective on curiosity, as discussed below, offers a means by which to frame curiosity as a form of metacognition, which is a primary goal of the current study.

Contemporary Models of Curiosity

Malone (1981), in an effort to apply findings from the curiosity literature to real-world educational settings in order to provide optimal learning environments for study, proposed that humans can experience both sensory and cognitive curiosity. The former is said to be due to surprising or complex visual patterns in nature, and the latter is said to be due to the prospect of modifying higher-order cognitive knowledge structures. As reviewed by Malone, humans strive to create complete knowledge structures through information-seeking and curiosity behaviors, thus making their cognitive concepts of the world complete, consistent, and parsimonious. Of primary interest to the current study is the proposal that a primary function of curiosity and information-seeking behaviors is to fill gaps in knowledge. Indeed, Malone outlined that one of the main purposes of curiosity is to serve as a signal of opportunity to the individual. In other words, that their current knowledge structure lacks completeness, but that there is an opportunity to discover and learn new information that can be used to reinforce existing knowledge. Although not explicitly stated by Malone, curiosity may be a metacognitive signal that

the experiencer uses to sense the gap in knowledge and engage in actions (e.g., information-seeking behaviors) in order to acquire or access the needed information.

Indeed, following from this line of thinking, that curiosity can serve as a signal to the experiencer that there is a gap in knowledge but that this gap is possible to resolve, Lowenstein (1994) proposed that people might experience curiosity due to metacognitive monitoring and control processes. When we detect that there is an information gap (formally called *information gap theory*), we seek to fill that void. The size of the information or knowledge gap is inversely related to the intensity of curiosity, as information that is perceived to be very far from access results in low levels of curiosity (e.g., the learner feels that they know an incredibly small amount about the topic and thus feels overwhelmed about the amount of information that must be learned in order to fill the void). However, when people feel that they are on the verge of discovering, learning, or accessing the information, and that there is a small perceived knowledge gap, they experience very high or intense feelings of curiosity, which in turn urges them to search further until resolution is found.

Building on the aforementioned theories of curiosity, such as curiosity drive theories (e.g., Berlyne, 1950; 1958; 1960; 1962; 1966), optimal arousal theories (e.g., Litman, 2005; Litman & Jimerson, 2004; Silvia, 2012), and Lowenstein's (1994) information gap theory, Litman and Jimerson (2004) developed the *interest-deprivation theory of curiosity*. They propose that curiosity can emerge due to either the individual feeling as though they are deprived of information and wish to reduce or resolve that gap, or due to feeling a general interest that is not caused by a specific deficit or threat, but rather an enjoyment of learning something new. The latter form of curiosity, called

curiosity as a feeling-of-interest (CFI) involves positive feelings of interest and joy, along with the anticipation of learning something new. Litman and Jimerson proposed that CFI is experienced when the person does not feel as though they are suffering from a lack of knowledge, but rather that it would be pleasurable to discover something new and avoid boredom. In a similar vein as the drive-reduction theories and optimal-arousal theories, CFI is associated with the anticipated pleasure of finding out new information.

Of primary interest to the current study, though, is the form of curiosity that Litman and Jimerson (2004) coined *curiosity as a feeling-of-deprivation* (CFD), during which the individual senses a lack of knowledge. They proposed that CFD is a much more intense curiosity experience that involves the individual sensing that the needed information is substantive, meaningful, and could increase their subjective feelings of competence. This might involve learning the answer to a complex question, a valuable fact, or perhaps the solution to a difficult problem. Crucially, they hypothesize that CFD is related to very intense feelings of curiosity that more strongly motivate information-seeking behaviors. As the individual perceives themselves as being closer to obtaining the piece of missing information, they experience more intense levels of curiosity, as the needed information is on the verge of being accessed (Noordewier & van Dijk, 2020). Indeed, as supported by other researchers, these instances of CFD are often associated with intense levels of curiosity and altered decision-making processes, such as being more willing to expend limited resources to access the target information (e.g., FitzGibbon et al., 2020; Metcalfe et al., 2017) or be more willing to wait and discover the answer (Metcalfe et al., 2021).

Such support for the proposal that curiosity can stem from a discrepancy between what is known and what one desires to know can be found in Kang et al. (2009) who examined how curiosity affects information-seeking behaviors. In a series of experiments, Kang et al. presented participants with general knowledge questions (e.g., “What instrument was invented to sound like human singing?”) and asked them to rate how curious they were to discover the answer (e.g., “Violin”). The primary question of interest concerned how curiosity would affect participants’ decision-making processes. In one condition, participants were given 25 hypothetical tokens that they could use to discover the answer to the question (note that there were 50 questions in total). In another condition, after attempting to answer the question, participants were told that they could wait to discover the correct answer, with the length of the waiting interval being between 5 to 25 seconds, or they could skip to the next question, which would result in not seeing the correct answer. If participants do indeed feel a need to resolve feelings of curiosity due to perceived knowledge gaps, then there should be an increased likelihood that they will be willing to spend limited resources, such as tokens, or wait longer to discover the answer, as this would resolve the intense need to fill the void.

Indeed, Kang et al. (2009) found that participants were significantly more likely to spend limited resources and wait longer to discover the unknown answer when they reported high levels of curiosity. In other words, intense levels of curiosity were associated with altered decision-making processes, such that participants were motivated to engage in information-seeking behaviors in order to resolve the feelings of curiosity due to perceived knowledge gaps. Further, these instances of experiencing

high levels of curiosity that in turn lead to information-seeking behaviors also have ramifications for subsequent memory access. In a follow-up experiment, Kang et al. demonstrated that memory for items which prompted the highest levels of curiosity on the initial test were better remembered later. Specifically, after completing the initial memory task of being presented with general knowledge questions and attempting to answer them, participants were brought back into the lab 11 to 16 days later for a follow-up memory test consisting of the same general knowledge questions. For the items that prompted the highest levels of curiosity during recall failure during the initial memory task, participants were significantly more likely to recall them on the follow-up memory task. These results clearly suggest that curiosity, specifically intense levels of curiosity, are associated with enhanced subsequent memory, perhaps because curiosity is a metacognitive signal that encourages both internal and external information-seeking behaviors in order to access the needed information.

Litman and Jimerson's (2004) proposal that curiosity can occur due to feelings of deprivation, or perceived gaps in knowledge, along with empirical results demonstrating the changes in decision-making processes due to curiosity during retrieval failure, have great implications for the current study. As previously mentioned, Berlyne (1950; 1960) proposed that curiosity may likely occur due to a conflict between the familiar and novel elements of a stimulus, prompting the individual to engage in information-seeking behaviors in order to resolve the uncertainty. By incorporating the information-gap theory, this proposal by Berlyne is plausible, as the individual may sense that, despite the current stimulus having familiar components, there are unknown aspects about it, signaling to the individual that they have a gap in knowledge. Indeed, as discussed by

Litman (2019), the self-awareness that one possesses a knowledge gap is a type of metacognitive judgement, as it requires that one reflects on their current level of knowledge and then engage in search processes, either internal or external, in order to fill that gap and learn more about the novel elements of the current stimulus that also somehow feels familiar.

Further, it may be that this perceived knowledge gap in and of itself may not be the only necessary factor in prompting feelings of curiosity that in turn motivate information-seeking behaviors. It may be that the experiencer must also sense a feeling of closeness to the needed information. Indeed, as previously discussed, Lowenstein's (1994) information gap theory specifically proposed that one's perceived proximity to the needed knowledge is what determines the intensity of the curiosity signals. Specifically, people may feel the most curious about information that is not currently accessible but feels as if it is *almost* accessible. As the perceived gap in knowledge narrows, people feel extremely curious, as the information is close and almost retrievable, and therefore feel motivated to discover the needed information (see also Noordewier & van Dijk, 2020). In consideration of this, it may be that the metacognitive sensations signaling to the individual that they are close to accessing the information are the necessary component for triggering feelings of curiosity, henceforth referred to as *curiosity as a feeling-of-closeness*, as information that does not feel close to being accessed will not signal to the individual that it is within the realm of possibility to discover, and they will therefore not feel curious to discover the missing information.

Metacognition

One area of research that offers insight on how feelings of closeness for inaccessible information might influence information-seeking behaviors to resolve such a knowledge gap is the study of metacognition, as researchers have established links between momentary retrieval-failure-based metacognitive states and feelings of closeness. Metacognition is the ability to think about one's own cognitions, knowledge, and memory processes (e.g., Koriat, 2007; Nelson & Narens, 1990; Rhodes, 2019), and is thought to consist of two components: monitoring and control. *Monitoring* involves the act of reflecting on one's memories (e.g., reflecting on whether one has learned enough information to pass an upcoming exam) while *control* involves the self-regulation of learning (e.g., based on monitoring processes, one does not feel that their knowledge is sufficient to pass an upcoming exam and will therefore spend more time and resources studying). Based on this, when someone reflects on their knowledge and feels that a current gap in knowledge is close to being resolved, they might feel very curious to discover that piece of information and spend limited resources to engage in information-seeking behaviors.

This proposed relationship between metacognitive sensations and the control process of engaging in information seeking-behaviors is viewed as a highly adaptive function, particularly for retrieval-failure-based metacognitive states, such as tip-of-the-tongue (TOT) states and forms of familiarity-detection during retrieval failure, like déjà vu states. For example, Schwartz and Cleary (2016) proposed that an adaptive function of retrieval-failure-based metacognitive states might be to engage in further information-seeking behaviors. Although the first attempt to conjure up details concerning the

current stimulus failed, a cognitive signal or sensation can emerge, indicating to the experiencer that *something* is held within memory and that it is potentially accessible given further search. Therefore, the metacognitive sensations that prompt further information-seeking behaviors may be similar to curiosity in that they encourage the experiencer to spend additional effort and limited resources to discovering the missing piece of information. Indeed, as discussed below, retrieval-failure-based metacognitive states, such as TOT and déjà vu, have been associated with changes in decision-making processes, some of which are adaptive information-seeking behaviors (e.g., Cleary et al., 2021; Metcalfe et al., 2017; Schwartz, 2002). Thus, these metacognitive states may further overlap with feelings of curiosity, offering insight into the potential mechanisms that drive this signal.

Tip-of-the-tongue States

As previously discussed, a key component in whether one feels a sense of curiosity for a piece of missing information is feelings of closeness. As the individual detects that they are closer to accessing the needed information, they become more curious and thus more likely to engage in information-seeking behaviors (e.g., Lowenstein, 1994; Noordewier & van Dijk, 2020). Recent empirical findings, speaking to the necessary component of feelings of closeness for information-seeking behaviors to occur concern the metacognitive state of TOT, suggest that feelings of closeness are a crucial component of TOT states (Rousseau & Kashur, 2021). During a TOT state, one feels as if the target information is right on the verge of being accessible but is momentarily unretrievable. Some have proposed that the TOT state is actually an adaptive function of the retrieval system, as it alerts the individual to the fact that

information is stored within memory, thus motivating them to further search for the inaccessible piece of information (e.g., Schwartz & Cleary, 2016; Schwartz & Metcalfe, 2011). Indeed, research has shown that participants tend to spend additional time searching for the answer while in TOT states (Schwartz, 2001) and that they are more likely to request a list of multiple-choice options when in a TOT state for an unknown answer than when not (Cleary et al., 2021).

Perhaps one reason that participants experiencing TOT states show increased information-seeking behaviors is because this metacognitive state is associated with feelings of closeness. Indeed, in a recent study, Rousseau and Kashur (2021) examined the relationship between TOT states and feelings of closeness. In their experiment, participants were given general knowledge questions, such as “What is the name of the gold-plated, humanoid robot in Star Wars?” Participants were given these questions either in a one-on-one Zoom meeting with the researcher or during a small group Zoom meeting consisting of the researcher and other participants, during which they attempted to generate the correct answer (e.g., “C-3PO”). When examining how often participants reported experiencing a TOT state for the general knowledge question, Rousseau and Kashur found that participants working in a collaborative group were significantly more likely to report experiencing a TOT state (note that participants in the group condition did not share with each other whether they were experiencing a TOT state; they only attempted to recall the answer together.).

In further examining the data, Rousseau and Kashur (2021) initially hypothesized that TOT states may be like a social contagion, in that someone is more likely to experience a TOT themselves if the person (or people) they are with are concurrently

experiencing a TOT. However, when conducting a follow-up experiment in which a confederate participated in the small group Zoom meetings and acted as if they were experiencing a TOT, there was no increase the likelihood of the participants reporting a TOT. Based on this, Rousseau and Kashur proposed that perhaps people attempting to remember information collectively are more likely to infer that retrieval is possible, as more people attempting to solve a problem increases the chances of doing so (similar to the logic behind crowdsourcing). Based on this metacognitive appraisal, it may be that participants experience stronger feelings of closeness to the target information, and that the presence of a TOT state further increases these feelings of closeness, as it signals to the experiencer that recall is possible and imminent. Thus, these findings suggest an initial link between retrieval-failure-based metacognitive states and feelings of closeness. When people experience internal signals that a piece of momentarily inaccessible information is potentially within their grasp (i.e., a TOT signal), they may feel as if they are very close to accessing that information, which may in turn drive them to be more curious about the information and engage in information-seeking behaviors.

Indeed, in a compelling example of how TOT states are associated with information-seeking behaviors in order to resolve a perceived knowledge gap, Metcalfe et al. (2017) conducted an experiment in which feelings of curiosity and their relation to TOTs were examined. Although prior research has demonstrated that people will engage in information-seeking behaviors when they perceive a knowledge gap, and that TOTs are associated with changes in decision-making processes, little research had been conducted explicitly examining the relationship between TOT states and feelings of curiosity. An earlier study conducted by Litman et al. (2005) provided initial evidence

for a relationship between feelings of curiosity and TOT states, such that participants were more curious to discover the unretrievable answer to a general knowledge question and were subsequently more willing to use limited resources to discover that answer. However, their design choices did not allow for a clear connection between in-the-moment decision-making processes and TOT states, as participants were given the opportunity to discover the answers at the end of the experiment as opposed to in the moment of retrieval failure. Additionally, Litman et al. only had 12 general knowledge questions in their experiment. Thus, Metcalfe et al. conducted a study following up on Litman et al.'s results, but with a larger number of stimuli and a design that allowed for the examination of in-the-moment changes to decision-making processes as a result of TOT states.

In their study, Metcalfe et al. (2017) presented participants with 82 general knowledge questions. After attempting to answer each question, participants were asked to indicate whether they were in a TOT state and also whether they were curious to discover the answer. However, participants were informed that, although they could see the correct answer at a later time, they could only see the correct answer for up to 10% of the questions. Although this design did not allow for immediate resolution to participants' perceived knowledge gaps, this still allowed for Metcalfe et al. to examine how TOT states might be associated with altered decision-making processes and feelings of curiosity in the moment. Indeed, their results showed that, when failing to recall the correct answer, participants were twice as likely to want to see the answer when they were in a TOT state as opposed to a non-TOT state, suggesting that

participants are more curious when experiencing a retrieval-failure-based metacognitive state, such as TOT.

Based on the findings of Rousseau and Kashur (2021) and Metcalfe et al. (2017), there appears to be a relationship between retrieval-failure-based metacognitive states, feelings of closeness, feelings of curiosity, and changes in information-seeking behaviors. These findings are in agreement with the proposals of curiosity theorists, specifically in relation to Litman and Jimerson's (2004) curiosity as a feeling-of-deprivation, which is the form of curiosity that occurs due to a perceived knowledge gap. When an individual detects that there is currently a gap in knowledge or that some piece of information is currently inaccessible, this signals to them that they should continue searching for that piece of information, either inwardly (e.g., continuing to search one's memory for a piece of momentarily inaccessible information) or outwardly (e.g., if one's memory search does not succeed, then seeking external sources, such as the internet, to access the information). In the case of retrieval-failure-based metacognitive states, such as TOT, the missing information is located within one's knowledge-base, and thus the individual uses the presence of the TOT state as a basis to justify a continued search of their memories. The TOT state signals that the experiencer is close to accessing the information, which in turn leads to intense curiosity as a feeling-of-closeness and increased information-seeking behaviors.

Indeed, evidence to support such information-seeking behaviors during TOT states can be found in Cleary et al. (2021), who provided clear evidence for how the presence of a TOT state encourages the experiencer to engage in information-seeking behaviors, such as risking limited resources in order to obtain the target information and

resolve the gap in knowledge. In their study, it was found that the presence of a TOT state can adaptively influence participants' decision-making processes while taking a test (Cleary et al., 2021). In taking the perspective that TOT states are adaptive, motivating the individual to continue searching memory as they signal to the experiencer that something is held within the knowledge-base, Cleary et al. examined how the presence of a TOT state can be used strategically on tests. Specifically, when participants failed to retrieve the target answer for a general knowledge question, they were given the opportunity to see multiple-choice options of the potential answers. However, they were told that failing to correctly select the answer from the multiple-choice options would result in a loss of points. Based on other research examining knowledge gaps and information-seeking behaviors (e.g., FitzGibbon, 2020; Gruber et al., 2014; Kang et al., 2009; Wade & Kidd, 2019), participants should be more likely to spend limited resources when experiencing a perceived gap in knowledge, as they are curious to discover the missing piece of information. Indeed, Cleary et al. found that when participants were in a reported TOT state, they were more likely to choose to see the multiple-choice options despite the potential loss of points, and were also more likely to correctly choose the answer. These findings suggest that there is an association between retrieval-failure-based metacognitive states (specifically the TOT experience), feelings of curiosity, feelings of closeness, and alterations in information-seeking decision-making processes.

Despite the growing body of literature suggesting that retrieval-failure-based metacognitive experiences affect decision-making processes, and the emerging findings that have established a link between curiosity as a feeling-of-closeness and TOT states,

explicit attempts to connect the curiosity and metacognition literatures are sparse. However, doing so may offer theoretical implications pertaining to the cognitive mechanisms driving curiosity signals. Further, as called for by Litman (2009) in his paper proposing a theoretical framework in which curiosity and metacognition interact, there is currently a gap in the literature examining how curiosity and metacognition might interact when the participant fails to remember visual and auditory information, as current research has only focused on instances in which the individual fails to remember the answers to general knowledge questions.

These gaps in the literature, though, may be better informed by reexamining a proposal put forth by Berlyne decades ago (e.g., Berlyne, 1950; 1960). Berlyne proposed that some forms of curiosity might emerge when the individual is presented with a novel stimulus that contains familiar elements, creating a strange juxtaposition between old and new elements. Although this specific proposal has yet to be investigated, there is a form of metacognition that may offer a springboard by which to elegantly examine Berlyne's proposal, while also offering theoretical insight into the mechanisms underpinning curiosity. Specifically, familiarity-detection, a metacognitive signal which is thought to occur due to feature-matching processes, may be a driving factor behind feelings of curiosity.

Familiarity-detection during Retrieval Failure

Familiarity is the feeling of having encountered something before but being unable to conjure up specific details concerning the event. However, despite being unable to access specific details of the encoded event, people can use the sensation of familiarity to recognize that they have indeed encountered this situation before. A

classic example of this phenomenon can be found in Mandler's (1980) *butcher-on-the-bus* phenomenon, in which someone encounters a person on the bus, but only feels a vague sense of familiarity with that person. They fail to recall specific details concerning the person, such as their name or how they know each other, instead only feeling a sense of familiarity for them. Because the individual detects a sense of familiarity, though, they may continue searching through their memories until they remember that the person on the bus is actually their local butcher. In this example, the sensation of familiarity occurred despite the individual failing to recollect specific details concerning the butcher on the bus and prompted further searching of memory in order to resolve the momentary lapse in memory. In other words, familiarity-detection was associated with curiosity as a feeling-of-closeness, which prompted the individual to engage in information-seeking behaviors. A goal of the present study is to examine the possibility that familiarity-detection during retrieval failure might lead to information-seeking behaviors. Such evidence pointing toward such a relationship is discussed below along with the theoretical mechanisms underlying feelings of familiarity, which may be similar to those that produce feelings of curiosity.

Feature-based Familiarity Detection

Within the past few decades, there has been a growing body of work suggesting that cognitive feature-matching processes play a large role in participants' subjective sense of familiarity during retrieval failure, and that this in turn can affect participants' decision-making processes (e.g., Cleary, 2004; Cleary et al., 2009; Huebert et al., 2021; Ryals & Cleary, 2012). As feature-based familiarity detection has been associated with changes in decision-making processes, there may be a similar relationship with curiosity

behaviors. Therefore, studying the relationship between feelings of curiosity and feelings of familiarity might offer insight into the mechanisms that drive curiosity.

The idea that feature-matching provides the basis by which familiarity-detection arises is a long-held theoretical assumption within cognitive psychology, with some calling it a pre-theoretical assumption (e.g., Tulving & Bower, 1974), and has influenced the development of many formal models of recognition memory (see Clark & Gronlund, 1996, for a review). The basic idea behind feature-matching is that when someone is presented with a stimulus, all available information is combined to create a test probe. In order to determine whether the current stimulus is old or new, the test probe's features are compared (or matched) against all of the stored memory traces, which are the prior experiences that an individual has encoded. Although the specific mechanisms by which this process takes place vary across theoretical models, the common assumption is that when a test probe's features positively match with the features stored within the memory traces, this indicates a high level of featural overlap, which signals to the individual that the current test probe is likely to be "old." In other words, if the current stimulus shares many features with information stored within memory, this stimulus will feel highly familiar, which will affect the person's decision-making processes on how to interact with the stimulus.

In order to empirically study this theoretical feature-matching process by which familiarity-detection is assumed to occur, and to identify the potential features that might be held within memory traces, research has approached the topic by using a retrieval-failure-focused paradigm known as the recognition without identification (RWI) paradigm (Cleary & Greene, 2000; 2001). In this paradigm, participants are typically presented

with a study list consisting of whole stimuli, such as words, scenes, objects, or songs (e.g., Cleary & Greene; 2000; 2001; Cleary et al., 2009; Cleary et al., 2004; Kostic & Cleary, 2009; McNeely-White & Cleary, 2019; Peynircioğlu, 1990). At test, participants are presented with fragmented stimuli, some of which correspond to whole stimuli presented at study. Their task is to discriminate between the fragmented test stimuli that were and were not presented during the study phase. During retrieval failure (that is, when participants fail to recollect the original stimulus from study), participants may experience feelings of familiarity with the fragmented test stimulus, particularly for the fragmented test stimuli that correspond to a whole stimulus presented during study. This is the *recognition without identification* phenomenon. Although participants fail to recollect specific details concerning the original encoding event, the fragmented test stimulus contains isolated features that were presented at study, which trigger feelings of familiarity due to the feature-matching process that is proposed to take place. These feelings of familiarity in turn affect the participant's decision-making process on whether or not to endorse the test stimulus as "old" or "new," with higher feelings of familiarity (which have been shown to increase with increasing featural overlap) being more likely for stimuli which are endorsed as "old." The proposal that feature-matching processes give rise to feelings of familiarity has been supported by numerous studies using the RWI and related paradigms, with researchers repeatedly demonstrating that, even when participants fail to recollect specific details concerning the encoding event, featural similarity between study and test items increases perceived feelings of familiarity, with features of different types within both visual and auditory modalities.

Feature-Matching with Visual Features

The RWI paradigm was first used by Peynircioğlu (1990) to examine feelings of familiarity and feature-matching processes using visual letters of word. In her study, Peynircioğlu first presented participants with a study list consisting of visual words, such as RAINDROP or AMYTHEST, and then a test list consisting of word fragments. Some of these word fragments corresponded to items presented during study (e.g., R_I_ _R_P) while others did not (e.g., S_Q_E_ _E). When participants failed to identify the word fragment, they were still able to discriminate between word fragments that did and did not correspond to whole words presented during study based on feelings of familiarity. Indeed, Cleary and Greene (2000) replicated this pattern of results under a variety of manipulations, providing further support for the notion that features within the test cue (e.g., R_I_ _R_P) are matched with the features stored within memory traces, and that if there is a high degree of similarity, then this produces a strong familiarity signal, which in turn affects participants' decision-making processes on a recognition memory test, such as whether or not to endorse an item as "old" or "new".

Other aspects of visual word features have also been investigated using a variant of the RWI paradigm known as the recognition without cued recall (RWCR) paradigm (Cleary, 2004). In the RWCR paradigm, participants are given a test cue, such as *bashful*, that potentially corresponds to a word presented at study, such as *bushel*. When participants fail to recall the studied word that resembles the current word, they are able to use familiarity-detection to discriminate between test cues that do and do not correspond to items presented during study. Indeed, Cleary (2004) demonstrated that when cued recall failed for a visual word cue (e.g., failing to recall the word *bushel* when

tested on the word *bashful*), participants were still able to discriminate between test cues that did and did not resemble studied words, suggesting that graphemic features are held within memory traces and can be used in familiarity-detection. Further, Cleary found that manipulating semantic and phonological similarity between study and test words also lead to an RWCR effect.

Additional research using the RWCR paradigm on visual word features has also provided support for the idea that increasing featural overlap subsequently increases subjective feelings of familiarity with the test cue (Ryals & Cleary, 2012). In their experiment, participants were tested on non-words such as *POTCHBORK*, which corresponded to either zero, one (e.g., *PITCHFORK*), or four graphemically similar studied words (e.g., *PITCHFORK, POCKETBOOK, PULLCORK,* and *PATCHWORK*). When the test cue corresponded to four graphemically similar studied words, the RWCR effect was significantly larger as demonstrated by the test cue receiving significantly higher familiarity ratings compared to cues that only had one or zero graphemically similar studied words. The findings of Ryals and Cleary suggest that, when the features of a novel visual cue had originally been encountered within a recently seen study word, participants will be more likely to experience higher levels of familiarity with the cue despite failing to recollect details concerning the original encoding experience(s). Further, as the level of feature-overlap between a test cue and the studied items increases, this subsequently increases the participant's perceived level of familiarity with the cue.

Beyond visual word features, other types of visual features have also been examined in terms of their contributions to feature-matching processes and familiarity-

detection during retrieval failure. For example, Cleary et al. (2004) demonstrated that geons, which are thought to be the basic geometric components that comprise everyday objects, might be present within memory traces and thus contribute to the computation of familiarity. Indeed, in an experimental design similar to that of Cleary and Greene (2001), in which whole items were presented at study and then fragmented versions at test, Cleary et al. presented participants with black and white line drawings of common objects, such as a coffee mug or a violin. At test, participants were presented with fragmented versions of these everyday objects, with some corresponding to whole, unfragmented items that were presented at study. Among the test cues that did correspond to items presented at study, Cleary et al. found that participants, during recall failure, were significantly more likely to find the test cue familiar compared to items that did not correspond to whole objects presented at study. Although participants failed to identify the fragmented object at test, the feature-matching process that occurred allowed for them to use familiarity-detection to discriminate between studied and unstudied items, further suggesting that visual information, such as geons, can be used to influence recognition memory-based decision-making processes.

Finally, other visual features that have been considered to take part in the feature-matching process that gives rise to feelings of familiarity are spatial relations. Specifically, the spatial layouts of a scene, such as the orientation and proximities of pieces of furniture within a room or landmarks within an outdoor environment, have been shown to be held within memory traces and elicit feelings of familiarity during retrieval failure. In a study using the method by which study and test stimuli overlap in some aspect of their features, Cleary et al. (2009) showed that, when participants were

presented with novel black and white drawings of scenes during test, they were significantly more likely to find the scenes that spatially resembled those presented at study familiar than those that did not. Despite being unable to recall the study scene that spatially resembled the current test scene, participants were still able to use familiarity-detection in their decision-making processes to discriminate between scenes that did and did not contain experimentally familiarized spatial layout features. These findings have been replicated in a number of studies, all suggesting that spatial layouts are held within memory traces and used in the feature-matching process. (e.g., Cleary & Claxton, 2018; Cleary et al., 2019; Cleary et al., 2018).

Feature-Matching with Auditory Features

A great deal of research has been conducted using the RWI and RWCR paradigms to examine how visual feature-types are involved in the feature-matching process that gives rise to familiarity-detection and enable decision-making processes to still occur even during retrieval failure. Beyond the visual domain, though, there have also been efforts in examining how auditory features might play a role in familiarity-detection during recall failure. Specifically, research has demonstrated that features such as auditory word phonemes, which are perceptually unique units of sound (e.g., the sounds of *b*, *g*, and *ch* in the words *bat*, *grass*, and *chicken*), and musical features, such as rhythm and pitch, are involved in the feature-matching process (Cleary et al., 2007; Kostic & Cleary, 2009; McNeely-White et al., 2021).

In a study following up on prior work examining visual wordform features, Cleary et al. (2007) created an auditory analog to the original experiment done by Peynircioğlu (1990), who used visual word fragments (e.g., R_I_ _R_P), by isolating auditory

phonemes from spoken words. During study, participants heard voice recordings of whole words, such as *RAINDROP* or *AMETHSYT*. At test, they were presented with auditory word fragments (e.g., the study word *RAINDROP* fragmented into the spoken sounds of *R_I_ _R_P* during the test phase). When participants failed to identify the fragmented test cue, they were more likely to provide higher familiarity ratings to auditory word fragments that corresponded to whole words spoken at study than those that did not, which is similar to findings of Peynircioğlu. Indeed, these findings established that auditory features, such as phonemes, are used in the feature-matching process that gives rise to familiarity-detection, while also suggesting that memory-based phenomenon occurring during retrieval failure found in the visual realm may also be found the auditory realm.

The research done by Cleary et al. (2007) prompted follow-up studies examining other potential auditory feature-types that may be held within memory traces and subsequently contribute to the computation of familiarity signals. Specifically, musical feature-types have been a recent focus of research using the RWI paradigm. In one of the first studies examining the role of musical features in familiarity-detection with music, Kostic and Cleary (2009) first conducted an experiment using fragmented songs to establish that an RWI effect could be found, such that participants could use familiarity-detection during retrieval failure to discriminate between studied and unstudied musical pieces. In their experiment, Kostic and Cleary digitally spliced notes from a song to create auditory song fragments (similar to how Cleary et al. spliced phonemes from a word to create auditory word fragments). They found that, when participants were tested on fragment songs that corresponded to unretrieved whole

songs presented at study, they were significantly more likely to report higher feelings of familiarity than if the fragmented test songs did not correspond to whole songs presented at study. These findings indeed suggest that musical features can be used in the feature-matching process to produce familiarity-signals in order to enable the experiencer to discriminate between studied and unstudied items.

In an effort to further examine *which* types of musical features are involved in the feature-matching process and give rise to familiarity-detection, Kostic and Cleary (2009) conducted additional experiments in which they isolated rhythm and pitch, which are thought to be two of the primary features used in music cognition (e.g., Krumhansl, 2000; Schellenberg et al., 2014). Rhythm was isolated by extracting the exact rhythm from each original song clip and tapping out the pattern on a wooden block instrument. Conversely, pitch information was isolated by extracting each original song clip's note order and attaching the notes to an arbitrary, unstudied rhythm. During their experiments, participants heard the whole, unaltered song clips at study (e.g., "Mary Had a Little Lamb") before completing the test list, which consisted of the isolated musical features, some of which corresponded to whole, unaltered studied songs. Across multiple experiments, Kostic and Cleary found that, when participants failed to identify the isolated musical feature at test, they were significantly more likely to find the feature familiar if it had come from a whole, unaltered studied song clip presented at study than if it had not. Further, although the difference did not reach significance, there was an emerging pattern such that participants found isolated tonal features more familiar than isolated rhythm features.

In following-up on Kostic and Cleary's (2009) findings, McNeely-White et al. (2021) conducted a series of experiments in which the isolated features were presented at study as opposed to test in order to assess how specific features might affect familiarity-detection with the whole, unaltered song segments at test. In their experiments, McNeely-White et al. presented participants with either isolated tonal sequences or rhythms at study, which were later embedded within whole songs at test. Their results showed that, when a whole, unaltered test song contained a familiarized feature-type, participants were significantly more likely to find the test song familiar despite experiencing retrieval failure. However, there was no significant difference between the familiarity ratings provided to test song clips that contained either experimentally familiarized isolated tonal sequences or rhythms, suggesting that there is no difference in these feature-types' contributions to the familiarity-signal computation. Although there was reason to predict that test songs containing experimentally familiarized tonal sequences would receive higher familiarity ratings during retrieval failure than test songs containing experimentally familiarized rhythms, based on the pattern of results obtained by Kostic and Cleary, no such difference was found.

A Means of Manipulating Perceived Familiarity Level

Collectively, the patterns of results obtained using the RWI and RWCR paradigms have implications not only for basic memory research, such as what types of features might be held within memory traces, but also for more applied research. Specifically, the RWI and RWCR effects suggest that, when the current stimulus shares overlapping features with stored memory traces, even when participants experience retrieval failure, they are able to use familiarity-detection in their recognition memory

decision-making processes. As previously discussed, if the current test cue contains letter features (Cleary & Greene, 2000; 2001; Peynircioğlu, 1990), visual graphemic or phonological features (Cleary, 2004; Ryals & Cleary, 2012), geons (Cleary et al., 2004), spatial relation features (Cleary & Claxton, 2018; Cleary et al., 2009; Cleary et al., 2018; Cleary et al., 2019), auditory phonological features (Cleary et al., 2007), or musical features of rhythm and pitch (Kostic & Cleary, 2009; McNeely-White et al., 2021), participants are able to use familiarity-detection during retrieval failure to discriminate between cues that do and do not resemble items presented at study. In short, familiarity-detection enables recognition memory decision-making processes to occur even during retrieval failure. The presence of heightened familiarity signals to the participant that they have indeed encountered the current stimulus before, despite not being able to recollect details about the original encoding event. One might wonder, then, what other ways familiarity-detection might be involved in decision-making processes.

Of specific interest to the current study is whether familiarity-detection might be related to increased curiosity as a feeling-of-closeness and subsequently information-seeking behaviors. In consideration of prior research, there is reason to believe that curiosity might emerge when the current stimulus contains familiarized elements, as proposed by Berlyne (1950; 1960). Berlyne proposed that some forms of curiosity may emerge when the individual is presented with a novel stimulus that contains familiarized elements, which he termed relative novelty. Indeed, Berlyne's (1956) results suggested that there is a link between familiarity and curiosity, such that participants were the most curious about familiar animals. Although Berlyne's proposal about relative novelty has

yet to be empirically examined, the RWI paradigm and familiarity-detection offer a unique means by which to investigate such a proposal, as familiarity-detection can occur when the current stimulus contains familiarized features (i.e., there is a form of relative novelty). An example can be seen in the ways in which these paradigms have been applied to the study of déjà vu, as déjà vu has been studied through a paradigm involving the juxtaposition of familiar and novel elements.

Further, the RWI paradigm and familiarity-detection allow for an expansion of the curiosity literature, as other modalities can be investigated beyond trivia questions. As previously discussed, Litman (2009) called for further investigations of the phenomenology of curiosity for information presented in visual and auditory modalities, as this is a more ecologically valid form of curiosity. Indeed, a specific form of familiarity-detection actually allows for such an investigation while also offering a unique means by which to examine Berlyne's (1950; 1960) relative novelty proposal. Specifically, the déjà vu experience offers a means by which to collectively address multiple unanswered research questions concerning the broad connections between the curiosity and metacognition literatures, potential mechanisms behind curiosity, factors that drive information-seeking behavior, and how individuals might experience curiosity as a feeling-of-closeness for visual and auditory information.

Déjà vu Experiences

Déjà vu is a jarring metacognitive state in which the individual feels as if they have experienced the current situation before despite knowing otherwise (Brown, 2003). The individual feels a strong sense of familiarity with the current scenario, most often a scene, but knows that this specific scenario has not been experienced previously.

According to researchers who study the déjà vu experience from an empirical perspective, déjà vu may represent a form of familiarity-detection that occurs in everyday life due to featural overlap between the current scenario and prior experiences (e.g., Cleary, 2008; Cleary et al., 2009; Cleary & Claxton, 2018; Cleary et al., 2018; Cleary et al., 2019). Indeed, research using the RWI paradigm has found support for the proposal that déjà vu occurs due to the current novel situation sharing similar elements with a previously encountered situation, which harkens back to Berlyne's (1950; 1960) relative novelty proposal. In an initial attempt to empirically study the mechanisms that might support déjà vu experiences, Cleary et al. (2009) presented participants with black and white line drawings of scenes that configurally resembled scenes that had been presented at study. When participants failed to recall the name of the study scene that configurally mapped onto the current test scene, they provided significantly higher familiarity ratings if the test scene did indeed correspond to a scene presented at study, thus establishing that participants can use familiarity-detection to discriminate between studied and unstudied items, and were also significantly more likely to report experiencing a sense of déjà vu for the test scene. This was the first demonstration of déjà vu being studied in an empirical setting, providing support for the idea that spatial featural overlap between the current and previously encountered scenes results in high feelings of familiarity and an increased probability of reporting déjà vu. Of interest to the current study, though, is the circumstances under which déjà vu is thought to occur. Specifically, déjà vu is more likely to occur when the current novel stimulus contains a pattern of experimentally familiarized elements. Based on this, might a sense of curiosity also be present in déjà vu experiences?

Following up on Cleary et al.'s (2009) findings, Cleary et al. (2012) used fully immersive 3D scenes in a virtual reality environment to examine how dynamically unfolding scenes that spatially resemble scenes presented at study might prompt a sense of déjà vu, as this is a more ecologically valid scenario. As in Cleary et al. (2009), Cleary et al. (2012) manipulated the spatial configuration of scenes to control for the amount of featural overlap between study and test scenes by creating all environments on a grid (see Figure 2 below). Thus, during the test phase, participants were presented with test scenes that potentially adhered to the spatial layouts that were studied during the encoding phase but were contextually different (e.g., the test scene of a reception area contained the exact spatial layout as the study scene of an aquarium). Like in Cleary et al. (2009), even during retrieval failure, participants were significantly more likely to find the test scene familiar if it spatially resembled a scene presented at study and were also significantly more likely to report a sense of déjà vu for the scene.

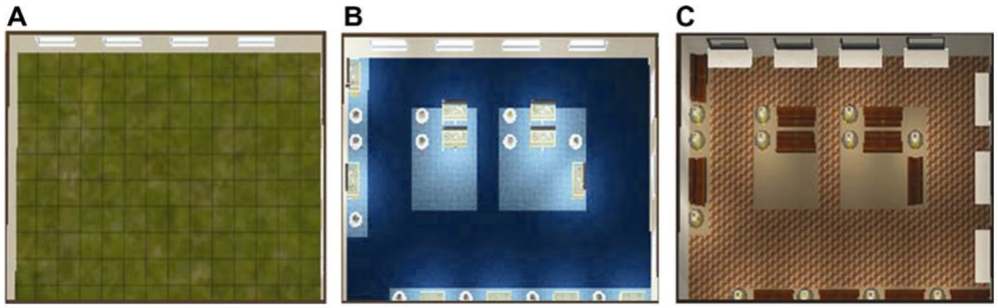

Figure 2. Taken from Cleary et al. (2012). Panel A represents the grid on which all study and test scenes were created in order to carefully manipulate spatial relation featural overlap. Panel B represents a sample study scene (Aquarium) which contains spatial relations that are embedded within the sample test scene (Reception Area) in Panel C.

In consideration of the empirical research conducted on the metacognitive experience of déjà vu, this strange metacognitive state might indeed offer a unique means by which to examine Berlyne's (1950; 1960) proposal that curiosity might emerge when the individual is presented with a novel stimulus containing a familiar

pattern of elements, as this indicates to the individual that they have a gap in their knowledge structures and should seek out information to fill this void. Indeed, as previously discussed, the déjà vu experience is a retrieval-failure-based metacognitive state that meets these criteria, as it is thought to occur when a *novel* scenario contains *familiarized* spatial elements. Therefore, this metacognitive state offers a unique means by which to investigate Berlyne's proposal while also addressing Litman's (2009) call for further research to be conducted on curiosity and metacognitive states for visual and auditory information, as the literature has thus far only examined trivia questions, and real-world experiences are more dynamic, involving visual and auditory modalities. Examining the relationship between déjà vu experiences and increased curiosity as a feeling-of-closeness would further our theoretical understanding of what circumstances lead to intense feelings of curiosity.

Déjà Entendu

Further, although the déjà vu experience is a retrieval-failure-based metacognitive state for visual information, there is a subtype of the experience for auditory information, which is déjà entendu. Déjà entendu is the feeling of having heard something before despite knowing that it is completely new (Brown, 2003; 2004; Cleary & Brown, 2022; McNeely-White & Cleary, 2019). Being a subtype of the déjà vu phenomenon, this retrieval-failure-based metacognitive phenomenon that occurs for auditory stimuli has been shown to occur due to similar feature-matching and familiarity-detection processes as déjà vu. In the first study to empirically examine déjà entendu, McNeely-White and Cleary (2019) proposed that déjà entendu may share similar mechanisms with déjà vu, such that it may occur due to familiarity-based recognition

that emerges when the current test stimulus overlaps in features with memory traces, specifically auditory features. To create auditory analogs to the visual scenes used in déjà vu experiments, McNeely-White and Cleary used NPR's Piano Puzzlers, which are re-writes of original famous songs, such as children's tunes, pop-songs, or folk songs, in the style of a classical composer, such as Mozart, Chopin, or Bach, by composer Bruce Adolphe. The original song is embedded within the Piano Puzzler such that some of the original features are intact yet masked by the composer's genre. For example, the familiar song "The Girl from Ipanema" might be turned into a Piano Puzzler by embedding phrases of the original harmony within a novel song written in the style of Brahms. Based on Adolphe's methods of creating Piano Puzzlers, McNeely-White and Cleary proposed that this juxtaposition between old and new songs may make Piano Puzzlers a strong candidate for eliciting déjà entendu, as each Piano Puzzler retains the original melody within a novel context.

In their experiment, participants were first presented with the original versions of well-known songs before being presented with the test list of Piano Puzzlers (McNeely-White & Cleary, 2019). During the test list, half of the Piano Puzzlers corresponded to original songs presented at study while the other half did not. Participants were asked to indicate whether they were experiencing déjà entendu for the Piano Puzzler, whether they found the song familiar, and whether they could identify the song. During instances of recall failure, it was found that déjà entendu was associated with strong feelings of familiarity compared to instances of non-déjà entendu. Despite the featural overlap between study and test songs, though, there was no difference in the probability of reporting déjà entendu for Piano Puzzlers that were and were not studied in their

original forms. However, an interesting pattern of results emerged, such that, when examining familiarity ratings as a function of déjà entendu and study status, a significant RWI effect was found only during instances of déjà entendu. If the Piano Puzzler had been studied in its original format, and also prompted a sense of déjà entendu at test, participants were significantly more likely to find the song familiar, suggesting a link between music RWI and déjà entendu.

Although McNeely-White and Cleary's (2019) results provide a first empirical examination of the déjà entendu phenomenon, a shortcoming of the stimuli used is that the feature overlap between study and test was not scientifically manipulated. Based on the work examining déjà vu experiences, there must be high feature overlap between study and test stimuli in order to prompt a sense of déjà vu (e.g., Cleary et al., 2012). Although Piano Puzzlers offered a unique means by which to examine the potential mechanisms of déjà entendu, they were not carefully created in order to ensure a high degree of feature overlap. Thus, in a follow-up study, McNeely-White and Cleary (in progress) conducted a series of experiments in which the stimuli used were the isolated rhythm and pitch sequences created by Kostic and Cleary (2009). In their experiments, McNeely-White and Cleary presented participants with a study list of isolated song features, such as the isolated rhythm of "Mary Had a Little Lamb." At test, participants heard whole, unaltered songs, half of which contained familiarized features presented at study. Indeed, McNeely-White and Cleary found that, when participants experienced retrieval failure, they were significantly more likely to report a sense of déjà entendu for the test song if it contained unidentified features from the study phase than if it did not, suggesting that déjà entendu does indeed share similarities with déjà vu in that it is

more likely to occur when the current test stimulus shares features with information stored within memory.

Based on the pattern of results examining both déjà vu (e.g., Cleary et al., 2009; Cleary et al., 2012; Cleary et al., 2018) and déjà entendu (McNeely-White & Cleary, 2019; in progress), these retrieval-failure-based metacognitive states may offer a unique means by which to examine Berlyne's (1950; 1960) proposal that curiosity may occur when the individual is presented with a novel stimulus containing familiar elements. The RWI paradigm used to study these metacognitive experiences uses a methodology to create such situations, as participants are presented with novel test stimuli containing experimentally familiarized features. Additionally, the paradigm allows for the direct examination of what makes some situations prompt strong feelings of curiosity, such that intense feelings of familiarity due to high feature overlap may motivate the individual to engage in information-seeking behaviors in order to resolve the curiosity as a feeling-of-closeness. Although research has yet to be conducted examining how déjà vu states might encourage information-seeking behavior, there are indications throughout the literature that it does bias decision-making processes in other regards, such as conferring illusory feelings of prediction.

A growing body of work has begun to suggest that high feelings of familiarity are related to illusory feelings of prediction for future events. In their 2018 study, Cleary and Claxton examined how déjà vu might be associated with such feelings of predictions, as it is a commonly held association that people feel as if they can predict future events while experiencing déjà vu in real life. To investigate this, Cleary and Claxton used the paradigm of prior déjà vu research (e.g., Cleary et al., 2012) to present participants with

first-person navigation videos, and asked participants to indicate whether they were experiencing déjà vu and also how strongly they felt as if they could predict the direction of the next turn that the camera would take in exploring the test environment. Because of the nature of the RWI paradigm, which is used to study déjà vu, it was theoretically plausible that participants might be able to predict the next turn, as some of the test scenes adhered to the exact spatial layouts and virtual navigations presented in the contextually different study scenes. However, despite participants indicating that, during déjà vu, they strongly *felt* as if they could predict which way the scene would unfold, these feelings were illusory, with predictive ability being only at chance.

This illusory feeling of prediction bias has also been found during instances of déjà entendu. McNeely-White and Cleary (in progress) examined whether participants might feel as if they could predict the characteristics of a yet-to-be-heard note during déjà entendu. In a series of experiments, participants were presented with test songs, some of which contained experimentally familiarized features. Each test song was stopped short of a final note, though, and participants were asked whether they found the song to be familiar, whether they were experiencing déjà entendu, and also whether they felt as if they could either predict whether the next note would ascend or descend in pitch (Experiment 1) or whether the next note would play out of the left or right computer speaker (Experiment 2). During instances of retrieval failure, participants were significantly more likely to indicate that they felt as if they could predict the characteristics of the proceeding note while experiencing déjà entendu than when not. However, their accuracy was at chance for predicting the contour of the song. Further, the experiment examining whether participants felt as if they could predict the speaker

location of the next note was inherently illusory, as the song's notes adhered to a random pattern of left-right speaker locations, and was thus impossible to predict. Despite this, participants were biased to believe that they could predict the next note's speaker location.

Research examining decision-making processes during déjà vu and déjà entendu have demonstrated that participants exhibit altered patterns of decision-making behaviors while experiencing these retrieval-failure-based metacognitive states that result as a function of familiarity-detection (e.g., Cleary & Claxton 2018; Cleary et al., 2018; McNeely-White & Cleary, in progress). Based on these findings, might déjà vu and déjà entendu be similarly associated with increased feelings of curiosity and alterations in information-seeking behaviors? Currently, there is a major gap in the literature speaking to this empirical question, and thus examining how feature overlap and familiarity-detection might contribute to feelings of curiosity during instances of déjà vu and déjà entendu might also better inform our theoretical understanding of what circumstances result in feelings of curiosity and affect information-seeking behaviors.

Current Study

Within the past 15 years, there has been a recent shift in focusing on how metacognitive states associated with retrieval failure influence one's decision-making processes, such as how TOT states increase one's inclinations to engage in risk-taking behaviors (Cleary et al., 2020) or test-taking behaviors (Cleary et al., 2021). Additionally, research on déjà vu and déjà entendu experiences has demonstrated an association between these strange metacognitive states and feeling as if one can predict what will happen next, despite showing no ability to do so. Although recent

research has begun to examine how TOT states are associated with curiosity and information-seeking behaviors (e.g., Litman et al., 2005; Metcalfe et al., 2017), no such effort has been made towards examining how déjà vu states might similarly prompt curiosity as a feeling-of-closeness and information-seeking behaviors during retrieval failure. Examination of this potential relationship between déjà vu states and feelings of curiosity would be of theoretical and applied interest. From a theoretical perspective, the mechanisms behind what causes curiosity are poorly understood (e.g., Litman, 2005; 2019) yet this state has serious effects on learning and memory outcomes, with people showing increased information-seeking behaviors for items that prompt intense levels of curiosity and subsequently are more likely to remember that information weeks later (e.g., Gruber et al., 2014; Kang et al., 2009; Murphy et al., 2021; Wade & Kidd, 2019). Therefore, examining what circumstances lead to these high feelings of curiosity have clear implications from an applied perspective, as this will inform our understanding of how to maximize learning in educational environments.

The origins of this curiosity, though, are not well-understood, yet hints throughout the literature have suggested that curiosity may occur when a novel situation contains familiar elements (Berlyne 1950; 1960), which prompts the individual to engage more with the situation and discover more information, as this information fills a void in their knowledge structures. Therefore, a potential candidate for examining the relationship between feelings of familiarity and curiosity might be déjà vu and déjà entendu experiences, as they are thought to occur when the current, new test situation contains experimentally familiarized elements (e.g., Cleary et al., 2012; McNeely-White & Cleary, 2019; in progress). The current study aimed to examine the relationship between

familiarity-detection during retrieval failure, specifically déjà vu and déjà entendu experiences, and curiosity as a feeling-of-closeness.

Another overarching goal of the present study was to examine whether these metacognitive sensations of familiarity-detection during retrieval failure might be associated with a desire to seek out the potential source of the sensations, and if so, whether these sensations are associated with alterations in participants' decision-making processes. One approach to examining whether such alterations might occur is to provide participants with limited opportunities to receive information concerning the current situation, following from Metcalfe et al. (2017). In their study examining TOTs and curiosity, Metcalfe et al. provided participants with the opportunity to see the correct answer for up to 10% of the questions, with results demonstrating that participants were significantly more likely to use their opportunities during TOT states than non-TOT states. Based on this, the current experiments incorporated a similar limited-opportunity mechanism, by which participants were given limited opportunities throughout the test phases to receive information concerning the current visual (Experiment 1) or auditory (Experiment 2) stimulus, which would in turn resolve their perceived knowledge gap.

Another approach to investigating potential changes in information-seeking behaviors is to examine the number of commission errors made on trials associated with déjà vu or déjà entendu, in addition to participants' reaction time on recall prompts (see Schwartz et al., 2000 and Schwartz et al., 2000 for the association between TOT states, commission errors, and increased reaction time). If participants are indeed curious about the source of the familiarity-signals and demonstrate changes in information-seeking behaviors, one manifestation of this might be increased

commission errors during déjà vu and déjà entendu states. As opposed to an omission error, in which participants do not provide any information when prompted to recall relevant details concerning the current test stimulus (i.e., they leave the prompt blank), a commission error is an incorrect answer (e.g., the target song name was “London Bridge,” but the participant typed in “Mary Had a Little Lamb”). If there is an association between metacognitive sensations of familiarity-detection during retrieval failure and alterations in information-seeking behaviors in order to resolve perceived knowledge gaps, perhaps participants attempting to resolve these gaps might make more commission errors, which might reflect increased internal searchers of memory for relevant information. Further, if participants are indeed engaging in more information-seeking behaviors during metacognitive states, specifically déjà vu and déjà entendu, and are attempting to internally search memory for relevant information concerning the source of the sensation, might this manifest as more time spent on recall prompts, as participants might be spending more time internally searching for the answer?

To investigate these research questions, Experiment 1 examined how déjà vu experiences might be associated with increased curiosity as a feeling-of-closeness to discover missing information. As seen in work done by Cleary and Claxton (2018), déjà vu is associated with changes in decision-making processes, specifically the feeling of being able to predict the outcome of an unfolding scene despite showing no ability to do so. There may be a similar change in decision-making processes during déjà vu in that participants might feel an increased inclination to engage in information-seeking behaviors in order to resolve intense moments of curiosity as a feeling-of-closeness. Therefore, in Experiment 1, I hypothesized that instances in which a participant reported

a sense of déjà vu would also be accompanied by stronger feelings of curiosity in addition to an increased inclination to expend limited resources to discover the corresponding study scene. These findings will not only establish a link between feelings of curiosity for visual stimuli (a needed area of study, as called for by Litman, 2009) and déjà vu experiences, but will also inform us of whether curiosity is indeed more likely to occur due to feature-matching processes and familiarity-detection. If Berlyne's (1950; 1960) proposal is correct, then novel test scenes that do indeed contain experimentally familiarized elements from study will be more likely to prompt intense levels of curiosity during retrieval failure, as the participant will feel the intense need to resolve the temporary curiosity as a feeling-of-closeness (Litman & Jimerson, 2005). I should additionally find that participants will be more likely to expend limited resources in order to discover the study scene whose features are embedded within the novel test scene that prompted the feeling of curiosity, as a similar behavior is demonstrated during TOT states (Metcalfe et al., 2017).

Experiment 2 served as an attempt to examine how feelings of curiosity for auditory stimuli might occur during déjà entendu. As previously discussed, Litman (2009) called for additional research examining the mechanisms and consequences of curiosity for auditory information; therefore, Experiment 2 used auditory stimuli to examine the relationships between familiarity-detection, déjà entendu, and feelings of curiosity, and whether participants will show an increased tendency to engage in information-seeking behaviors in order to resolve the temporary feelings of knowledge gaps during retrieval failure. Further, the amount of feature overlap between the features of the current test cues and those stored within memory traces was examined,

as prior work has shown that incrementally increasing the amount of feature overlap between study and test stimuli subsequently increases the perceived level of familiarity with the test cue, theoretically because of the feature-matching processes that are thought to occur to produce the familiarity signal (e.g., Huebert et al., 2021; McNeely-White et al., 2021; Ryals & Cleary, 2012). As there is more information to match onto, the test cue will be perceived as being more familiar. Might a similar phenomenon occur with feelings of curiosity, such that participants feel even more curious to discover the missing information when the current test cue more strongly resembles something stored within memory, as this increases the magnitude of the relative novelty?

Chapter 2 – Experiment 1 (Déjà vu and Feelings of Curiosity)

Based on prior research examining feelings of curiosity and information-seeking behaviors, there is reason to hypothesize that participants will feel increased levels of curiosity during the retrieval-failure-based metacognitive experience of déjà vu (e.g., Metcalfe et al., 2017; Schwartz & Cleary, 2016). Therefore, in Experiment 1, I examined whether participants experiencing déjà vu also report higher feelings of curiosity. This was be done by using the RWI paradigm and prior methodologies used in déjà vu experiments (e.g., Cleary et al., 2012; Cleary et al., 2018), such that participants were presented with first-person virtual tours during the test phase, with half of them containing familiarized spatial features presented during the study phase. I hypothesized that the retention of familiarized spatial features in test scenes would increase participants' perceived feelings of curiosity during retrieval failure, providing support Berlyne's (1950; 1960) relative novelty proposal, compared to when the test scenes did not contain familiarized spatial features. Further, I hypothesized that participants would be more likely to use limited opportunities to see the corresponding study scene that the current test stimulus mapped onto when the current test stimulus contained experimentally familiarized features. This was done by allowing participants to indicate on up to 20% of the trials that they wanted to see the study scene that resembled the current test stimulus. Additionally, I hypothesized that there would be a relationship between déjà vu and high levels of curiosity, such that participants would provide higher curiosity ratings for the current test scene while also experiencing déjà vu compared to when they were not experiencing déjà vu. Finally, if there is a

relationship between déjà vu and feelings of curiosity, I hypothesized that participants would be more likely to indicate that they want to use limited resources to see the corresponding study scene when they are in déjà vu compared to when they are in non-déjà vu.

Method

Participants

Participants consisted of 72 undergraduates from Colorado State University. Previous research examining the déjà vu experience and biased decision-making processes (Cleary & Claxton, 2018) found large effect sizes (d_z = 1.33 in Experiment 2 and d_z = 1.54). However, as no research has yet been conducted on feelings of curiosity for visual stimuli, let alone during déjà vu experiences, a conservative power analysis was used, with power set to .90, an α of .05, and a medium effect size (d = .50). Using G*Power (Faul et al., 2007), which indicated that the necessary sample size to demonstrate such an effect would require 44 participants. All participants received course credit for participating in this experiment.

Materials

Stimuli consisted of the 64 study-test pairs of configurally similar scenes that have been used in prior research examining the déjà vu phenomenon (Cleary et al., 2012; Cleary & Claxton, 2018). These scenes were designed such that the test scenes contained the exact spatial layout of elements from the contextually unique study scenes (see Figure 3). In addition to using the 64 study-test pairs, audio files created by Cleary et al., 2012 were used during the study phase to indicate what the name of the

scene is (e.g., “This is a junkyard. Junkyard” was played while participants viewed the junkyard study scene).

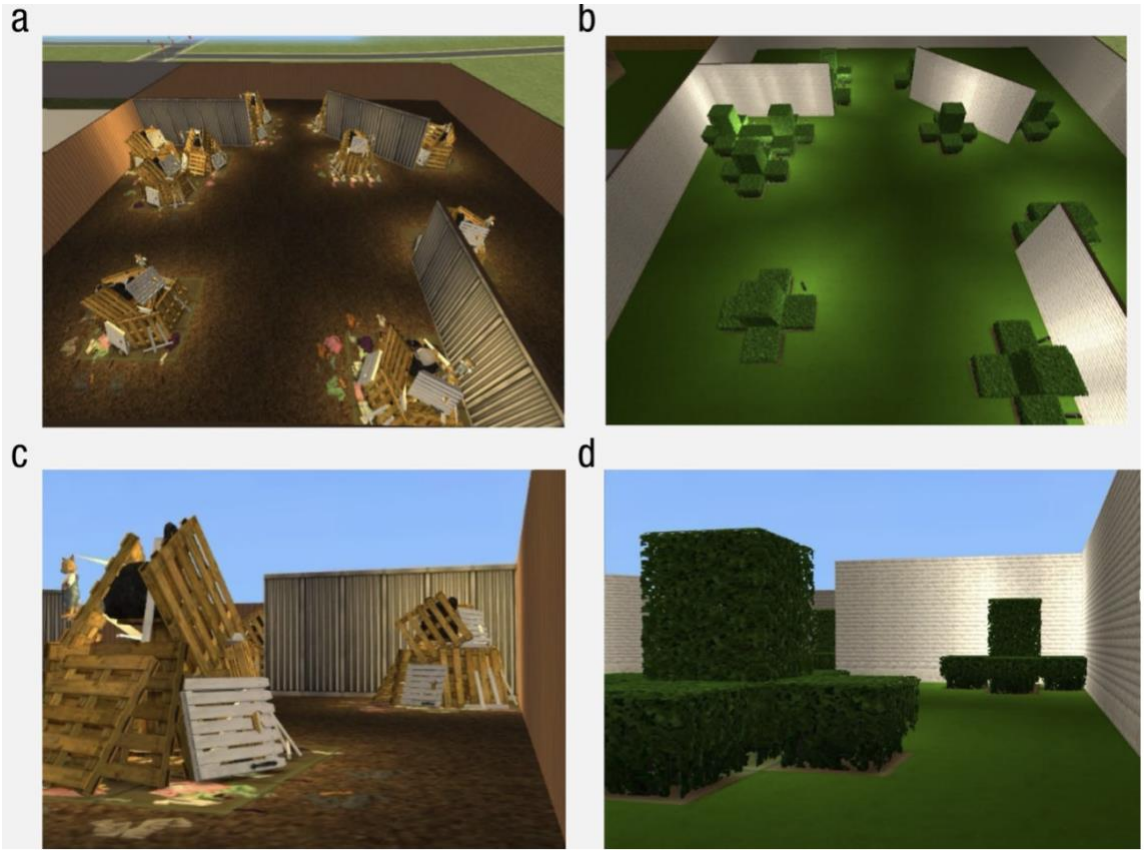

Figure 3. Taken from Cleary & Claxton (2018). Panels a and b represent a bird’s-eye view of the junkyard study scene, which spatially corresponds to the hedge garden test scene. Panels c and d represent the participant’s first-person perspective of these study-test pairs.

Two counterbalanced versions of the experiment were created such that each test scene fell into the studied and unstudied condition. For example, in the first counterbalanced version, the garden hedges test scene spatially corresponded to a scene presented at study (the junkyard study scene), while in the second counterbalanced version, the garden hedges test scene did *not* spatially correspond to a scene presented during study.

Procedure

Participants were randomly assigned to one of the two counterbalanced versions of the experiment. The 64 study-test pairs were divided into two study-test blocks. Each study list consisted of 16 virtual tour videos accompanied by an audio file indicating what the name of the scene is. After each study list came the test list, which consisted of 32 contextually unique virtual tour videos. Half of these test scenes corresponded to spatially similar studied scenes while the other half did not. All video tours were randomly presented.

Prior to beginning the experiment, participants were presented with instructions explaining their tasks throughout the experiment (see Appendix A). After reading through the instructions, participants began the first study block. Their only task was to watch each virtual tour video carefully and try to remember the name. Once the study list of 16 virtual tour videos was complete, participants were then given specific instructions before completing the first test block. They were instructed that they would now view a new list of virtual tour videos, none of which would have been seen during the study phase. However, they were told that some of the virtual tours might remind them of a scene presented during the study phase. After each video, they would be asked a series of questions concerning the scene. First, they would be asked to indicate whether they are experiencing déjà vu, which was defined as “the feeling of having been someplace or done something before, without being able to pinpoint why, and despite knowing that the current situation is new.” Next, they would be asked to indicate how familiar the current test scenes feels on a scale from zero (*Not at all familiar*) to 10 (*Extremely familiar*). Next, they would be told that they should indicate how curious they

feel about the current test scene on a scale from zero (*Not at all curious*) to 10 (*Extremely curious*). They would then be prompted to indicate whether they can think of a scene from the study phase that reminds them of the current test scene, and if so, to type in the name of that scene. Finally, they would then be told that they will have limited opportunities to discover information concerning the current test scene, namely to which study scene (if any) it corresponds. However, they would only be able to see this information on approximately 20% of the trials, which was six opportunities per test block.[1] If they indicated that they did want to see the corresponding study scene, then an image of the study scene would be displayed along with the name of the scene (note that if the current test scene did not correspond to a scene presented at study, then the text "This scene does NOT correspond to a scene presented at study" was displayed).

After receiving the test instructions, participants then began the first test block, consisting of 32 randomly ordered virtual tour videos, half of which corresponded to scenes presented during the study phase. After watching each virtual tour video, participants were asked a series of questions concerning the test scene, with each question appearing one at a time on the screen. The first question asked participants to indicate whether they were experiencing déjà vu (Y=Yes, N=No). The second question next asked them to indicate how familiar the test scene felt on a scale of zero to 10 (*0 = Not at all familiar, 10 = Extremely familiar*). The third prompt asked participants to indicate how curious they felt about the test scene on a scale of zero to 10 (*0 = Not at*

[1] Although Metcalfe et al. (2017) allowed their participants in their TOT experiment to view the target answer on 10% of the trials, there is little precedent in the literature concerning how scarce these limited resources must be. Therefore, in the current experiment, the number of limited opportunities to receive relevant information concerning the current test stimulus (20%) was determined in an attempt to create a sense of scarcity so that participants would carefully allocate these opportunities.

all curious, 10 = Extremely curious). A fourth question asked participants whether they could think of a scene from the study phase that reminded them of the current test scene, and if so, to type in the name of that scene. Finally, participants were then asked to indicate whether they would like to use their limited opportunities to see information about the current test scene (Y=Yes, N=No). For this question, the remaining number of opportunities was displayed at the bottom of the screen so that the participant was reminded of how many opportunities they had left (e.g., the text "3 opportunities remaining" was displayed at the bottom left-hand side of the screen) along with the trial number (e.g., the text "Trial 4/16" was displayed at the bottom right-hand side of the screen). If the participant indicated that they did indeed want to see information concerning the test scene, then a still-image of the configurally similar study scene was displayed in the middle of the screen along with the name of the study scene. However, if the test scene did not correspond to a configurally similar study scene, then the text "This scene does NOT correspond to a scene presented at study" was displayed in the middle of the screen. To prevent participants from simply saying "No, do not use limited opportunities" as a means of completing the experiment faster, the text "Please wait…" appeared if the participant indicated "No" on this prompt, which lasted the same duration (3 seconds) as if they had indicated "Yes."

After completing the first study-test block, participants then proceeded to the second study-test block. Additionally, the limited-opportunities counter was reset, such that participants were able to use their limited resources to see information pertaining to the current test scene on up to 6 of the trials in the second test phase.

Results

In both Experiments 1 and 2, data were analyzed using null hypothesis significant testing (NHST) and Bayesian methods of analysis. In addition to reporting *p*-values and standard effect sizes produced from NHST, Bayes factors (*BF*s) are also reported; these were computed using JASP with the JZS prior, as it requires the fewest prior assumptions about the range of true effect sizes (Rouder et al., 2009). The recommendations proposed by Wagenmakers et al. (2007) concerning strength of evidence provided by Bayes factors were used, such **I** considered a Bayes factor to provide either anecdotal evidence (*BF* = 1-3), substantial evidence (*BF* = 3-10), strong evidence (*BF* = 10-30), very strong evidence (*BF* = 30-10), or extreme evidence (*BF* > 100) in favor of one hypothesis, either the null or alternative, over the other. A Bayes factor of 1 was considered to provide no evidence for either the null or alternative hypothesis. In the results sections reported below, BF_{10} was reported when arguing in favor of the alternative hypothesis while BF_{01}, which is the reciprocal of BF_{10}, was reported when arguing for the null hypothesis.

Identification Rates

Trials were hand-labeled as being either an instance of identification success (e.g., the participant typed in the target scene's name), partial identification success (e.g., the participant typed in "wedding" for the target scene Arbor), or identification failure (e.g., the participant typed in the incorrect target name or left the prompt blank). The trials labeled as identification failure were further subdivided as being an instance of a commission error (the participant typed in the incorrect information) or an omission

error (the participant did not type in anything or typed “Don’t know,” “Can’t remember,” etc.).

On average, when presented with a novel test scene that did indeed spatially resemble a scene presented during the study phase, participants correctly provided the name of the corresponding study scene on 26% (*SD* = .19) of the trials. As the primary analyses of interest to the current study focus on instances of retrieval failure, this left an average of 87% of test trials for analyses (note that trials containing novel test scenes that did *not* spatially resemble studied scenes could not possibly be labeled as having an instance of identification success, as there was nothing in the study phase relevant to the test scenes, and were therefore included in the analyses reported below).

In examining the trials on which retrieval failed, the probabilities of participants making commission errors were computed first by taking the total number of trials on which a commission error occurred for a spatially similar scene and dividing by the total number of trials on which a commission *or* omission error occurred for a spatially similar scene. The same method was done to compute the probabilities of participants making a commission error for spatially dissimilar scenes. Participants were marginally more likely to produce a commission error on trials corresponding to spatially similar study scenes (*M* = .18, *SD* = .15), compared to trials that did not correspond to spatially similar study scenes (*M* = .16, *SD* = .13), $t(71) = 1.88$, *SE* = .01, $p = .06$, $d = .14$, $BF_{10} = .69$. When considering only the participants who actually made commission errors, a similar pattern emerged, such that participants were marginally more likely to make a commission error on trials that corresponded to spatially similar study scenes (*M* = .20,

SD = .15) compared to those that did not (*M* = .17, *SD* = .13), *t*(66) = 1.89, *SE* = .01, *p* = .06, *d* = .16, BF_{10} = .71.

Recognition without Identification

Focusing on trials in which participants failed to identify the corresponding spatially similar scene presented during the study phase, I now turn to participants' familiarity ratings. When participants were presented with a test scene that did correspond to an unidentified spatially similar study scene, they provided significantly higher familiarity ratings (*M* = 3.11, *SD* = 1.54) compared to when they were presented with a test scene that did not correspond to a spatially similar study scene (*M* = 2.72, *SD* = 1.44), *t*(71) = 4.81, *SE* = .08, *p* < .001, *d* = .26, BF_{10} = 2116.54 (see Figure 4 below). This pattern of results is similar to that of prior research, such that participants, although failing to identify the prior experience, are able to discriminate between items that were and were not studied based on subjective familiarity.

Déjà vu Reports and Memory Search Time

Overall, during retrieval failure, participants reported experiencing déjà vu on an average of 27% (*SD* = .15) of the trials. As discussed above, one manifestation of an increased inclination to engage in information-seeking behaviors during retrieval-failure-based metacognitive states might be increased time spent attempting to produce the target information. In looking at the amount of time spent on the Recall prompt for trials on which identification failed, there was a significant difference as a function of déjà vu state report. On trials associated with a déjà vu report, participants spent a significantly longer amount of time on the Recall prompt (*M* = 3419.73 ms, *SD* = 1894.73 ms), compared to trials associated with a non-déjà vu report (*M* = 1413.51 ms, *SD* = 1165.29

ms), $t(69) = 10.82$, $SE = 185.42$, $p < .001$, $d = 1.19$, $BF_{10} = 3.80 \times 10^{13}$. These findings are novel, as they suggest that participants experiencing déjà vu spend longer amounts of time searching for information in order to explain the strange metacognitive sensation, specifically in the form of internally searching memory for corresponding information.

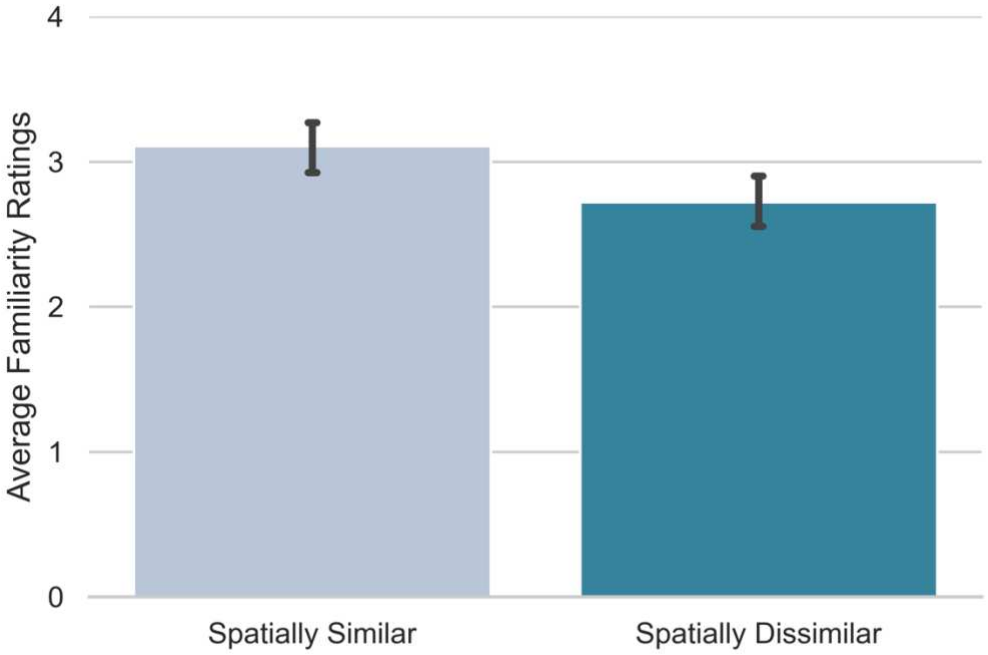

Figure 4. Average familiarity ratings provided during retrieval failure as a function of spatial similarity. Participants provided significantly higher familiarity ratings for test scenes that did correspond to a spatially similar study scene compared to those that did not. Error bars represent standard error of the mean.

Probability of a Déjà vu Report Given Study Status

In replication of prior research examining the déjà vu phenomenon, it was found that, while experiencing identification failure, participants were significantly more likely to experience a sense of déjà vu when the test scene corresponded to a spatially similar study scene ($M = .31$, $SD = .17$) compared to when it did not ($M = .26$, $SD = .15$), $t(69) = 3.15$, $SE = .02$, $p = .002$, $d = .31$, $BF_{10} = 11.60$ (see Figure 5 below). In examining

reaction time data, when participants reported experiencing déjà vu for the test scene, no difference was found in the amount of time spent on the Recall prompt when the test scene did (M = 3271.26 ms, SD = 2075.57 ms) versus did not (M = 3672.24 ms, SD = 2231.23 ms) correspond to a spatially similar study scene, $t(67) = -1.49$, SE = 269.24, p = .14, BF_{01} = 2.63.

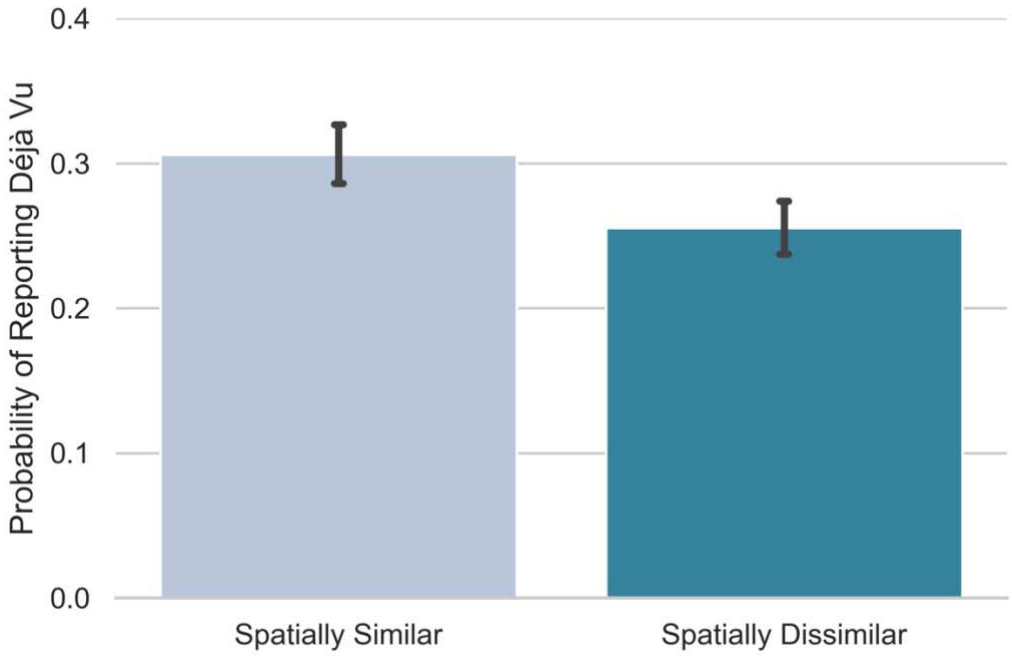

Figure 5. Probability of reporting déjà vu during retrieval failure as a function of spatial similarity. Participants were more likely to report experiencing a sense of déjà vu if the test scene did correspond to a spatially similar study scene compared to if it did not. Error bars represent standard error of the mean.

Déjà vu and Identification Errors

Although never previously examined in déjà vu research, there is emerging research suggesting that retrieval-failure-based metacognitive states encourage participants to make more attempts at retrieval relevant details from memory (Huebert et al., under review). Specifically, when participants are in reported TOT states for the answer to a general knowledge question, they provided significantly more partial

recollection responses compared to when they are in a non-TOT, despite showing no increase in accuracy among these attempts. Based on this, and the current experiment's object to determine whether déjà vu experiences are associated with increased information-seeking behaviors, I examined the relationship between déjà vu states and the types of errors made when attempting to recall the target information, as increased desires to engage in information-seeking behaviors during déjà vu might manifest as participants making more commission errors in an attempt to resolve the perceived knowledge gap.

When comparing the probability of participants making a commission versus omission error while experiencing déjà vu, there was no significant difference. Participants were not more likely to make a commission error (M = .46, SD = .30) than an omission error (M = .54, SD = .30) while experiencing déjà vu, $t(69) = 1.10$, SE = .07, p = .28, BF_{01} = 4.29. However, when considering the probability of participants making a commission versus omission error while experiencing non-déjà vu, a significant difference did emerge, such that non-déjà vu was significantly more likely to be accompanied by omission errors (M = .95, SD = .07) than commission errors (M = .05, SD = .07), $t(69) = 52.66$, SE = .02, $p < .001$, $BF_{10} = 5.60 \times 10^{53}$. This perhaps suggests that the presence of a déjà vu state might prompt one to internally search memory more so than when they are not experiencing such a sensation.

To further investigate the occurrence of commission errors as a function of déjà vu reports, when only considering trials on which commission errors were made, participants were significantly more likely to report a sense of déjà vu (M = .46, SD = .30) compared to non-déjà vu (M = .05, SD = .07), $t(69) = 11.66$, SE = .04, $p < .001$, d =

1.77, BF_{10} = 9.86 x 10^{14}. When limiting this analysis to include only participants who actually did make commission errors (i.e., excluding those who only made omission errors during retrieval failure), the magnitude of the effect became larger. When participants reported a sense of déjà vu, they were significantly more likely to make a commission error (M = .49, SD = .27) compared to when they did not report a sense of déjà vu (M = .05, SD = .07), $t(65)$ = 12.42, SE = .04, p < .001, d = 2.15, BF_{10} = 6.58 x 10^{15} (see Figure 6 below).

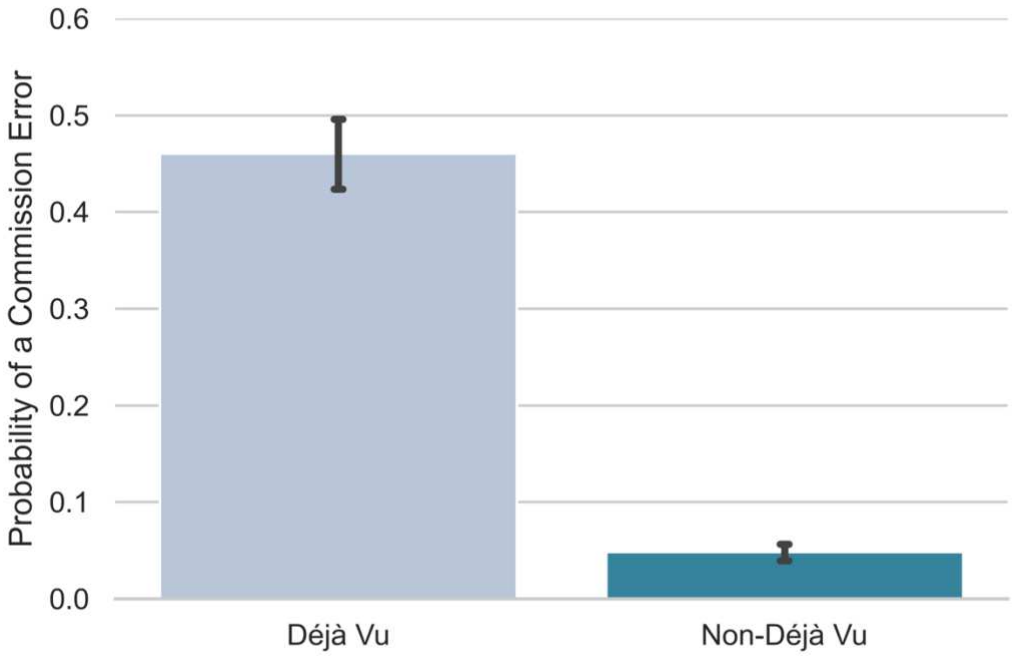

Figure 6. Probability of a commission error as a function of déjà vu reports. The occurrence of a commission error was significantly more likely during reported déjà vu states than non-déjà vu states. Error bars represent standard error of the mean.

Although the directionality of this relationship is not yet known, the pattern of results suggests that participants experiencing déjà vu may feel an internal drive to search for information concerning the current situation, as seen by the increased association with

making a commission error. Additional research examining the time course of these processes is needed, though, to make stronger claims concerning this association.

Familiarity Ratings Accompanying Déjà vu Reports

When participants reported experiencing a sense of déjà vu during retrieval failure, they provided significantly higher familiarity ratings for the test scene (M = 6.26, SD = 1.50) than when they were in a non-déjà vu state (M = 1.66, SD = 1.32), $t(69)$ = 20.96, SE = .22, $p < .001$, d = 3.25, BF_{10} = 2.30 x 10^{28}. These findings are similar to prior research (e.g., Cleary et al., 2018) demonstrating the association between déjà vu and feelings of familiarity, such that participants subjectively experience intense levels of familiarity for the current environment when they are concurrently experiencing déjà vu.

Feelings of Curiosity

Curiosity Ratings as a Function of Study Status

Turning now to the data of primary interest, the patterns of results pertaining to feelings of curiosity are reported. To assess whether Berlyne's (1950; 1960) relative novelty proposal might be supported, a paired-samples t-test was conducted examining the curiosity ratings provided for test scenes that either did or did not contain experimentally familiarized features from the study phase, regardless of identification status. Indeed, in support of Berlyne's proposal that curiosity might emerge when an individual is presented with a novel stimulus containing familiarized elements, participants provided significantly higher curiosity ratings to test scenes that contained familiarized features from study (M = 3.68, SD = 2.11) compared to those that did not

(M = 3.26, SD = 2.29), t(71) = 3.60, SE = .12, p < .001, d = .19, BF_{10} = 40.73 (see Figure 7 below).

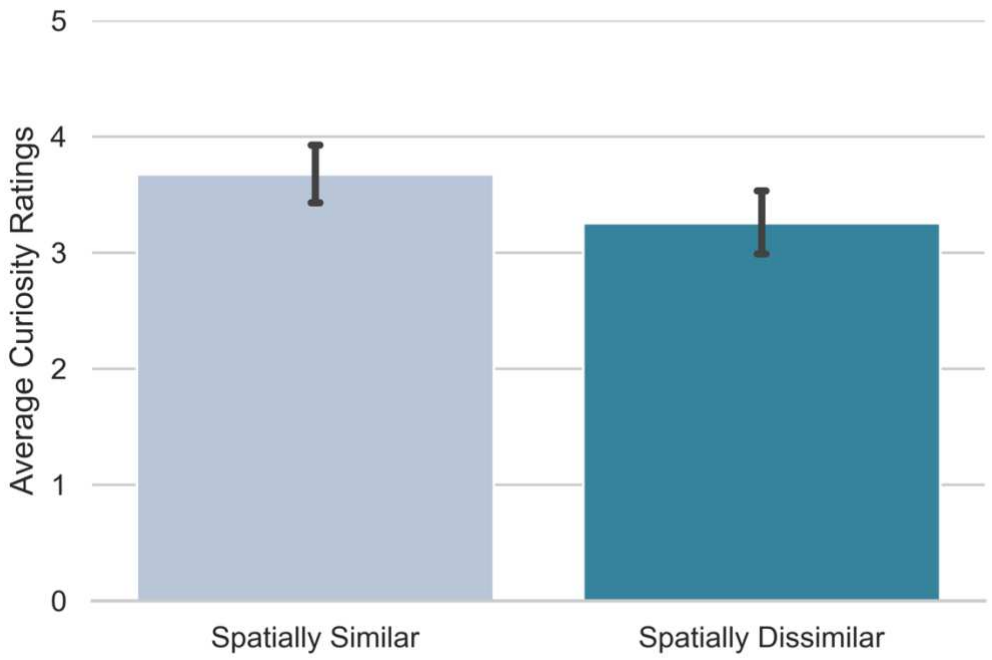

Figure 7. Average curiosity ratings provided as a function of spatial similarity, regardless of identification status. The average curiosity ratings provided for scenes that did spatially resemble a study scene were significantly higher than the scenes that did not spatially resemble a study scene. Error bars represent standard error of the mean.

However, when examining only trials on which retrieval failed, no significant difference between the curiosity ratings was found. Test items containing familiarized features did not receive higher curiosity ratings (M = 3.37, SD = 2.20) compared to those that did not contain experimentally familiarized features (M = 3.26, SD = 2.30), t(71) = 1.45, SE = .08, p = .15, BF_{01} = 2.86 (see Figure 8 below).

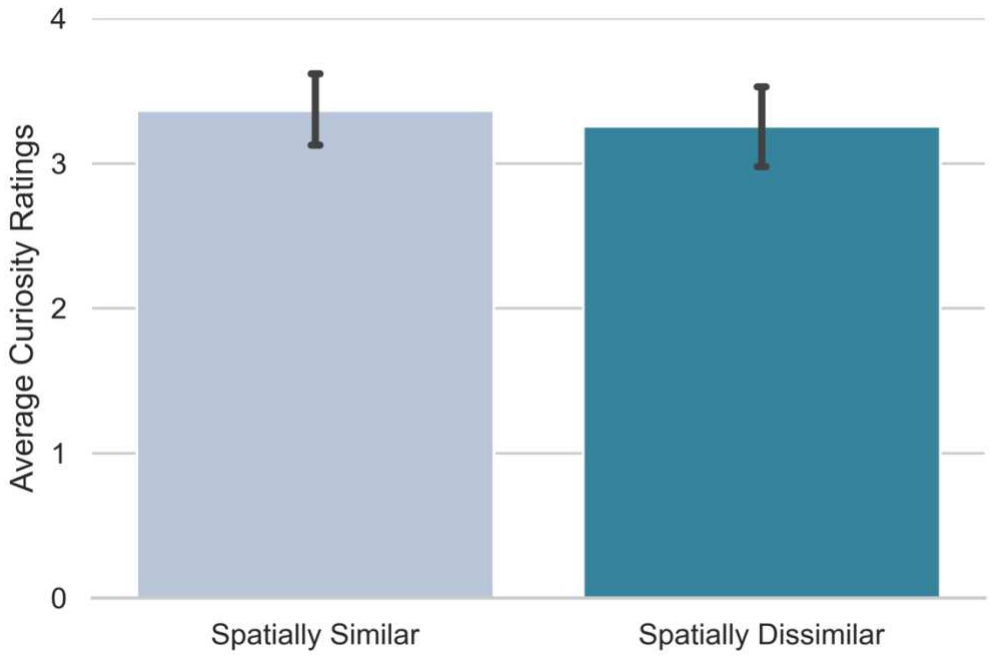

Figure 8. Average curiosity ratings provided during retrieval failure as a function of spatial similarity. There was no difference in the average curiosity ratings provided for scenes that did versus did not spatially resemble a study scene. Error bars represent standard error of the mean.

Curiosity Ratings as a Function of Identification Errors

As previously reported, participants were more likely to make commission errors while experiencing déjà vu than non-déjà vu suggesting, that déjà vu states are associated with increased internal search processes, as participants were motivated to attempt to produce more relevant information, perhaps due to the strange metacognitive signal. To assess whether commission errors, which can be viewed as failed attempts to retrieve relevant information, are associated with increased curiosity, a paired-samples *t*-test was conducted, examining whether participants provided higher curiosity ratings for trials on which they made commission errors as opposed to omission errors. Indeed, participants provided significantly higher curiosity ratings for trials on which they made commission errors ($M = 4.84$, $SD = 2.43$) compared to trials on which they made

omission errors (M = 3.34, SD = 2.25), t(66) = 6.03, SE = .25, p < .001, d = .64, BF_{10} = 1.62 x 10^5 (see Figure 9 below).

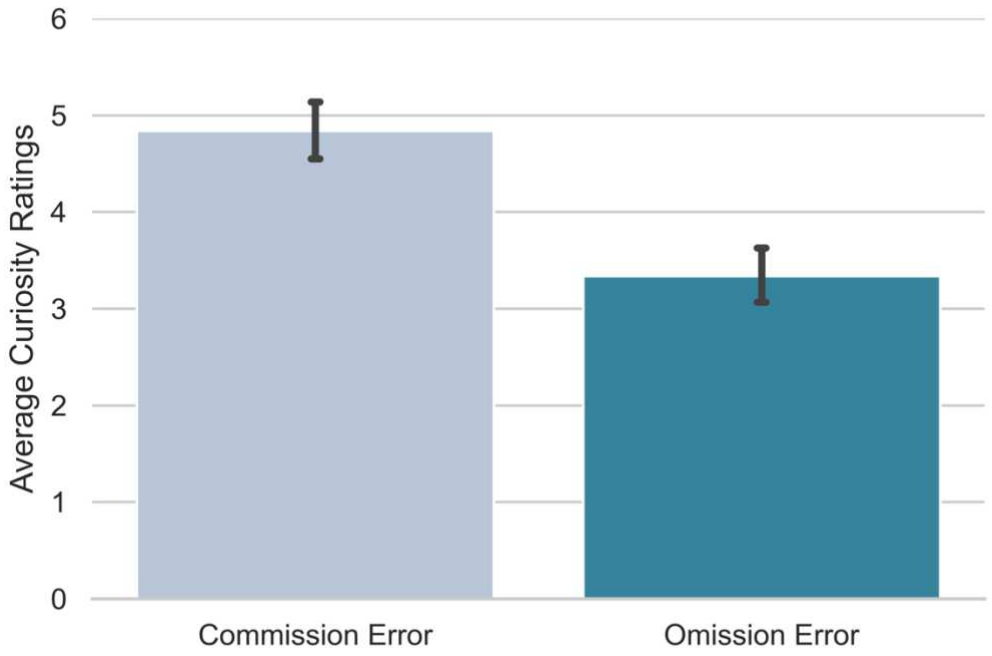

Figure 9. Average curiosity ratings provided as a function of identification error-type (commission or omission). Participants provided significantly higher curiosity ratings on trials associated with a commission error than an omission error. Error bars represent standard error of the mean.

Curiosity Ratings as a Function of Déjà vu States

To assess the relationship between feelings of curiosity and déjà vu states, a paired-samples *t*-test was conducted examining the curiosity ratings provided during retrieval failure as a function of reported déjà vu state. When participants reported that they were experiencing déjà vu for the test scene, they provided significantly higher curiosity ratings (M = 4.91, SD = 2.32) compared to when they were in a non-déjà vu state (M = 2.87, SD = 2.33), t(69) = 7.80, SE = .26, p < .001, d = .89, BF_{10} = 2.07 x 10^8 (see Figure 10 below). Additionally, when examining the reaction time data measuring how long participants remained on the curiosity rating prompt, a significant difference

was also found, such that participants spent more time on the curiosity prompt during reported déjà vu states (M = 2002.51 ms, SD = 1196.20 ms) compared to when they were in reported non-déjà vu states (M = 1697.54 ms, SD = 1203.22 ms), $t(69) = 4.45$, SE = 68.60, $p < .001$, $d = .25$, BF_{10} =584.32. These patterns of results suggest that déjà vu is associated with increased feelings of curiosity and may be prompting longer searches of memory, as suggested by the longer reaction time on the curiosity prompt.

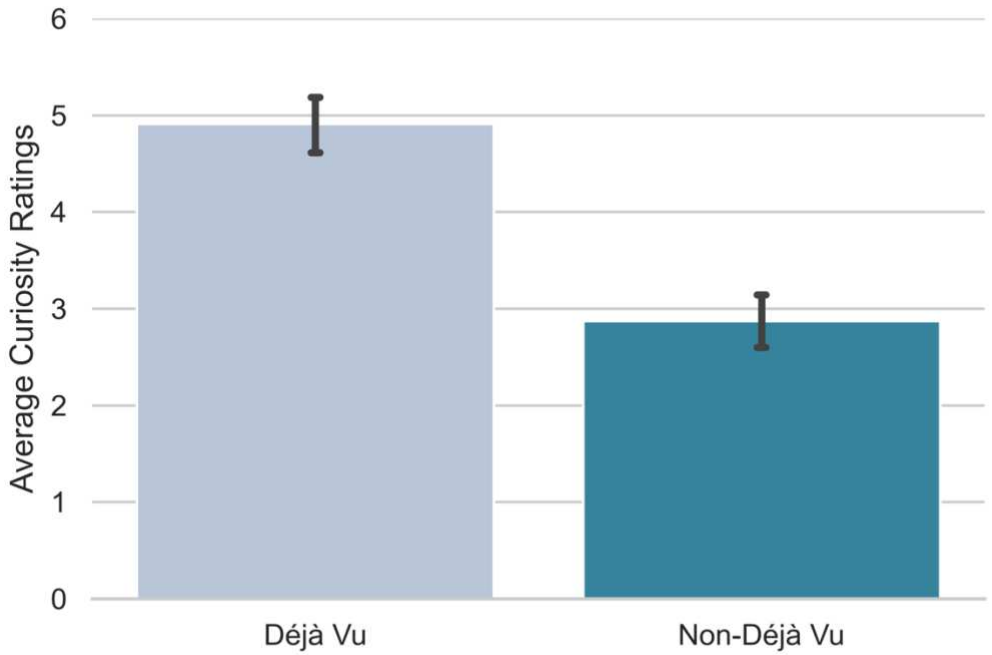

Figure 10. Average curiosity ratings provided during retrieval failure as a function of reported déjà vu state. Participants provided significantly higher curiosity ratings during reported déjà vu states than non-déjà vu states. Error bars represent standard error of the mean.

Prior research has demonstrated that the presence of a metacognitive state, specifically déjà entendu (McNeely-White & Cleary, 2019), is necessary in order for the RWI effect to emerge, such that participants can only discriminate between items that do and do not contain experimentally familiarized features while experiencing déjà entendu. Given that the current patterns of results suggest that participants do not

provide significantly different curiosity ratings as a function of study status, but *do* as a function of déjà vu state, might a similar phenomenon be present? Specifically, might curiosity ratings differ between test scenes that do and do not contain experimentally familiarized elements but only when participants concurrently report experiencing déjà vu? To investigate this, a 2 Study Status (Studied, Unstudied) x 2 Déjà vu Status (Déjà vu, non-Déjà vu) repeated-measures ANOVA was conducted, revealing a significant main effect of déjà vu status, $F(1, 67) = 57.05$, $MSE = 4.77$, $p < .001$, $n_p^2 = .46$, $BF_{10} = 1.17 \times 10^{21}$ (see Figure 11 below). However, there was no main effect of study status, $F(1, 67) = .03$, $MSE = .62$, $p = .88$, $n_p^2 = .00$, $BF_{01} = 7.38$, or a significant interaction, $F(1, 67) = .08$, $MSE = .68$, $p = .78$, $n_p^2 = .00$, $BF_{01} = 5.41$.

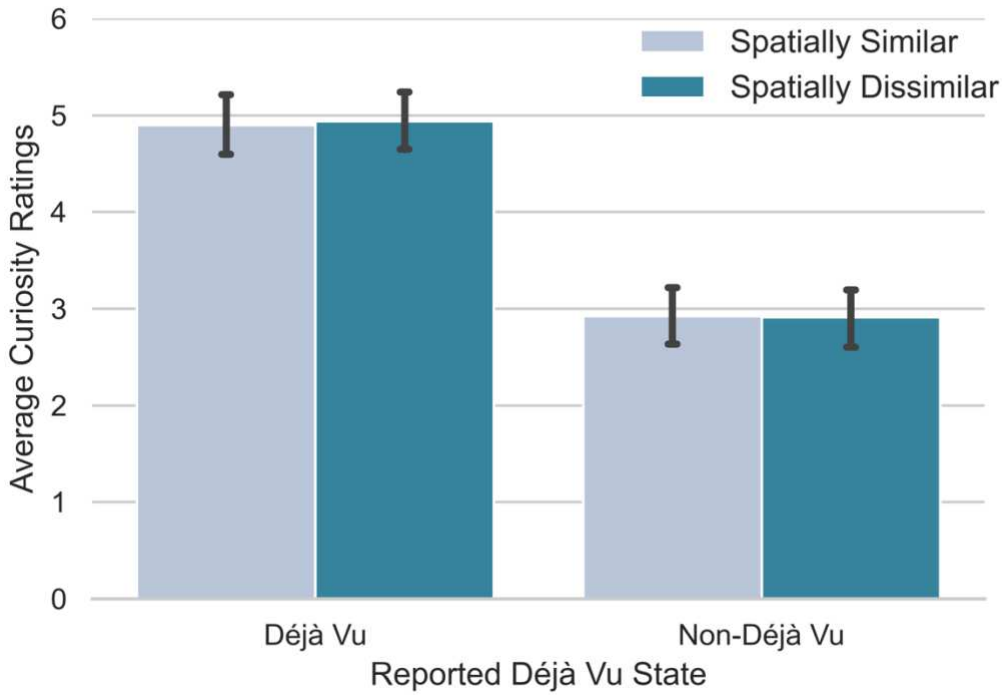

Figure 11. Average feeling of curiosity ratings as a function of study status and reported déjà vu states. Curiosity ratings were significantly higher among déjà vu reports compared to non-déjà vu reports; however, there was no difference as a function of study status. Error bars represent the standard error of the mean.

Feelings of Familiarity and Feelings of Curiosity

I now turn to the results pertaining to the relationship between participants' feelings of familiarity and feelings of curiosity during retrieval failure. The correlation between familiarity and curiosity ratings was computed for each participant. These correlations were then analyzed using a one-sample *t*-test with a test value of zero. Overall, participants' familiarity ratings (M = 3.44, SD = 2.17) and curiosity ratings (M = 2.96, SD = 1.39) were significantly correlated, $t(68) = 11.00$, $p < .001$, $d = 1.32$, $BF_{10} = 6.29 \times 10^{13}$, with an average correlation of .45 (SD = .34). This pattern of results suggests that there is a positive correlation between feelings of familiarity and curiosity during retrieval failure, with feelings of curiosity increasing as the participant subjectively feels more intense levels of familiarity with the test cue (note, though, that the directionality is not yet known).

To examine whether the experimental manipulation of study status had the same effect on curiosity ratings as it did on familiarity ratings, a 2 Rating-Type (Curiosity Rating, Familiarity Rating) x 2 Study Status (Studied, Unstudied) repeated-measures ANOVA was conducted. Although not originally hypothesized, a significant interaction emerged, $F(1, 71) = 9.44$, $MSE = .15$, $p = .003$, $n_p^2 = .12$, $BF_{10} = 0.32$ (although note that the Bayes factor does not provide conclusive support). When examining the influence of Study Status, a significant main effect was found, $F(1, 71) = 15.02$, $MSE = .30$ $p < .001$, $n_p^2 = .18$, $BF_{10} = .67$ (see Figure *12*). However, the Bayes factor did not provide conclusive support for this main effect. As previously reported above, when comparing the familiarity ratings provided during retrieval failure, participants provided significantly higher ratings when the test scene contained experimentally familiarized

features presented at study (M = 3.11, SD = 1.54) compared to when it did not (M = 2.72, SD = 1.44), $t(71) = 4.81$, SE = .08, $p < .001$, d = .26, BF_{10} = 2116.54. However, there was no difference in curiosity ratings provided for test scenes that did contain experimentally familiarized features (M = 3.37, SD = 2.20) versus did not (M = 3.26, SD = 2.30), $t(71) = 1.45$, SE = .08, p = .15, BF_{01} = 2.86.

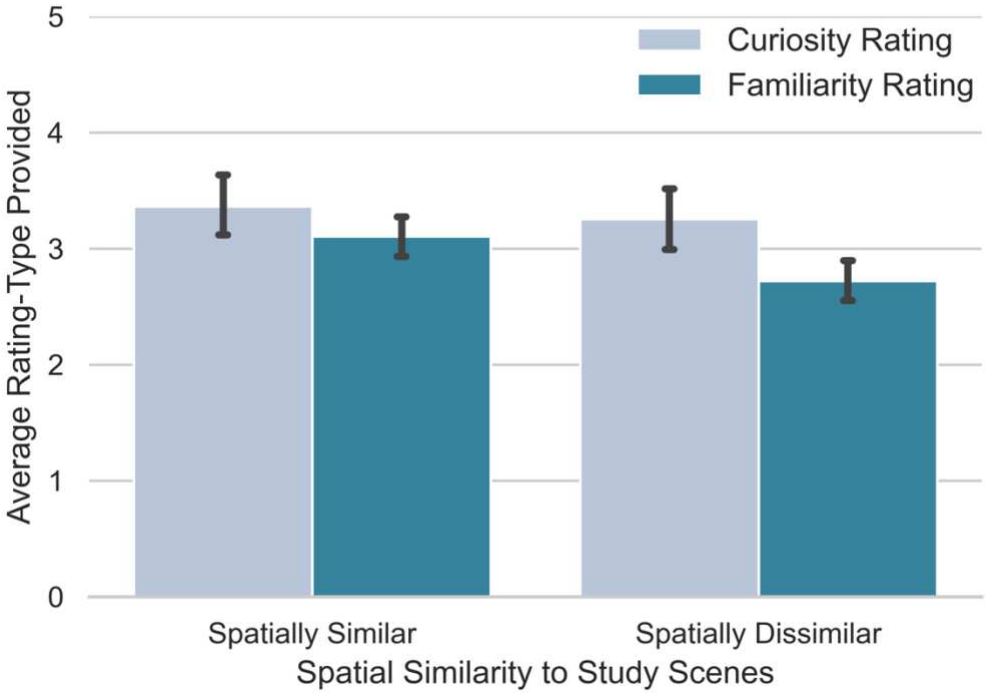

Figure 12. Average ratings, either familiarity or curiosity, provided for familiarity and curiosity judgements as a function of spatial similarity to scenes presented during the study phase. Error bars represent the standard error of the mean.

As hypothesized, there was not a significant main effect of Rating-Type, $F(1, 71) = 3.55$, MSE = 3.19, p = .06, n_p^2 = .05, BF_{01} = .10. However, this effect appears to be approaching significance and the Bayes factor appears to favor the alternative hypothesis (note that the inverse of the BF_{01} reported is BF_{10} = 9.70, which provides substantial evidence for the alternative hypothesis). Upon conducting post-hoc t-tests, this effect becomes clearer. When the test item contained studied spatial features,

participants did not provide significantly different familiarity ratings (M = 3.11, SD = 1.54) compared to curiosity ratings (M = 3.37, SD = 2.20), $t(71) = 1.27$, SE = .20, p = .21, BF_{01} = 3.57. However, when comparing test items that did not contain studied spatial features, a significant difference did emerge, such that participants provided significantly higher curiosity ratings (M = 3.26, SD = 2.30) compared to familiarity ratings (M = 2.72, SD = 1.44), $t(71) = 2.36$, SE = .23, p = .02, d = .27, BF_{10} = 1.73. However, the Bayes factor only provides anecdotal evidence for this effect. Collectively, these patterns of results suggest that curiosity ratings tended to be higher than familiarity ratings regardless of study status.

Information-Seeking Behaviors

The data pertaining to information-seeking behaviors are now examined. Overall, 44% of the participants used some but not all of their limited opportunities (N = 32), 51% used all of their limited resources (N = 37), and only 4% used none of their limited resources (N = 3; note that these three participants are necessarily excluded from the analyses presented below, as they did not use any of their limited resources). Upon analyzing resource allocation as a function of study-test block, participants used an average of 5.13 (SD = 2.10) resources on the first test block and an average of 4.65 (SD = 2.30) resources on the second test block. However, due to the experiment allowing participants to indicate “Yes, use limited resources” even when they were out of opportunities (note that during these instances, participants were shown the text “Out of limited opportunities”), 21 participants (29%) indicated “Yes, use limited resources” more than six times per test block. For example, one participant indicated “Yes, use limited resources” on eight of the test trials despite only having a total of 6 opportunities

to actually view the corresponding study scene. These instances of wanting to expend resources that did not exist were included in the analyses reported below, as one may argue that those instances in which the participant indicated “Yes, use limited resources” despite not having any might be the most intense in terms of curiosity, familiarity, and/or déjà vu.

Resource Allocation as a Function of Identification Status

To examine how participants allocated their limited resources to receive information concerning the current test scene, the proportions of trials on which participants indicated “Yes, use limited resources” that were labeled as identification failure versus success were computed. Overall, the proportion of trials on which identification failed and was accompanied by a “Yes” response (M = .15, SD = .06) was significantly lower than the proportion of trials on which identification succeeded in some form, either fully or partially, and was accompanied by a “Yes” response (M = .33, SD = .32), $t(65) = -4.41$, SE = .04, $p < .001$, $d = -.80$, $BF_{10} = 497.91$. Based on this pattern of results, it may be that participants were motivated to receive confirmatory feedback on their identification attempts.

Resource Allocation as a Function of Study Status

To examine how study status might influence participants’ information-seeking behaviors, specifically whether test scenes containing experimentally familiarized features might more strongly motivate participants to expend limited resources to discover the source of the familiarity, the proportions of trials on which participants indicated “Yes, use limited resources” that either did or did not contain experimentally familiarized features were computed, regardless of identification status. Overall, the

proportion of trials on which the test scene did contain experimentally familiarized features that were accompanied by a “Yes” response (M = .18, SD = .08) was significantly greater than the proportion of trials on which the test scene did not contain experimentally familiarized features that were accompanied by a “Yes” response (M = .14, SD = .06), $t(68) = 3.13$, SE = .01, p = .003, d = .50, BF_{10} = 10.83 (see Figure 13 below). This pattern of results suggests that test scenes containing experimentally familiarized elements more strongly motivated participants to use their limited resources than scenes that did not contain experimentally familiarized features.

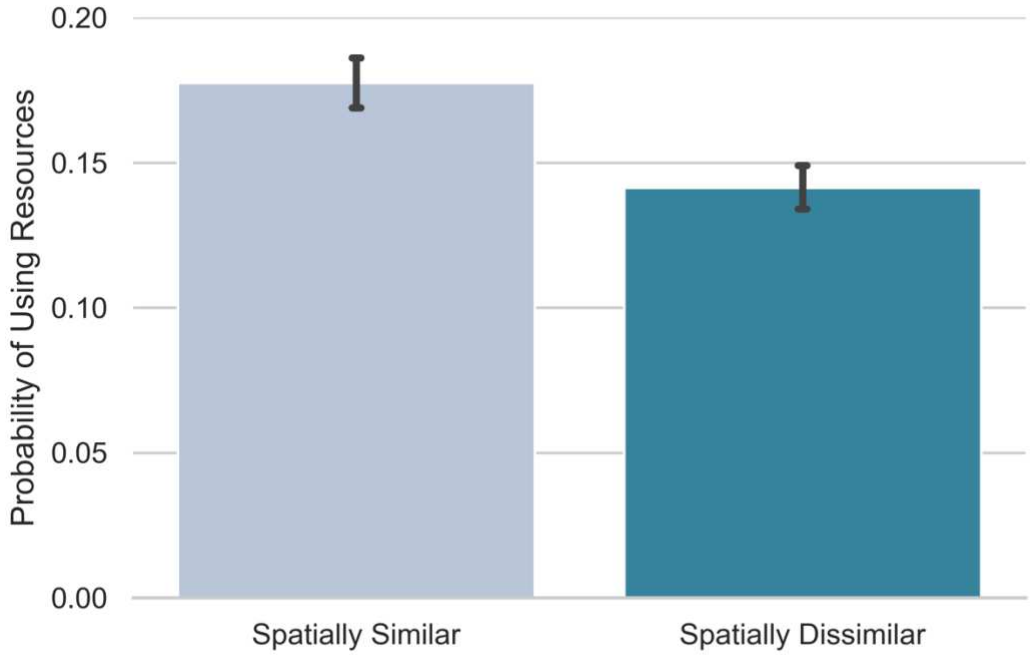

Figure 13. Probability of participants deciding to use their limited resources as a function of spatially similarity to scenes presented during the study phase, regardless of identification status. Error bars represent standard error of the mean.

In conducting the same analysis as reported above, but with the exclusion of trials on which identification succeeded (i.e., only focusing on trials in which identification failed), there was no difference in the proportion of trials that did contain

experimentally familiarized features and received a “Yes” response (M = .16, SD = .10) and those that did not contain experimentally familiarized features and received a “Yes” response (M = .14, SD = .06), $t(68) = 1.59$, SE = .01, p = .12, BF_{01} = 2.29 (see Figure 14 below). However, the means are in the predicted direction and the Bayes factor only provides anecdotal evidence in support of the null hypothesis.

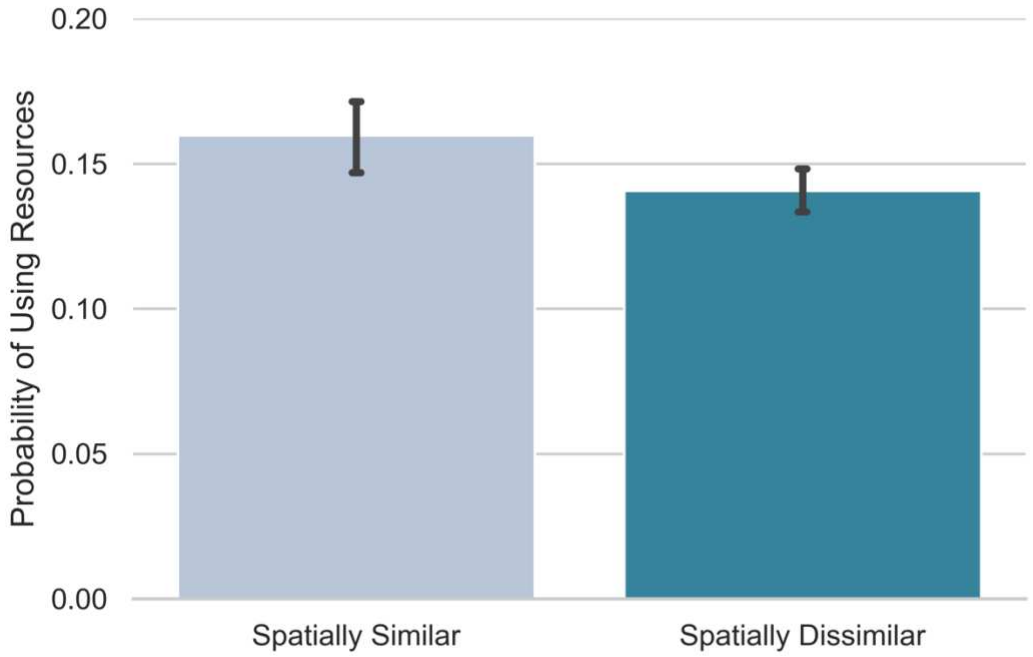

Figure 14. Probability of participants deciding to use their limited resources as a function of spatially similarity to scenes presented during the study phase, during instances of retrieval failure. Error bars represent standard error of the mean.

To examine how identification status might play a role in resource allocation when the test item did indeed contain experimentally familiarized features, the proportion of trials on which identification failed and participants indicated “Yes” (M = .16, SD = .10) was compared against the proportion of trials on which identification succeeded and participants indicated “Yes” (M = .33, SD = .32). There was a significant difference found, such that participants were more likely to use limited resources when

the test scene did correspond to an item presented at study and the participant successfully identified the study scene compared to when identification failed, $t(65) = 3.84$, $SE = .04$, $p < .001$, $d = .71$, $BF_{10} = 82.46$. Again, this pattern of results suggests that participants were motivated to use their limited resources in order to receive confirmatory feedback that their identification attempt was indeed correct.

Resource Allocation and Curiosity Ratings

The influence of participants' subjective curiosity ratings on their resource allocation was next examined. To do this, the average curiosity ratings provided during retrieval failure for trials on which participants decided to use their resources versus not use their resources were computed and compared using a paired-samples *t*-test. As hypothesized, participants were significantly provided significantly higher curiosity ratings for trials on which they decided to use their limited opportunities to receive information concerning the current test scene ($M = 5.03$, $SD = 2.63$) compared to when they decided against using their limited opportunities ($M = 3.12$, $SD = 2.22$), $t(68) = 8.78$, $SE = .22$, $p < .001$, $d = .77$, $BF_{10} = 9.83 \times 10^9$ (see Figure 15 below). These data are supportive of prior research suggesting that feelings of curiosity affect participants' information-seeking behaviors (e.g., Kang et al., 2009), such that their internal metacognitive signals drive them to seek out additional information concerning the current situation.

Resource Allocation and Feelings of Familiarity

To next examine the hypothesis that participants will provide higher familiarity ratings for trials on which they fail to identify the corresponding study scene and decide

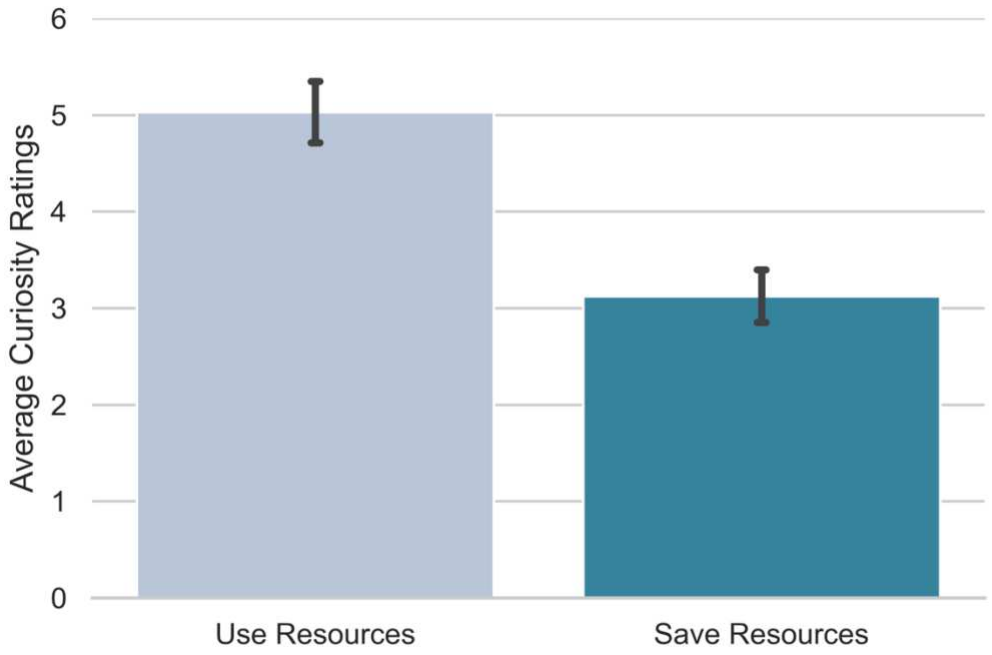

Figure 15. Average curiosity ratings provided during retrieval failure as a function of participants' decisions to use or save their limited opportunities. Error bars represent the standard error of the mean.

to use their resources versus not, the average familiarity ratings for such trials were computed. A paired-samples *t*-test revealed that participants did indeed provide significantly higher familiarity ratings during retrieval failure for trials on which they decided to use their limited opportunities to receive information about the current test scene ($M = 4.58$, $SD = 1.86$) compared to when they decided to not use their opportunities ($M = 2.65$, $SD = 1.46$), $t(68) = 8.71$, $SE = .22$, $p < .001$, $d = 1.14$, $BF_{10} = 7.44 \times 10^9$ (see Figure 16 below). This pattern of results is novel, as it suggests that familiarity serves a potentially adaptive function in that is encourages participants to seek out additional information about the current situation when they experience internal metacognitive signals that indicate high levels of familiarity.

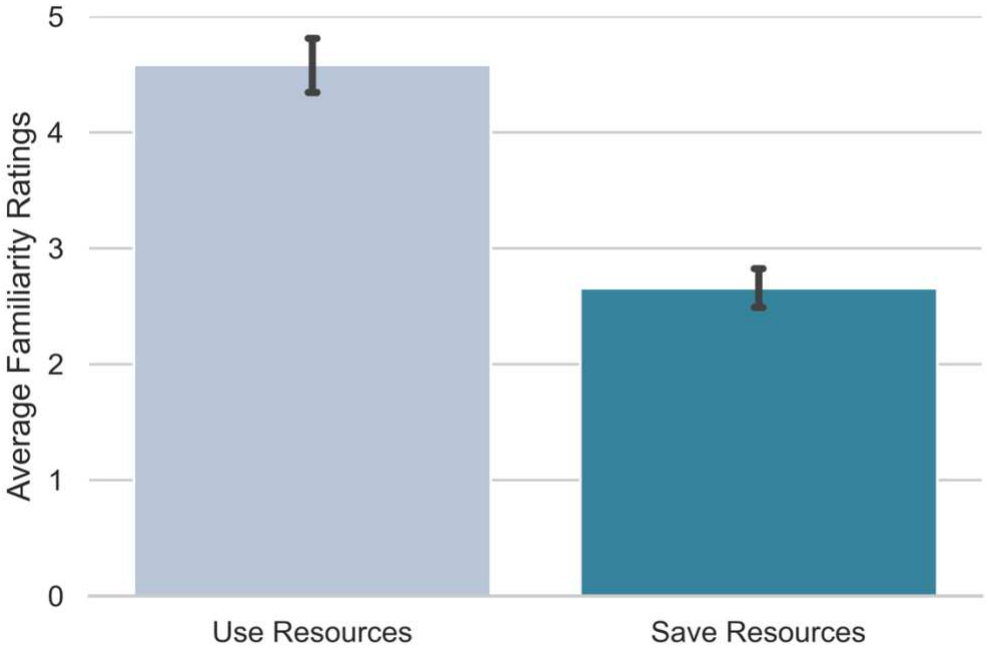

Figure 16. Average familiarity ratings provided during retrieval failure as a function of participants' decisions to use or save their limited opportunities. Error bars represent the standard error of the mean.

Resource Allocation as a Function of Déjà vu Reports

To examine the relationship between déjà vu states and the probability of participants using their limited resources to receive information concerning the current test scene, the proportions of trials on which participants decided to use their resources while concurrently experiencing déjà vu versus non were computed. Upon conducting a paired-samples *t*-test comparing these two proportions, it was found that participants were significantly more likely to indicate "Yes, use limited resources" while they were experiencing déjà vu ($M = .35$, $SD = .21$) compared to when they were not experiencing déjà vu ($M = .09$, $SD = .07$), $t(67) = 9.07$, $SE = .03$, $p < .001$, $d = 1.68$, $BF_{10} = 2.72 \times 10^{10}$ (see Figure 17).

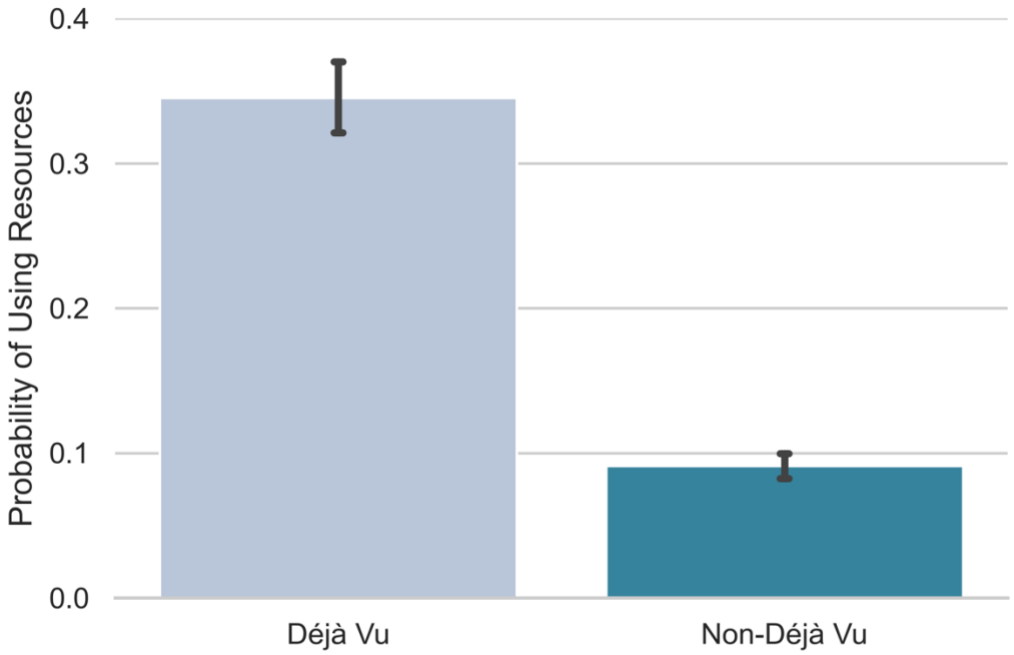

Figure 17. The probability of participants indicating "Yes, use limited resources" as a function of reported déjà vu state. Error bars represent the standard error of the mean.

When limiting the comparison to trials on which participants simultaneously experienced high levels of curiosity (i.e., provided a rating of 6 or higher). A paired-samples *t*-test indicated that, during instances of intense levels of curiosity, participants were significantly more likely to use limited opportunities to receive information concerning the current test scene while experiencing déjà vu ($M = .38$, $SD = .27$) compared to when they were not experiencing déjà vu ($M = .26$, $SD = .29$), $t(45) = 2.39$, $SE = .05$, $p = .02$, $d = .46$, $BF_{10} = 2.05$. Although this was a statistically significant difference, the magnitude of the effect was not greater when limiting the trials to instances of intense curiosity compared to when there was no such criterion, as hypothesized. Indeed, the effect size diminished (All Trials $d = 1.71$ versus High Curiosity Trials $d = .46$), and the Bayes Factor only provides anecdotal evidence for the alternative hypothesis.

Information-Seeking Behaviors when Opportunities Remained

The analyses reported above suggest that participants do indeed use their limited resources strategically, as demonstrated by the increased likelihood that participants will reserve their limited opportunities for trials on which they experience a sense of déjà vu, feel intense levels of curiosity, and/or are presented with a test scene containing experimentally familiarized spatial features. However, the analyses reported above included *all* trials, regardless of whether the participants had opportunities remaining. Although these analyses provide evidence for significant effects, there may be patterns of results that are hidden due to the inclusion of trials on which no more opportunities remained. For example, a participant may have been experiencing retrieval failure for a given test scene that contained experimentally familiarized spatial features, may have been experiencing déjà vu, high levels of familiarity, and high levels of curiosity; however, if they had already used up their limited resources on prior trials that they would have otherwise used on the current trial, then this would an inaccurate depiction of their resource allocation behaviors. Therefore, the analyses reported below focus only on trials in which participants still had limited opportunities remaining (i.e., they had one or more opportunities remaining).

Resource Allocation as a Function of Study Status

To examine how participants used their limited resources for trials on which resources still remained as a function of whether the test scene did or did not correspond to a spatially similar study scene, the proportion of trials on which participants indicated “Yes, use limited resources” and the test scene did correspond to a spatially similar study scene was computed and compared against trials on which

participants indicated “Yes, use limited resources” but the test scene did not correspond to a spatially similar study scene, regardless of identification status. On average, participants were significantly more likely to use their limited resources if the test scene contained experimentally familiarized spatial features (M = .22, SD = .13) than if the test scene did not contain experimentally familiarized features (M = .18, SD = .11), $t(68)$ = 2.83, SE = .02, p = .006, d = .33, BF_{10} = 5.11. This pattern of results is similar to that reported above examining *all* trials (not limited to trials on which participants still had remaining resources), suggesting that participants are more likely to use their limited resources when the current novel stimulus contains a pattern of familiar elements compared to when it does not.

In examining how this behavior might emerge during instances of retrieval failure, the same analysis reported above was conducted, but this time only including trials on which identification failed. Overall, participants experiencing retrieval failure were not more likely to use their limited resources on trials corresponding to spatially similar study scenes (M = .21, SD = .15) compared to those that did not correspond to spatially similar study scenes (M = .18, SD = .11), $t(68)$ = 1.71, SE = .01, p = .09, BF_{10} = .53.

As previously reported above, participants may be reserving their limited opportunities for instances in which they are seeking confirmatory feedback on their provided answer and are thus more likely to use their resources on trials in which identification did indeed succeed. To determine whether a similar pattern of results would emerge when focusing on trials in which opportunities still remained, the proportions of trials on which the test scene did correspond to spatial features presented at study and participants indicated “Yes, use limited resources” were

computed as a function of identification success versus identification failure. Indeed, as previously found, participants were more likely to indicate “Yes, use resources” for trials on which identification succeeded (M = .38, SD = .36) compared to trials on which identification failed (M = .21, SD = .15), $t(65) = 3.60$, SE = .05, $p < .001$, d = .26, BF_{10} = 39.91.

Collectively, these patterns of results are similar to those presented above in which *all* trials were included, suggesting overall that participants tend to use their limited resources to discover information about novel test scenes that contain experimentally familiarized spatial features, and that they tend to use their resources in order to receive confirmatory feedback, as demonstrated by their increased use of resources on trials in which identification succeeded.

Resource Allocation and Curiosity Ratings

The relationship between subjective feelings of curiosity and participants’ decisions to use limited resources was demonstrated above when analyzing *all* trials, such that participants provided significantly higher curiosity ratings on trials for which they decided to use their limited resources compared to when they did not. To assess whether the magnitude of this relationship would increase when only focusing on trials in which participants still had opportunities remaining, a paired-samples *t*-test was conducted, comparing participants’ average curiosity ratings for trials on which they did versus did not decide to use their limited resources. Indeed, not only did participants provide significantly higher curiosity ratings for trials on which they decided to use their resources (M = 5.14, SD = 2.63) compared to when they decided against using their resources (M = 2.98, SD = 2.18), $t(68) = 9.71$, SE = .21, $p < .001$, d = .83, BF_{10} = 3.99 x

10^{11}, the magnitude of the effect also increased. When comparing curiosity ratings as a function of decisions on whether or not to use limited resources, including *all* trials, regardless of whether resources even remained, produced a *d* of .77. When limiting the analysis to only trials on which resources actually remained, the effect size increased (*d* = .83). Although both of these effect sizes can be categorized as large (see Lakens, 2013), it is worth noting that the effect size was diminished when including trials on which participants did not have any resources to utilize.

Resource Allocation and Feelings of Familiarity

As reported above, participants provided significantly higher familiarity ratings for trials on which they indicated "Yes, use limited resources" compared to those on which they indicated "No, do not use limited resources." To assess how this pattern of results might change when only including trials on which there were still resources to allocate, a paired-samples *t*-test was conducted, comparing the average familiarity ratings provided on trials during retrieval failure as a function of participants' decisions to use their limited resources. As was found previously, participants provided significantly higher familiarity ratings for trials on which they indicated "Yes, use limited resources" ($M = 4.61$, $SD = 1.88$) compared to those on which they indicated "No, do not use limited resources" ($M = 2.59$, $SD = 1.50$), $t(68) = 8.89$, $SE = .23$, $p < .001$, $d = 1.18$, $BF_{10} = 1.48 \times 10^{10}$. Unlike with the effect size increase with curiosity ratings, limiting trials to only those on which limited resources still remained did not dramatically increase the magnitude of the effect on familiarity ratings ($d = 1.14$ compared to $d = 1.18$).

Resource Allocation as a Function of Déjà vu Reports

To assess the probability of participants using their limited resources during retrieval failure as a function of déjà vu state, I next report the analysis in which the trials analyzed were those on which participants still had remaining limited opportunities. As found before, participants were significantly more likely to indicate "Yes, use limited resources" during reported déjà vu states (M = .41, SD = .24) than non-déjà vu states (M = .12, SD = .13), $t(67) = 8.96$, SE = .03, $p < .001$, d = 1.55, $BF_{10} = 1.79 \times 10^{10}$. This pattern of results is similar to that reported above in which *all* trials were analyzed, further suggesting that déjà vu may serve as an adaptive signal for participants, as it may motivate them to engage in information-seeking behaviors.

When further limiting this analysis to only include trials on which participants also experienced intense levels of curiosity (i.e., they provided a curiosity rating of 6 or greater), a similar pattern emerged. Participants experiencing déjà vu were significantly more likely to indicate "Yes, use limited resources" (M = .50, SD = .31) than when they reported a non-déjà vu state (M = .31, SD = .32), $t(44) = 2.6$, SE = .06, p = .01, d = .60, BF_{10} = 3.82.

Summary

Taken together, the findings of Experiment 1 inform both the curiosity domain and the metacognition domain. The analyses reported above provide insight into the potential mechanisms of curiosity, specifically relating to Berlyne's (1950; 1960) relative novelty proposal that a novel situation containing familiarized elements should prompt strong feelings of curiosity. Indeed, such an effect was found in the current experiment, with test scenes corresponding to spatially similar study scenes receiving higher

curiosity ratings than those that did not spatially correspond to study scenes. Although this effect did not emerge when only considering instances of retrieval failure, the results support Berlyne's proposal. Further, the current experiment informs metacognition research, in that it suggests déjà vu may serve an adaptive role in encouraging increased feelings of curiosity and information-seeking behaviors. As discussed above, déjà vu was associated with significantly higher curiosity ratings as opposed to non-déjà vu, in addition to changes in information-seeking behaviors, such that participants were more likely to make commission errors during déjà vu than non-déjà vu in addition to spending more time attempting to retrieve the name of the corresponding study scene on the Recall prompt while experiencing déjà vu. Further, there was an increased inclination to expend limited resources to receive relevant information concerning the current test scene when participants experienced a sense of déjà vu compared to non-déjà vu. Collectively, these results provide support for the proposal that déjà vu may be an adaptive mechanism of the memory system, potentially signaling to the experiencer that they should continue to search memory for relevant information.

Chapter 3 – Experiment 2 (Déjà Entendu and Feelings of Curiosity)

The purpose of Experiment 2 was to examine how feelings of curiosity and information-seeking behaviors occur when the target information is auditory in nature. Indeed, as previously discussed, there is currently a gap in the literature examining how people experience feelings of curiosity for information that is auditory, such as music (Litman, 2009). Therefore, Experiment 2 examined how participants' feelings of curiosity and information-seeking behaviors manifest when they are experiencing déjà entendu, which is the feeling of having heard something before despite knowing otherwise. Further, the relationship between feelings of familiarity and feelings of curiosity was examined by incrementally increasing the level of featural overlap between the current test cue and stored memory traces. Prior research has shown that increasing the amount of featural overlap between study and test, with both visual and auditory stimuli, subsequently increases participants' perceived familiarity with the cue (e.g., Huebert et al., 2021; McNeely-White et al., 2021; Ryals & Cleary, 2012). For example, McNeely-White et al. conducted an experiment in which participants heard isolated song rhythms during the encoding phase. Each unique rhythm was played either once or three times throughout the encoding phase, thus creating either one or three memory traces containing those isolated rhythm features. At test, participants were presented with the whole, unaltered versions of these songs. When participants experienced retrieval failure, they were significantly more likely to find the test song familiar if it contained an experimentally familiarized rhythm sequence compared to if it did not. The features held within the test song matched the features held within memory traces. Further,

participants were even more likely to find the test song familiar if its rhythm had been familiarized three times during the encoding phase than if it had only been familiarized once. When the test songs matched onto even more memory traces (i.e., there were three memory traces containing the current test song's rhythm features), this produced a stronger familiarity signal during retrieval failure. The primary interest of the current study was whether this increased feature familiarization method might similarly increase feelings of curiosity during retrieval failure as it does with feelings of familiarity. If so, how might this affect information-seeking behaviors?

To further examine how increased feature overlap between the features of the current test cue and those stored within memory traces might subsequently increase participants' feelings of curiosity during retrieval failure and how this relates to the auditory retrieval-failure-based metacognitive experience of déjà entendu, Experiment 2 was conducted. Specifically, the RWI paradigm was used to extend the findings of Kostic and Cleary (2009), McNeely-White et al. (2021), and McNeely-White and Cleary (in progress) to examine whether participants experiencing déjà entendu also report higher feelings of curiosity, and how this might be affected be the amount of featural overlap. Towards this end, participants were first presented with unaltered, whole songs (e.g., "Mary Had a Little Lamb") either once or three times at study.

During the test phase, participants were then presented with isolated tonal sequences, some of which corresponded to whole songs presented at study while others did not. I hypothesized that the embedding of familiarized auditory features at test would increase participants' perceived feelings of curiosity during retrieval failure. Further, I hypothesized that participants would feel the most intense levels of curiosity

when the test cue mapped onto multiple memory traces created during the encoding phase compared to only one. Additionally, I hypothesized that participants would be more likely to use limited opportunities to discover information about the current test cue (i.e., hear the whole, unaltered song along with seeing the name of the song) when the test cue mapped onto multiple memory traces as opposed to only one. This was done by allowing participants to indicate on 25% of the trials that they would like to receive information about the current test cue. Additionally, I hypothesized that there would be a relationship between déjà entendu and high levels of curiosity, such that participants would provide higher curiosity ratings for the current auditory test cue while experiencing déjà entendu. Finally, if there is a relationship between déjà entendu and feelings of curiosity, I hypothesized that participants would be more likely to indicate that they want to use limited resources to receive information about the current auditory test cue when they are experiencing déjà entendu than when they are not.

Method

Participants

Participants consisted of 145 undergraduate students from Colorado State University. Three participants were lost from data analysis, though, due to either computer errors (i.e., two participants pressed the Windows key, which caused the E-Prime program to crash) or not understanding the instructions, leaving a sample size of 142. A power analysis had previously been conducted, which was based on the sample sizes used in the experiments of McNeely-White et al. (2021) and McNeely-White and Cleary (in progress). These prior experiments typically found small RWI and déjà entendu effect sizes when the test clip contained familiarized isolated rhythm

sequences. However, experimentally familiarized tonal sequences sometimes produce larger RWI and déjà entendu effect sizes. For example, McNeely-White and Cleary (in progress) found that, compared to when the current test song clip did not contain any familiarized features from study, participants were significantly more likely to find the song familiar if it contained an isolated tonal sequence that had been familiarized once (d = .23) or three times (d = .41). Additionally, when comparing test song clips that contained familiarized tonal sequences presented zero versus three times at study, they found a medium déjà entendu effect size (d = .41). Given the experimental design differences between the current experiment (the study list consisted of whole song clips while the test list was a pure list of only isolated tonal sequences), a conservative power analysis was conducted using G*Power (Faul et al., 2007), with power set to .90, an α of .05, and a small effect size (d = .30), which indicated that a sample size of 119 would be sufficient to detect such an effect. As no prior research has been conducted examining feelings of curiosity and déjà entendu, I aimed to have a sample size of at least 140 participants, which was achieved.

Materials

The stimuli consisted of 84 of the piano song clips and their isolated tonal sequences created by Kostic and Cleary (2009) and later used by McNeely-White et al. (2021), which are all well-known pieces, such as children's melodies and pop songs. To isolate tonal information, the notes were extracted from each song and played in their original order but according to a different, unstudied rhythm. All tonal sequences adhered to the same arbitrary rhythm except for a few, whose original rhythms were too

similar to the arbitrary rhythm. In these instances, the tonal sequences adhered to a new, unstudied arbitrary rhythm.

Three counterbalanced versions of the experiment were created such that each whole, unaltered song clip fell into each of the exposure conditions across participants. For example, in the first counterbalanced version, the isolated tonal sequence test song clip “Copacabana” corresponded to a whole, unaltered song clip presented once at study. In the second version, “Copacabana” did not correspond to any whole, unaltered song clip presented at study. Finally, in the third version, “Copacabana” corresponded to a whole, unaltered song clip presented three times at study.

Procedure

Participants were randomly assigned to one of the three counterbalanced versions of the experiment. The 84 song segments were divided into seven study-test blocks, with each study list consisting of eight unique whole, unaltered song clip presentations. Four of these whole, unaltered song clips were presented one time throughout the study list (the Exposure1X condition), while the other four were presented three times throughout the study list, resulting in 12 total presentations (the Exposure3X condition). These repetitions were randomly dispersed throughout the study list, such that participants heard 16 whole, unaltered song clips at study. At test, there were 12 unique isolated tonal sequences presented, such that 1/3 corresponded to the Exposure1X condition, 1/3 corresponded to the Exposure3X condition, and the remaining 1/3 did not correspond to any whole, unaltered song presented during the study phase (the Exposure0X condition). All of the isolated tonal sequences were randomly presented throughout the test list.

Prior to beginning the experiment, participants were presented with instructions explaining the experimental task (see Appendix B). They also heard an example of a whole, unaltered song clip, which was not presented during the study block, and then its isolated tonal sequence. After reading through the instructions, the participants then began the first study block, which consisted of 16 whole, unaltered song clips. After listening to each song clip, they were asked to try and identify it, such as by typing in the name of the song, the lyrics, or any other information they could conjure up.

Once the study list of 16 whole, unaltered song clips was completed, participants were then given specific instructions for the first test block. They were instructed that they would now hear a new list of audio clips, but this time the audio clips would be of isolated tonal sequences, some of which would be from songs presented at study while some would not. After each isolated tonal sequence, they would be asked a series of questions. First, they would be asked to indicate whether they are experiencing déjà entendu, which was defined as “The feeling of having heard something before despite knowing it is new.” Next, they were told that they would be asked to indicate how familiar the current isolated tonal sequence feels on a scale of zero (*Not at all familiar*) to 10 (*Extremely familiar*). Next, they would be asked to indicate how curious they feel about the isolated tonal sequence on a scale of zero (*Not at all curious*) to 10 (*Extremely curious*). They would then be asked to indicate whether they can identify the isolated tonal sequence, and if so, to type in the name of the song. Finally, they were told that they would have limited opportunities to discover information concerning the isolated tonal sequence, specifically to which studied whole, unaltered song (if any) that it corresponds. However, they would only be able to use these limited opportunities on

25% of the trials.[2] If they indicated that they did want to receive information concerning the isolated tonal sequence, then the whole, unaltered song would play along with its name being displayed on the screen. However, if there was no whole, unaltered song that was presented at study, then the text "This isolated tonal sequence does NOT correspond to a song clip heard at study" was displayed.

After receiving the test instructions, participants then began the first test block, consisting of 12 isolated tonal sequences, all of which were randomly ordered. After hearing each isolated tonal sequence, participants were asked a series of questions concerning the song fragment, with each question appearing one at a time on the screen. First, participants were asked to indicate whether they were experiencing déjà entendu (Y=Yes, N=No). The second question prompted them to indicate how familiar the isolated tonal sequence felt on a scale of zero to 10 (*0 = Not at all familiar, 10 = Extremely familiar*). The third question prompted participants to indicate how curious they felt about the isolated tonal sequence on a scale of zero to 10 (*0 = Not at all curious, 10 = Extremely curious*). The fourth question asked participants whether they could identify the name of the isolated tonal sequence, and if so, to type in that information. Finally, participants were asked to indicate whether they would like to use their limited opportunities to receive information about the current isolated tonal sequence (Y=Yes, N=No). For this question, the remaining number of opportunities was displayed at the bottom of the screen so that the participant was reminded of how many

[2] In Experiment 1, participants were allowed to use their limited opportunities on 20% of the trials, whereas in Experiment 2 they were allowed to do so on 25% of the trials. This discrepancy is largely due to the fact that each test list in Experiment 2 consisted of 12 isolated tonal sequences, which does not divide evenly by .20 (that would result in 2.4 opportunities per test list). Additionally, as previously stated, there is little precedent in the literature concerning how scarce these limited resources must be.

opportunities remained (e.g., “2 opportunities remaining” was displayed at the bottom left-hand side of the screen). Additionally, the trial number was also displayed so that the participant knew how many trials remained, enabling them to better manage their resources (e.g., “Trial 3/12” was displayed at the bottom right-hand side of the screen). If the participant indicated that they did indeed want to receive information concerning the isolated tonal sequence, then the whole, unaltered song began to play along with the visual text of the name of the song appearing in the middle of the screen (e.g., “*A Spoonful of Sugar*” appeared on the screen). However, if the isolated tonal sequence did not correspond to a whole, unaltered song presented at study, then the text “This song fragment does NOT correspond to a song clip heard at study” appeared in the middle of the screen. Finally, if the participant indicated that they did want to use their limited resources but had no more remaining, the text “Out of limited opportunities” appeared in the middle of the screen.

After completing the first study-test block, participants then proceeded to the second study-test block. Additionally, the limited-opportunities counter was reset, such that participants were able to use their limited resources to receive information pertaining to the current test song clip on 25% of the current test list’s trials. The same procedure was used for all seven study-test blocks.

Note that due to a programming error, participants in the second version of the experiment received the incorrect study list for the fifth study-test block. Therefore, this study-test block for these 31 participants was excluded from data analysis. Once the mistake was caught, the program was corrected and all subsequent participants in version two received the correct study list for the fifth study-test block.

Results

Identification Rates

Trials were hand-labeled as being either an instance of identification success (e.g., the participant typed in the correct target), partial identification success (e.g., the participant typed in "classical lullaby" for the song "*Rockabye Baby*"), or identification failure (e.g., they provided the incorrect song's name or left the prompt blank). The trials labeled as identification failure were further subdivided as being an instance of a commission error (the participant typed in incorrect information) or an omission error (the participant did not type in anything or typed "don't know," "can't remember," etc.).

On average, participants identified, either fully or partially, 22% (SD = .11) of the whole, unaltered songs presented at study and 9% (SD = .07) of the isolated tonal sequences presented at test. In examining the identification rates for songs presented at study, participants were significantly more likely to identify the whole, unaltered study song if it was presented three times as opposed to only once, $t(141) = 5.88$, $SE = .01$, $p < .001$, $d = .49$, $BF_{10} = 3.65 \times 10^5$ (see Table 1 for descriptive statistics). This pattern of is similar to that found by McNeely-White et al. (2021), who also demonstrated that participants are more likely to identify songs presented three times throughout the study list as there are more opportunities for identification success to occur compared to when the song is only presented once.

Table 1. Proportion of song clips correctly identified during the study and test phases.

	Study		Test	
Exposure Condition	*M*	*SD*	*M*	*SD*
0X	-	-	0.02	0.03
1X	0.19	0.12	0.09	0.09
3X	0.25	0.13	0.14	0.12

Turning to the identification rates for isolated tonal sequences presented during the test phase, a one-way ANOVA on Exposure Condition was conducted, which revealed a significant effect, $F(2, 282) = 120.39$, $MSE = .01$, $p < .001$, $n_p^2 = .46$, $BF_{10} = 3.49 \times 10^5$. When participants were presented with an isolated tonal test sequence that was previously presented one time as a whole, unaltered song during the study phase, they were significantly more likely to identify it than if it had not been previously presented during study as a whole, unaltered song, $t(141) = 10.00$, $SE = .01$, $p < .001$, $d = 1.01$, $BF_{10} = 1.31 \times 10^{15}$ (see Table 1 above for descriptive statistics). Further, if the isolated tonal test sequence had been previously presented as a whole, unaltered song three times during the study phase, participants were significantly more likely to identify it than if it had only been presented once, $t(141) = 7.16$, $SE = .01$, $p < .001$, $d = .49$, $BF_{10} = 1.98 \times 10^8$.

Identification Errors as a Function of Study Status

The commission error rates found in Experiment 1 suggested that participants are marginally more likely to make a commission error for items that do versus do not correspond to information presented during the study phase. To assess whether such an effect would emerge in the current experiment, the probabilities that participants would make a commission error on trials corresponding to whole, unaltered songs presented during the study phase versus trials that did not correspond to whole, unaltered songs presented during the study phase were computed. A significant effect was found, such that participants were significantly more likely to make a commission error for isolated tonal test sequences that did correspond to whole, unaltered songs presented at least once during the study phase ($M = .03$, $SD = .04$) compared to those

that did not (M = .02, SD = .05), t(141) = 2.70, SE = .004, p = .01, d = .22, BF_{10} = 3.08. When restraining this analysis to only include participants who did indeed make commission errors (note that 41 participants never made commission errors during the test phase and were therefore lost; therefore the degrees of freedom were only 100 instead of 141), the magnitude of the effect increased. Participants were significantly more likely to make a commission error on trials corresponding to whole, unaltered songs presented during the study phase (M = .05, SD = .04) compared to those corresponding to songs not presented at study (M = .03, SD = .05), t(100) = 2.73, SE = .001, p = .01, d = .44, BF_{10} = 3.64.

Further, when analyzing the probability of participants making a commission error as a function of exposure condition (Exposure0X, Exposure1X, Exposure3X), a one-way repeated-measures ANOVA revealed a significant effect, F(2, 282) = 3.44, MSE = .001, p = .03, n_p^2 = .02, BF_{10} = .62. When presented with an isolated tonal test sequence that did correspond to a whole, unaltered song presented once during the study phase, participants were significantly more likely to make a commission error (M = .03, SD = .05) compared to if it did not correspond to a whole, unaltered study song (M = .02, SD = .05), t(141) = 2.49, SE = .004, p = .01, d = .21, BF_{10} = 2.15. There was no significant increase in the probability of making a commission error when study exposure was increased from one to three instances (M = .03, SD = .05), t(141) = .20, SE = .004, p = .85, BF_{01} = 8.95. Overall, though, the significant effects are only anecdotally, at best, supported by the Bayes factors. Indeed, even when limiting these analyses to only include those who did make commission errors during the experiment, there was no evidence favoring the alternative hypothesis over the null, F(2, 200) =

3.47, MSE = .002, p = .03, n_p^2 = .03, BF_{10} = .82. Participants were more likely to make a commission error for an isolated tonal sequence that did correspond to a whole, unaltered song presented once during the study phase (M = .05, SD = .05) compared to when it did not (M = .03, SD = .05), t(100) = 2.51, SE = .01, p = .01, d = .40, BF_{10} = 2.16; however, the Bayes factor only provides anecdotal evidence for such an effect. Again, there was no significant difference in the probability of making a commission error for isolated tonal sequences originally exposed once versus three times (M = .05, SD = .05) during the study phase as a whole, unaltered song, t(100) = .20, SE = .01, p = .85, BF_{01} = 8.91. Collectively, these results suggest that when participants experienced retrieval failure for an isolated tonal clip, but sensed that *something* in memory was relevant to the current situation, they were potentially motivated to engage in internal information-seeking behaviors, which manifested at increased attempts to produce the target song's name, despite failing to do so.

Recognition without Identification

Prior research has demonstrated that, during retrieval failure, participants provide higher familiarity ratings to test songs that correspond to experimentally familiarized musical features from study versus those that do not (e.g., Kostic & Cleary, 2009; McNeely-White et al., 2021). To assess whether a similar pattern would emerge in the current study, participants' subjective familiarity ratings provided during retrieval failure for isolated tonal sequences that did correspond to a whole, unaltered study song were compared with the familiarity ratings provided during retrieval failure for isolated tonal sequences that did not correspond to a whole, unaltered study song. Indeed, a significant pattern emerged, such that, when participants failed to identify the isolated

tonal test sequence, they provided significantly higher familiarity ratings for tonal sequences that did correspond to previously-heard whole, unaltered study songs (M = 4.52, SD = 1.83) compared to those that did not (M = 3.32, SD = 1.61), $t(141) = 17.57$, SE = .07, $p < .001$, d = .67, $BF_{10} = 1.05 \times 10^{34}$ (see Figure 18 below). Despite participants failing to identify the isolated tonal test sequence, they were still able to discriminate, presumably based on internal familiarity signals, those that did correspond to information presented during the study phase from those that did not.

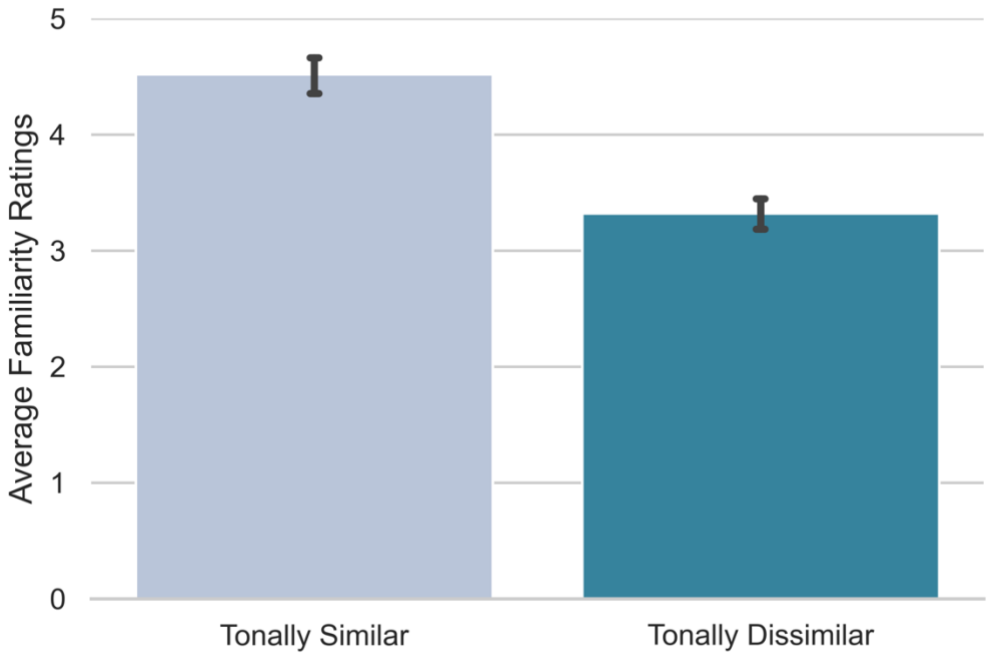

Figure 18. Average familiarity ratings provided during retrieval failure as a function of tonal similarity to a study song. Participants provided significantly higher familiarity ratings for isolated tonal sequences that corresponded to a whole, unaltered study song compared to those that did not. Error bars represent standard error of the mean.

Familiarity Ratings as a Function of Exposure Condition

Turning now to assess how increased exposure (Exposure0X, Exposure1X, Exposure3X) to the whole, unaltered song at study would affect subjective familiarity ratings provided during retrieval failure during the test phase, a one-way repeated-

measures ANOVA was conducted. Indeed, as has been shown in prior research examining musical feature repetition (McNeely-White et al., 2021), a significant effect emerged, $F(2, 282) = 201.81$, $MSE = .43$, $p < .001$, $n_p^2 = .59$, $BF_{10} = 2.64 \times 10^{51}$ (see Figure 19). When presented with an isolated tonal test sequence that corresponded to a whole, unaltered song presented once during the study phase ($M = 4.19$, $SD = 1.77$), participants provided significantly higher familiarity ratings compared to when the isolated tonal test sequence did not correspond to any whole, unaltered study song ($M = 3.32$, $SD = 1.61$), $t(141) = 13.04$, $SE = .07$, $p < .001$, $d = .51$, $BF_{10} = 6.98 \times 10^{22}$. Further, participants provided significantly higher familiarity ratings during retrieval failure for isolated tonal test sequences that corresponded to a whole, unaltered song presented three times at study ($M = 4.87$, $SD = 2.00$) compared to those that only corresponded to a whole, unaltered song presented once at study, $t(141) = 9.05$, $SE = .08$, $p < .001$, $d = .35$, $BF_{10} = 5.92, \times 10^{12}$.

Déjà Entendu

When experiencing retrieval failure, participants reported experiencing a sense of déjà entendu on 45% ($SD = .20$) of the trials. As previously discussed, one manifestation of increased information-seeking behaviors during retrieval-failure-based metacognitive states might be increased memory search times, as the participant may be internally searching for relevant information stored within memory. Such an effect was found in Experiment 1, with participants spending longer on the Recall prompt while experiencing déjà vu than non-déjà vu. A similar effect was found in Experiment 2. Examination of the reaction time data, as measured by the amount of time participants

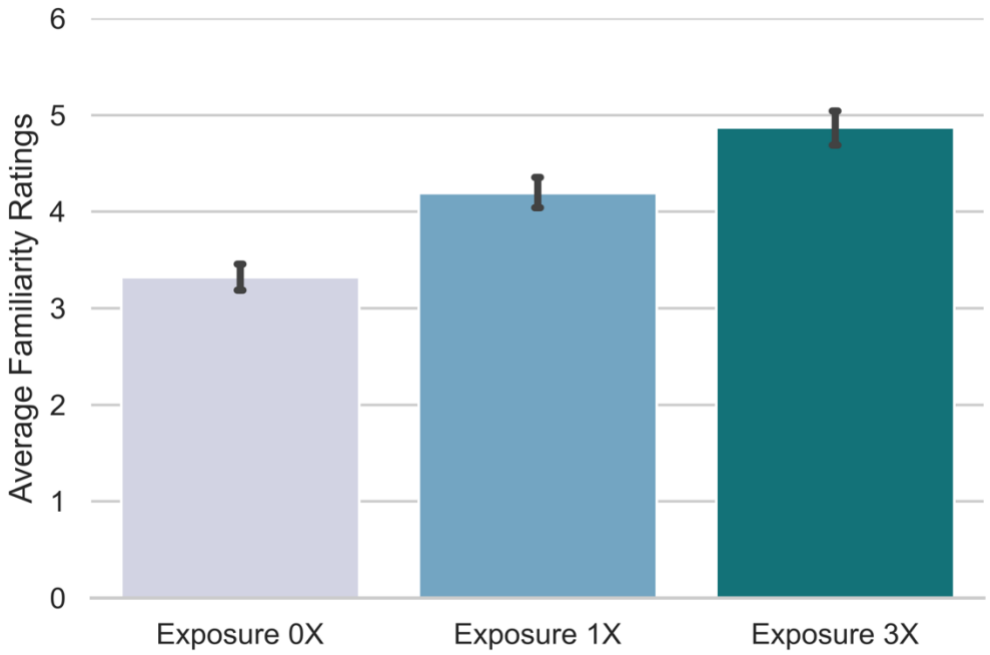

Figure 19. Average familiarity ratings provided during retrieval failure as a function of exposure condition. Error bars represent standard error of the mean.

remained on the Recall prompt before proceeding, revealed a significant difference as a function of reported déjà entendu state. Specifically, on trials associated with a déjà entendu report, participants spent a significantly longer amount of time on the Recall prompt (M = 2016.80 ms, SD = 1152.03 ms) compared to trials associated with a non-déjà entendu report (M = 1363.79 ms, SD = 689.59 ms), $t(141) = 8.52$, SE = 76.63, $p < .001$, $d = .63$, $BF_{10} = 3.04 \times 10^{11}$. These findings are similar to those of Experiment 1, such that the presence of a déjà entendu state might encourage the experiencer to spend more time searching for relevant information as to why they are experiencing this strange metacognitive sensation.

Probability of Déjà Entendu Given Study Status

Prior research examining the circumstances under which déjà entendu is more likely to be reported has shown that, when test stimuli contain experimentally

familiarized features, participants experiencing retrieval failure are more likely to report experiencing a sense of déjà entendu (McNeely-White & Cleary, in progress). However, this effect has not been consistently shown, as McNeely-White and Cleary (2019) did not demonstrate a straightforward pattern such as this when using *Piano Puzzlers* stimuli (although this may have been due to the nature of the stimuli themselves), and McNeely-White and Cleary (in progress), while finding such an effect in Experiment 1 only showed a marginally significant effect in Experiment 2. However, McNeely-White and Cleary (in progress) used a mixed-list design in addition to experimentally familiarizing isolated features at study instead of the whole, unaltered songs. The current study used a difference approach to examine the mechanisms underlying déjà entendu by presenting whole, unaltered songs at study and using a pure-list design consisting only of isolated tonal test sequences. In using this method, I found that participants, while failing to identify the isolated tonal test sequence, were significantly more likely to report experiencing a sense of déjà entendu if there had been a corresponding whole, unaltered song presented during the study phase (M = .51, SD = .22) compared to if there had not (M = .34, SD = .20), $t(141) = 14.63$, SE = .01 $p < .001$, d = .76, $BF_{10} = 6.97 \times 10^{26}$ (see Figure 20 below).

Probability of Déjà Entendu Given Exposure Condition

To further investigate the potential circumstances under which déjà entendu is more likely to occur, I now turn to examining the influence of exposure to the original whole, unaltered song at study. A one-way repeated-measures ANOVA examining the influence of Exposure Condition (Exposure 0X, Exposure 1X, Exposure 3X) on the

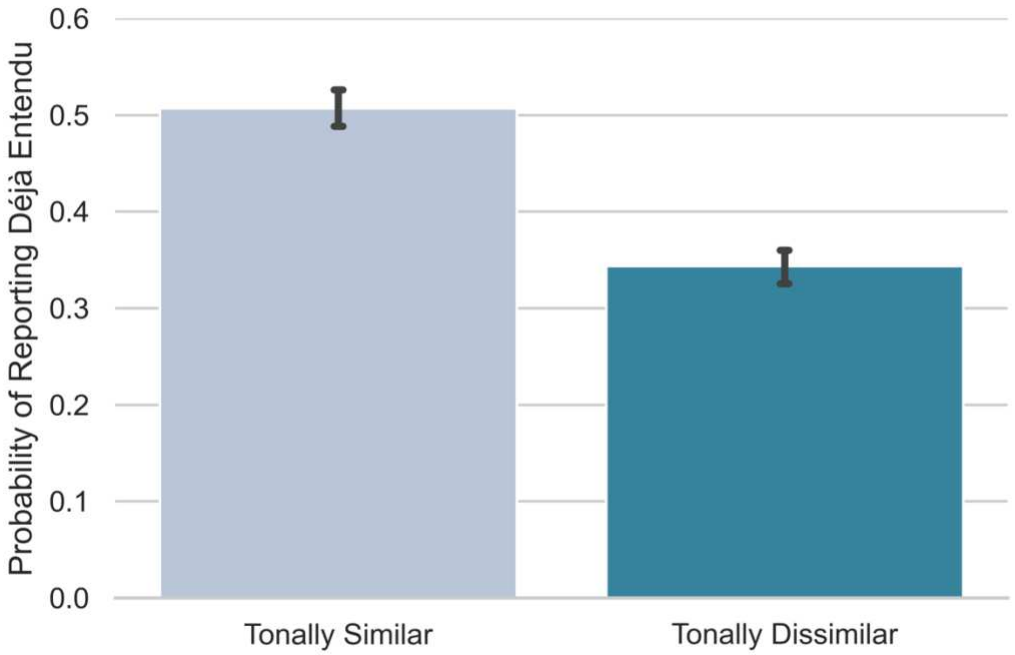

Figure 20. The probability of reporting déjà entendu during retrieval failure as a function of tonal similarity to a studied song. Participants were significantly more to report experiencing déjà entendu for isolated tonal sequences that did correspond to whole, unaltered songs presented at study compared to those that did not. Error bars represent the standard error of the mean.

probability of reporting déjà entendu for an isolated tonal test sequence during retrieval failure revealed a significant effect, $F(2, 282) = 145.25$, $MSE = .01$, $p < .001$, $n_p^2 = .51$, $BF_{10} = 3.42 \times 10^{40}$ (see Figure 21 below). Specifically, when participants were experiencing retrieval failure for an isolated tonal test sequence that corresponded to a whole, unaltered song presented at once at study, they were significantly more likely to report a sense of déjà entendu ($M = .46$, $SD = .22$) compared to when the isolated tonal test sequence did not correspond to any whole, unaltered song presented during the study phase ($M = .34$, $SD = .20$), $t(141) = 10.50$, $SE = .01$, $p < .001$, $d = .47$, $BF_{10} = 2.44 \times 10^{16}$. Further, when exposure to the whole, unaltered song was increased to three separate instances during the study phase, participants were significantly more likely to

report a sense of déjà entendu (M = .56, SD = .25) compared to if there was only one exposure instance during the study phase, $t(141) = 7.90$, SE = .01, $p < .001$, d = .40, BF_{10} = 1.02 x 10^{10}. These findings provide evidence for déjà entendu occurring due to feature-matching processes and that increasing the number of stored memory traces containing those features subsequently increases the probability that one will experience a sense of déjà entendu.

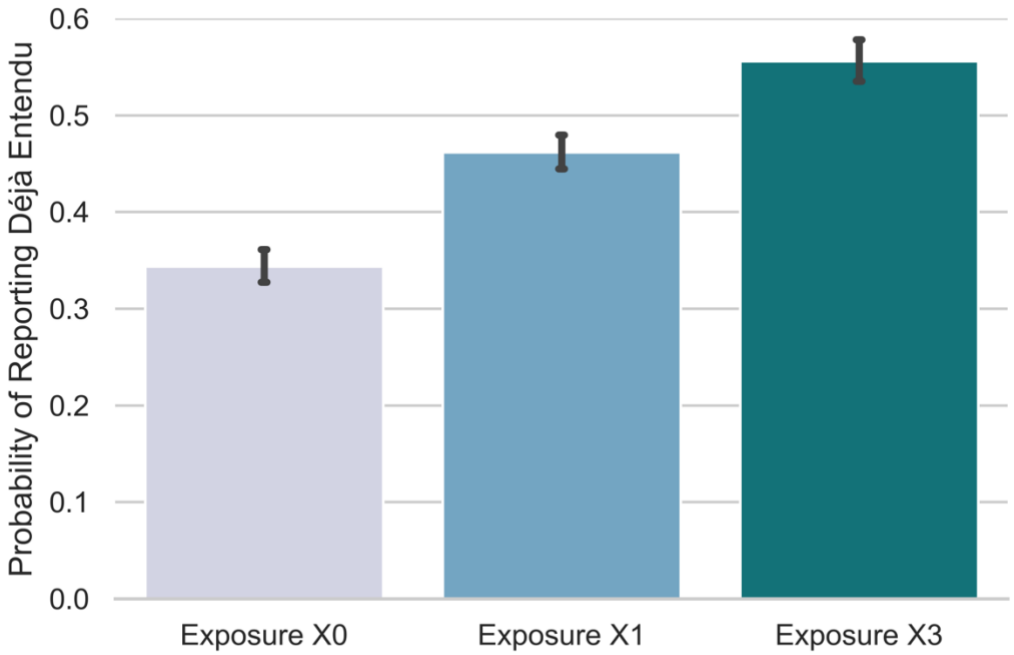

Figure 21. The probability of reporting déjà entendu during retrieval failure as a function of exposure condition. As exposure to the whole, unaltered song at study increased, so did the probability that participants reported experiencing a sense of déjà entendu. Error bars represent the standard error of the mean.

Identification Errors and Déjà Entendu

Experiment 1 demonstrated an association between commission errors and déjà vu states, such that participants were more likely to report experiencing déjà vu on trials associated with commission errors than to report non-déjà vu, which suggests that the presence of a retrieval-failure-based metacognitive state might encourage increased

internally memory search. A similar pattern was found in the current experiment, such that trials labeled as a commission error were more likely to be associated with déjà entendu reports (M = .08, SD = .13) compared to non-déjà entendu reports (M = .01, SD = .03), $t(141) = 7.00$, SE = .01, $p < .001$, d = .66, $BF_{10} = 7.24 \times 10^7$. When limiting this analysis to include only participants who did indeed make a commission error (41 participants did not make any commission errors during the test phase, and were therefore lost from this analysis), the effect became larger. For participants who at some point made a commission error during the experiment, these commission errors were more likely to be associated with a sense of déjà entendu (M = .11, SD = .14) compared to non-déjà entendu (M = .01, SD = .03), $t(100) = 7.54$, SE = .01, $p < .001$, d = .84, $BF_{10} = 4.03 \times 10^8$.

Feelings of Familiarity and Déjà Entendu

Prior research examining the déjà entendu phenomenon has demonstrated that participants tend to provide higher familiarity ratings while experiencing déjà entendu compared to when they are not (e.g., McNeely-White & Cleary, 2019). Indeed, a similar pattern was also found in the current study, such that, during retrieval failure, participants provided significantly higher familiarity ratings while experiencing déjà entendu (M = 6.73, SD = 1.15) compared to non-déjà entendu (M = 1.95, SD = 1.43), $t(141) = 36.64$, SE = .13, $p < .001$, d = 3.67, $BF_{10} = 1.76 \times 10^{70}$. Further, when analyzing the familiarity ratings provided during retrieval failure and reported déjà entendu as a function of exposure condition (Exposure0X, Exposure1X, Exposure3X), a one-way repeated-measures ANOVA revealed a significant effect, $F(2, 270) = 34.74$, MSE = .62, $p < .001$, n_p^2 = .21, $BF_{10} = 1.68 \times 10^{11}$ (see Figure 22 below). When participants were

experiencing retrieval failure and reported having a sense of déjà entendu, they provided significantly higher familiarity ratings for isolated tonal test sequences that corresponded to a whole, unaltered song presented during the study phase ($M = 6.79$, $SD = 1.41$) compared to those that did not ($M = 6.25$, $SD = 1.30$), $t(135) = 5.00$, $SE = .11$, $p < .001$, $d = .40$, $BF_{10} = 7494.17$. Additionally, compared to when the isolated tonal test sequence corresponded to a whole, unaltered song presented at study only once, participants provided significantly higher familiarity ratings to isolated tonal test sequences that corresponded to a whole, unaltered song presented three times at study ($M = 7.03$, $SD = 1.19$), $t(135) = 2.77$, $SE = .09$, $p = .01$, $d = .18$, $BF_{10} = 3.71$.

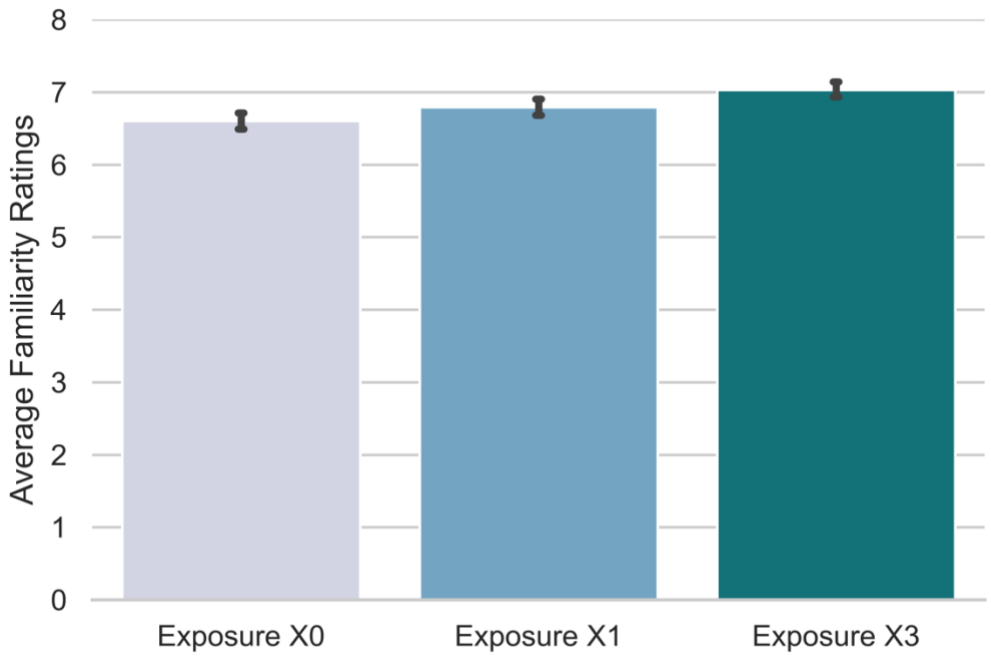

Figure 22. Average familiarity ratings provided during retrieval failure and reported déjà entendu states as a function of exposure condition. As exposure to the whole, unaltered study song increased, so did the average familiarity ratings provided during déjà entendu states. Error bars represent the standard error of the mean.

Feelings of Curiosity

Turning now to the data of primary interest, I now discuss the results on participants' curiosity ratings. When participants were presented with isolated tonal test sequences, regardless of identification status, they provided significantly higher curiosity ratings for those that corresponded to whole, unaltered songs presented during the study phase (M = 4.47, SD = 1.99) compared to those that did not (M = 3.87, SD = 2.10), $t(141) = 8.04$, SE = .07, $p < .001$, d = .29, $BF_{10} = 2.17 \times 10^{10}$ (see Figure 23 below). These data are novel, in that they provide a potential mechanism for why some pieces of music elicit stronger feelings of curiosity than others, specifically relating to Berlyne's (1950; 1960) relative novelty proposal. Pieces of novel music corresponding to the features of a previously heard song create a situation in which participants are curious to discover why an otherwise novel situation has components that are familiar, as those familiar components probably correspond to something in memory, and having a more complete knowledge structure concerning information that one has stored is desirable.

Further, when analyzing the curiosity ratings provided to isolated tonal test sequences that did versus did not correspond to whole, unaltered songs at study when identification failed, a similar pattern also emerged. On trials corresponding to whole, unaltered songs presented at study, participants provided significantly higher curiosity ratings (M = 4.40, SD = 2.04) compared to those that did not correspond to study songs (M = 3.83, SD = 2.14), $t(141) = 8.84$, SE = .06, $p < .001$, d = .27, $BF_{10} = 1.83 \times 10^{12}$ (see Figure 25). Again, these data are in support of Berlyne's (1950; 1960) relative novelty

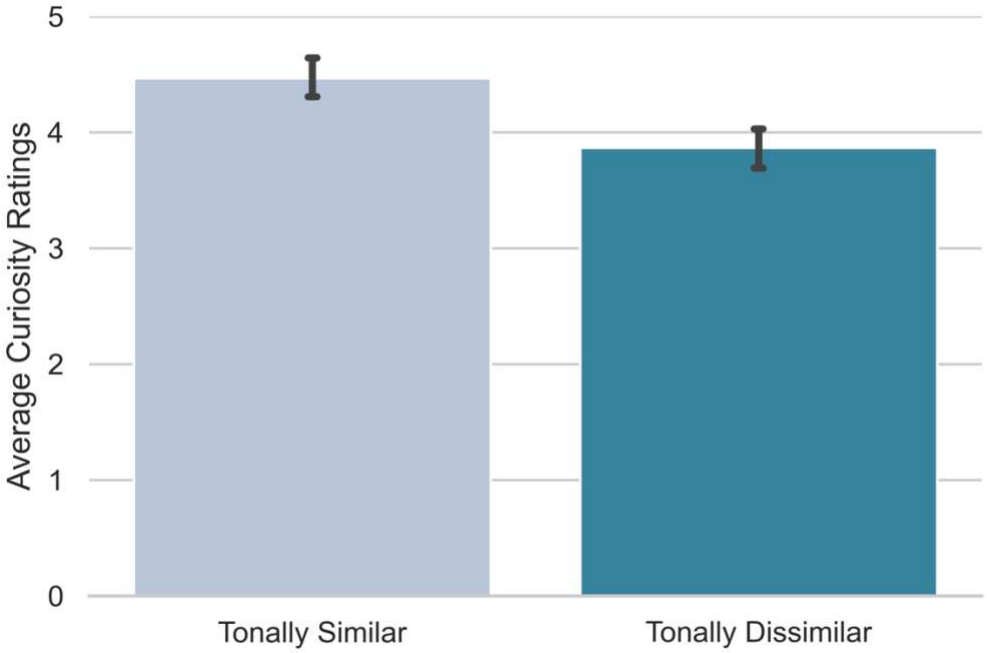

Figure 24. Average curiosity ratings provided to isolated tonal sequences that either did or did not correspond to whole, unaltered songs at study, regardless of identification status. Participants provided significantly higher curiosity ratings to those that did correspond to experimentally familiarized features compared to those that did not. Error bars represent the standard error of the mean.

proposal, such that people might feel more intense levels of curiosity when presented with a novel musical piece that corresponds to the features of a previously heard song.

Curiosity Ratings as a Function of Exposure Condition

To assess how the experimental manipulation of increasing exposure during the study phase might have affected subsequent feelings of curiosity for an isolated tonal sequence at test, regardless of identification status, a one-way repeated-measures ANOVA was conducted, revealing a significant effect, $F(2, 282) = 50.23$, $MSE = .41$, $p < .001$, $n_p^2 = .26$, $BF_{10} = 2.01 \times 10^{16}$ (see Figure 26 below). When participants heard an isolated tonal test sequence that corresponded to a whole, unaltered song presented once at study, they provided significantly higher curiosity ratings ($M = 4.30$, $SD = 2.01$)

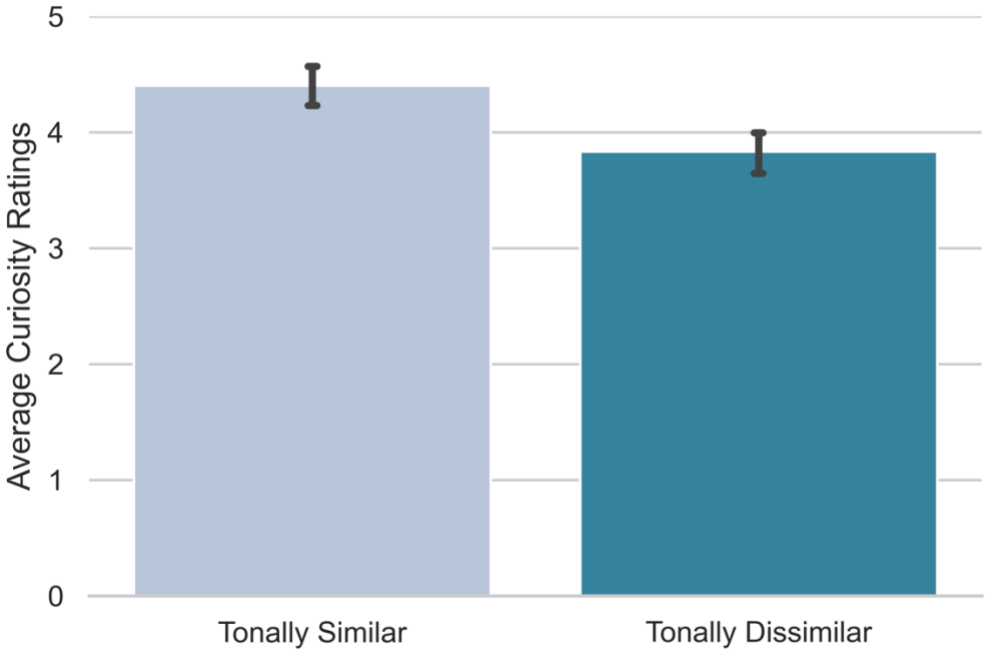

Figure 25. Average curiosity ratings provided to isolated tonal sequences that either did or did not correspond to whole, unaltered songs at study, during retrieval failure. Participants provided significantly higher curiosity ratings to those that did correspond to experimentally familiarized features compared to those that did not. Error bars represent the standard error of the mean.

compared to when they heard an isolated tonal test sequence that did not correspond to any whole, unaltered song presented at study ($M = 3.87$, $SD = 2.10$), $t(141) = 6.06$, $SE = .07$, $p < .001$, $d = .21$, $BF_{10} = 8.67 \times 10^5$. When exposure to the whole, unaltered song at study was increased to three instances, participants provided significantly higher curiosity ratings ($M = 4.63$, $SD = 2.04$) compared to when there was only one exposure instance, $t(141) = 5.01$, $SE = .06$, $p < .001$, $d = .16$, $BF_{10} = 7.82 \times 10^3$. These patterns of results suggest that, not only is curiosity heightened when listening to a novel musical piece that corresponds to the features of a previously heard song, but that the intensity of such curiosity increases when there is a stronger match between the current song's features and those previously heard.

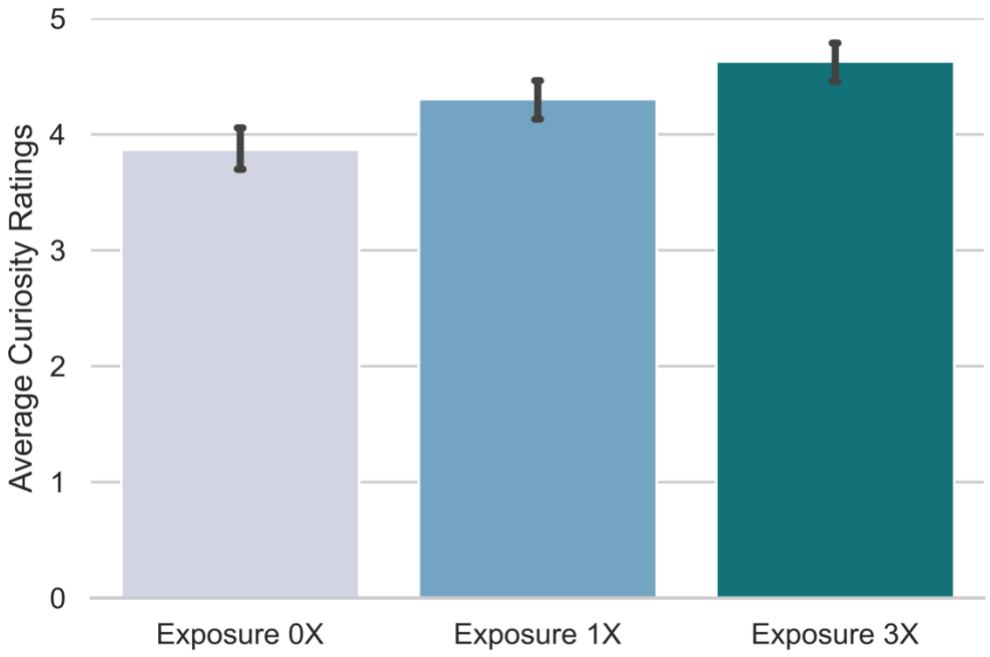

Figure 26. Average curiosity ratings provided to isolated tonal sequences as a function of exposure condition, regardless of identification status. Participants provided significantly higher curiosity ratings to isolated tonal test sequences as the prior experimental manipulation of exposure increased. Error bars represent the standard error of the mean.

In conducting the same analysis described above, but this time limiting the trials of interest to only instances of identification failure, a similar effect also emerged, $F(2, 282) = 52.98$, $MSE = .37$, $p < .001$, $n_p^2 = .27$, $BF_{10} = 1.44 \times 10^{17}$ (see Figure 27 below). When experiencing retrieval failure for an isolated tonal test sequence that did correspond to a whole, unaltered song presented once during the study phase, participants provided significantly higher curiosity ratings ($M = 4.24$, $SD = 2.07$) compared to when participants were experiencing retrieval failure for an isolated tonal sequence that did not correspond to any whole, unaltered song presented during the study phase ($M = 3.83$, $SD = 2.14$), $t(141) = 6.09$, $SE = .07$, $p < .001$, $d = .20$, $BF_{10} = 9.71 \times 10^5$. Additionally, when participants experienced retrieval failure for an isolated

tonal test sequence that corresponded to a whole, unaltered song presented three separate times at study, they provided significantly higher curiosity ratings (M = 4.58, SD = 2.09) compared to when the isolated tonal test sequence only correspond a whole, unaltered song presented once at study, $t(141) = 4.71$, $SE = .07$, $p < .001$, $d = .16$, $BF_{10} = 2.30 \times 10^3$. Collectively, these patterns of results provide support for Berlyne's (1950; 1960) relative novelty proposal, as the results suggest that a novel musical piece corresponding to previously familiarized features prompts stronger feelings of curiosity than if the whole situation were novel. Further, as exposure to those previously encountered musical features increases, which thus increases the number of stored memory traces, this creates a stronger match between the current novel musical piece's features and those stored in memory, which seems to also increase the perceived levels of curiosity. These findings also emerge when participants experience retrieval failure, in that they cannot identify the current musical piece, but they are able to sense that elements or features of the current piece correspond to information held within memory. These findings are novel, as they suggest that some forms of curiosity, specifically curiosity as a feeling-of-closeness, may emerge due to internal feature-matching processes.

Curiosity Ratings as a Function of Identification Error

The patterns of results in Experiment 1 relating to feelings of curiosity for visual stimuli suggested that the trials on which participants make commission errors are also associated with increased curiosity ratings, which may reflect an association between feelings of curiosity and increased internally-directed information-seeking behaviors. To assess whether a similar pattern would emerge in the current experiment using musical

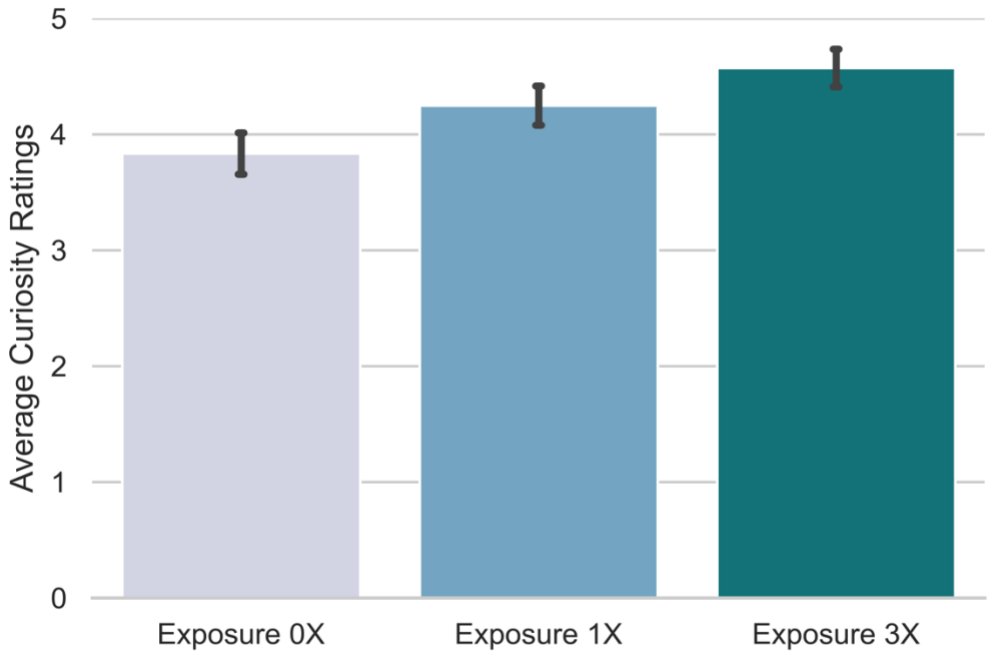

Figure 27. Average curiosity ratings provided to isolated tonal sequences as a function of exposure condition, during retrieval failure. Participants provided significantly higher curiosity ratings to isolated tonal test sequences as the prior experimental manipulation of exposure increased. Error bars represent the standard error of the mean.

stimuli, the curiosity ratings provided on trials associated with a commission error were compared with those associated with an omission error. Indeed, a similar pattern emerged, such that participants provided significantly higher curiosity ratings for trials associated with commission errors ($M = 6.68$, $SD = 2.62$) than trials associated with omission errors ($M = 4.50$, $SD = 1.97$), $t(100) = 7.99$, $SE = .27$, $p < .001$, $d = .93$, $BF_{10} = 3.34 \times 10^9$ (see Figure 28 below; note that 41 participants were lost from this analysis due to never making a commission error).

Curiosity Ratings and Déjà Entendu Reports

The patterns of results in Experiment 1 suggested that when participants report feeling a sense of déjà vu, they also provide significantly higher curiosity ratings compared to when they do not report feeling a sense of déjà vu. To assess whether a similar pattern

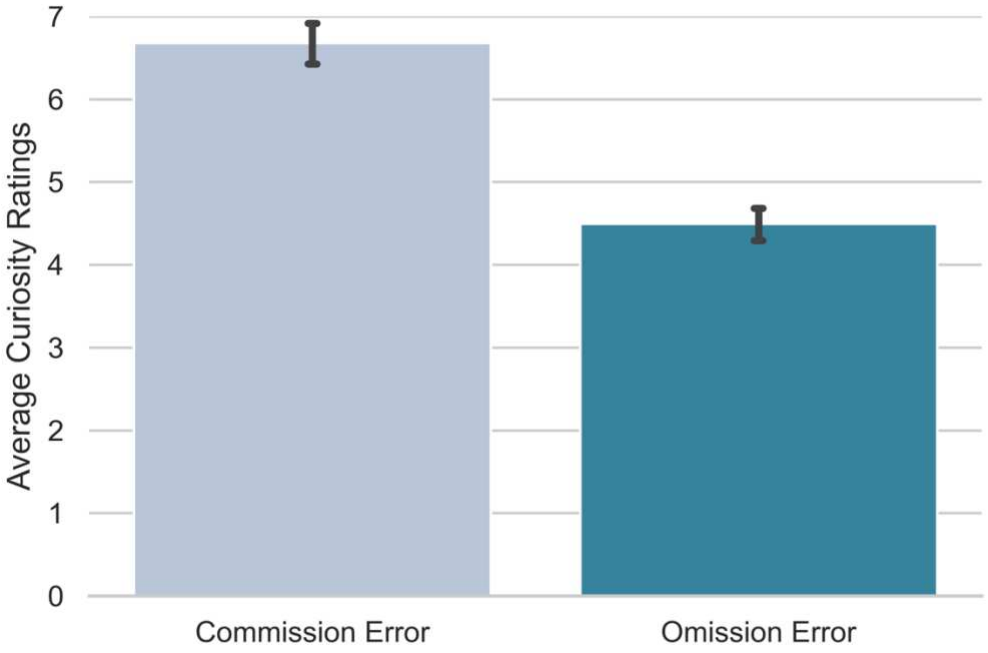

Figure 28. Average curiosity ratings provided on trials associated with either a commission or omission error. Participants provided significantly higher curiosity ratings for trials on which a commission error was made compared to those on which an omission error was made. Error bars represent the standard error of the mean.

would emerge in the current experiment, a paired-samples *t*-test was conducted, comparing the curiosity ratings provided to unidentified isolated tonal test sequences as a function of whether the participant also reported a sense of déjà entendu versus non-déjà entendu. Indeed, a similar pattern emerged. When participants reported a sense of déjà entendu for the unidentified isolated tonal test sequence, they provided significantly higher curiosity ratings ($M = 5.84$, $SD = 1.81$) compared to when they did not report a sense of déjà entendu ($M = 2.96$, $SD = 2.15$), $t(141) = 18.22$, $SE = .16$, $p < .001$, $d = 1.44$, $BF_{10} = 3.46 \times 10^{35}$ (see Figure 29 below). Further, in assessing the reaction time data, participants tended to spend more time on the curiosity prompt for trials on which they also reported a sense of déjà entendu ($M = 2314.40$ ms, $SD = 970.12$ ms) compared to trials on which they reported non-déjà entendu ($M = 1973.05$ ms, $SD =$

849.00 ms), $t(141) = 5.25$, $SE = 64.98$, $p < .001$, $d = .37$, $BF_{10} = 2.22 \times 10^4$. These patterns of results suggest that there is a relationship between feelings of curiosity and feelings of déjà entendu, and, although the directionality of the relationship has yet to be determined, perhaps the presence déjà entendu motivates a sense of curiosity and information-seeking behaviors, as evidenced by the increased time spent on the curiosity prompt during déjà entendu states.

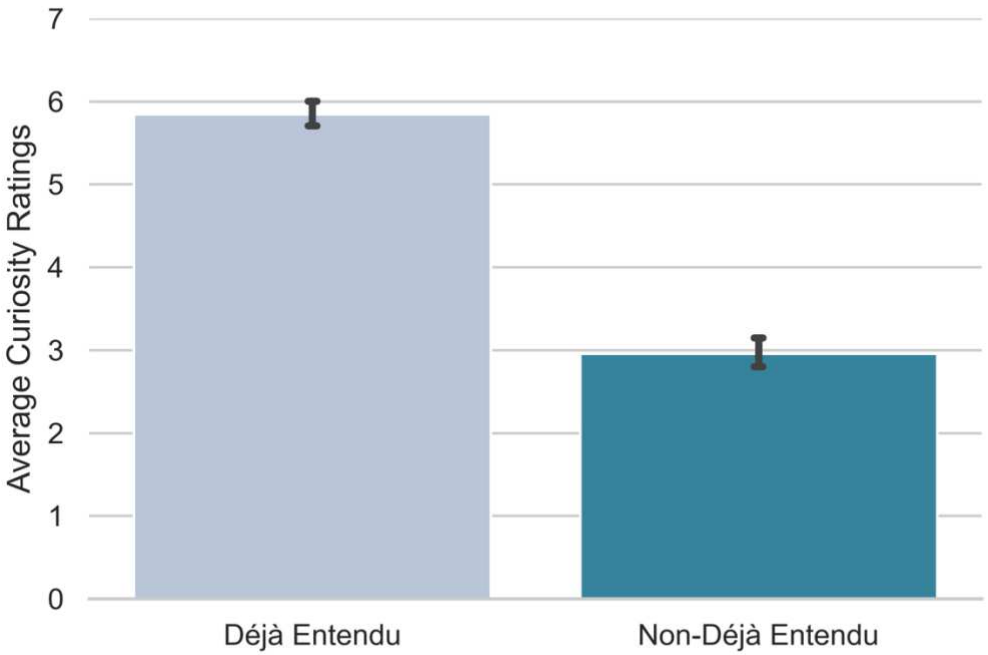

Figure 29. Average curiosity ratings provided during retrieval failure as a function of reported déjà entendu state. Participants provided significantly higher curiosity ratings for trials on which they also experienced a sense of déjà entendu compared to non-déjà entendu. Error bars represent the standard error of the mean.

As was similarly done in Experiment 1, the influences of reported Déjà Entendu State (Déjà Entendu, Non-Déjà Entendu) and Study Status (Studied, Unstudied) on feelings of curiosity during retrieval failure were assessed with a 2 x 2 repeated-measures ANOVA. A significant interaction was found, $F(1, 135) = 19.08$, $MSE = .37$, $p < .001$, $n_p^2 = .12$, $BF_{10} = 1.82$. As can be seen in Figure 30 below, a significant main

effect of Déjà Entendu State emerged, $F(1, 135) = 334.87$, $MSE = 3.05$, $p < .001$, $n_p^2 = .71$, $BF_{10} = 2.46 \times 10^{93}$, such that participants tended to provide higher curiosity ratings while experience déjà entendu compared to non-déjà entendu. When participants heard unidentified isolated tonal test sequences that did correspond to whole, unaltered songs presented during the study phase, they provided significantly higher curiosity ratings for trials on which they also experienced déjà entendu ($M = 5.19$, $SD = 1.87$) compared to non-déjà entendu ($M = 2.95$, $SD = 2.06$), $t(135) = 17.83$, $SE = .17$, $p < .001$, $d = 1.14$, $BF_{10} = 9.75 \times 10^{33}$. Similarly, when participants heard unidentified isolated tonal sequences that did *not* correspond to whole, unaltered songs presented during the study phase, they provided significantly higher curiosity ratings while experiencing a reported déjà entendu state ($M = 5.51$, $SD = 1.92$) compared to a reported non-déjà entendu state ($M = 3.00$, $SD = 2.11$), $t(135) = 16.72$, $SE = .15$, $p < .001$, $d = 1.24$, $BF_{10} = 2.52 \times 10^{31}$.

The 2 x 2 repeated-measures ANOVA also revealed a significant main effect of Study Status, $F(1, 135) = 11.44$, $MSE = .39$, $p < .001$, $n_p^2 = .08$, $BF_{10} = .17$. When participants were in a reported déjà entendu state for an unidentified isolated tonal test sequence, they provided significantly higher curiosity ratings if the isolated tonal test sequence corresponded to a whole, unaltered song presented during the study phase ($M = 5.92$, $SD = 1.87$) compared to if it did not ($M = 5.51$, $SD = 1.92$), $t(135) = 4.54$, $SE = .09$, $p < .001$, $d = .22$, $BF_{10} = 1.17 \times 10^3$. However, as can be seen in Figure 30 above, and as suggested by the significant interaction found, there was no difference in the curiosity ratings provided during reported non-déjà entendu states if they unidentified

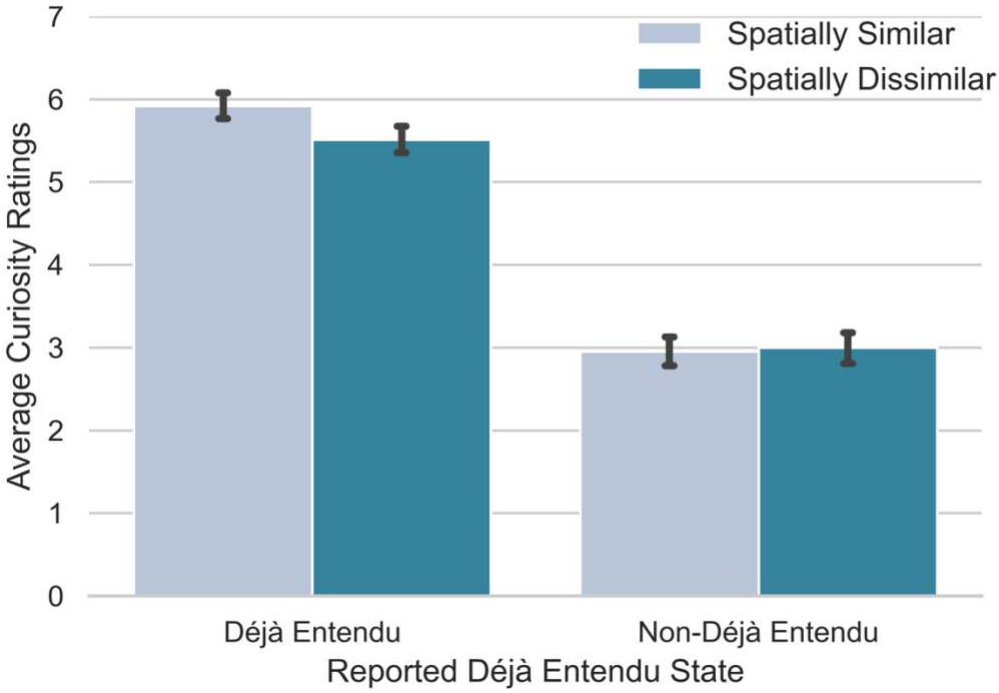

Figure 30. Average curiosity ratings provided during retrieval failure as a function of reported déjà entendu state and tonal similarity. Error bars represent the standard error of the mean.

isolated tonal test sequence did (M = 2.95, SD = 2.06) versus did not (M = 3.00, SD = 2.11) correspond to a whole, unaltered song presented during the study phase, $t(135)$ = -.85, SE = .06, p = .40, BF_{01} = 7.38. These findings are somewhat similar to the results of McNeely-White and Cleary (2019), who found that, when using *Piano Puzzlers* as test stimuli, the RWI phenomenon was contingent upon the presence of déjà entendu. Although that comparison was examining familiarity ratings and the current comparison is examining curiosity ratings, the necessity of a sense of déjà entendu to discriminate between stimuli that do versus do not musically resemble items presented at study is worth consideration.

Feelings of Curiosity and Feelings of Familiarity

To examine the relationship between participants' familiarity and curiosity ratings provided during retrieval failure, the correlation between the two ratings was computed for each participant and compared against a critical value of 0 using a one-samples *t*-test. Participants provided an average familiarity rating of 4.08 (SD = 1.70) and an average curiosity rating of 4.20 (SD = 2.04). These two measures were significantly and positively correlated, with an average correlation value of .61 (SD = .32), $t(138) = 22.73$, $p < .001$, $d = 1.93$, $BF_{10} = 1.10 \times 10^{45}$. This suggests that, as the perceived level of familiarity increases, so does the perceive level of curiosity (although note that the directionality is unknown, and it could be that as curiosity increases, so does the perceived level of familiarity).

As was done in Experiment 1, the influences of Judgement Type (Familiarity, Curiosity) and Study Status (Studied, Unstudied) on the provided rating were examined using a 2 x 2 repeated-measures ANOVA. As with Experiment 1, a significant interaction was found, $F(1, 141) = 65.14$, $MSE = .22$, $p < .001$, $n_p^2 = .32$, $BF_{10} = 4.89$. As can be seen in Figure 31 below, a significant main effect of Study Status emerged, $F(1, 141) = 271.94$, $MSE = .41$, $p < .001$, $n_p^2 = .66$, $BF_{10} = 2.42 \times 10^{17}$, such that participants tended to provide significant higher ratings for unidentified isolated tonal sequences that did versus did not correspond to whole, unaltered songs presented during the study phase. For example, when providing familiarity ratings, participants provided significantly higher ratings when the isolated tonal sequence did indeed correspond to a whole, unaltered song from study (M = 4.52, SD = 1.83) compared to if it did not (M = 3.32, SD = 1.61), $t(141) = 17.57$, $SE = .07$, $p < .001$, $d = .67$, $BF_{10} = 1.05 \times 10^{34}$.

Similarly, participants provided higher curiosity ratings to unidentified isolated tonal sequences that did (M = 4.40, SD = 2.04) versus did not (M = 3.83, SD = 2.14) correspond to whole, unaltered songs presented during the study phase, $t(141) = 8.84$, $SE = .06$, $p < .001$, $d = .27$, $BF_{10} = 1.83 \times 10^{12}$. However, like in Experiment 1, there was no significant main effect of Judgement Type (Familiarity, Curiosity), $F(1, 141) = 2.05$, $MSE = 2.73$, $p = .16$, $n_p^2 = .01$, $BF_{01} = 1.56$.

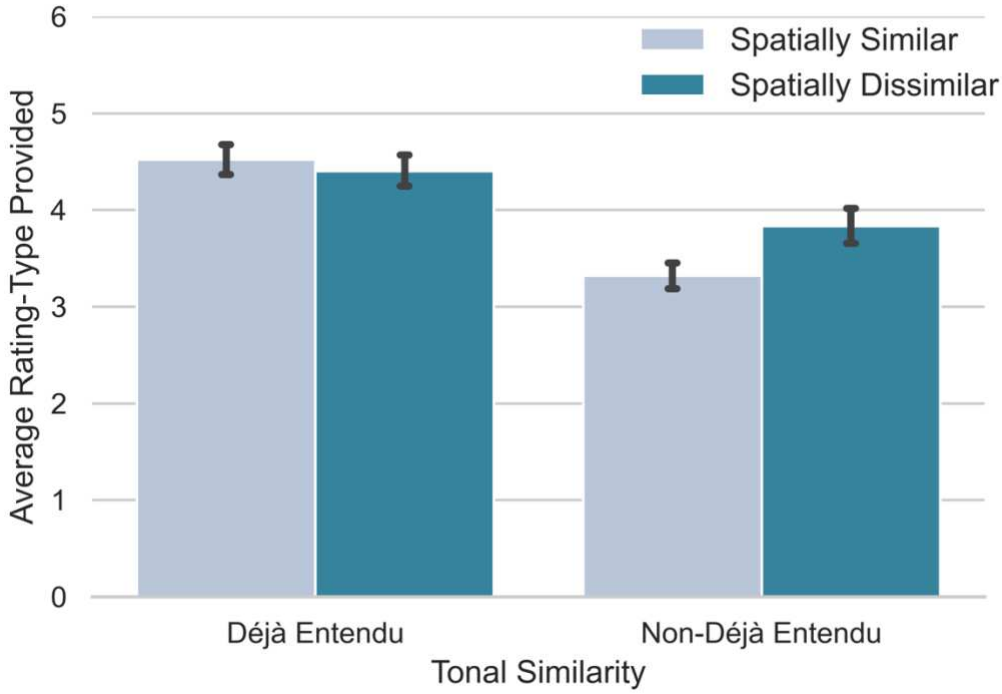

Figure 31. Average rating, either familiarity or curiosity, provided as a function of study status (studied, unstudied) and judgement type (familiarity, curiosity). Error bars represent the standard error of the mean.

Information-Seeking Behaviors

I now turn to the data concerning participants information-seeking behaviors. Overall, 56% of the participants used some but not all of their limited opportunities to hear the corresponding whole, unaltered study song (N = 79), 40% used all of their limited opportunities (N = 57), and only 4% used none of their limited opportunities

across all study-test blocks ($N = 6$; note that these six participants were necessarily excluded from the analyses reported below). On average, participants used 2.44 ($SD = 1.13$) of their limited resources on each of the seven study-test blocks[3]. As was found in Experiment 1, there were some participants who indicated "Yes, use limited resources" even when they had already expended their three opportunities on the given test block. Specifically, 73 participants (51%) indicated "Yes, use limited resources" more than three times per test block (note that, like in Experiment 1, these instances were met with the displayed text "Out of limited opportunities"). For example, one participant indicated "Yes, use limited resources" on five of the test trials despite only having a total of three opportunities to actually hear the corresponding whole, unaltered study song. These instances were included in the analyses reported below, as was done in Experiment 1.

Resource Allocation as a Function of Identification Status

To examine how participants allocated their limited resources to discover the whole, unaltered song segment to which the current isolated tonal test sequence corresponded, the proportions of trials on which participants indicated "Yes, use limited resources" that were labeled as either identification success or identification failure were computed. Like in Experiment 1, the trials on which identification succeeded was significantly more likely to be accompanied by a "Yes" response ($M = .31$, $SD = .26$) than trials on which identification failed ($M = .20$, $SD = .09$), $t(141) = 4.79$, $SE = .02$, $p <$

[3] Note that, due to the programming error described in the Methods section above, 31 of the participants received the incorrect Study Block #5, and therefore the fifth test block was excluded from all analyses for these participants. This exclusion was also taken into account when computing these participants' resource allocations. Additionally, one participant did not complete the seventh test block, and therefore did not use any of their limited opportunities on those trials.

.001, d = .36, BF_{10} = 3.24 x 10^3. Again, this may be suggesting that participants were motivated to receive confirmatory feedback on their identification attempts.

Following from this logic, the trials accompanied by a "Yes" response that were labeled as being either an instance of a commission error or an omission error were compared to determine whether participants might be using their limited resources to confirm whether their identification attempt was correct. Indeed, the trials on which participants made a commission error were significantly more likely to be associated with a "Yes, use limited resources" response (M = .43, SD = .40) compared to the trials on which an omission error was made (M = .20, SD = .08), t(100) = 5.54, SE = .04, p < .001, d = .63, BF_{10} = 5.18 x 10^4.

Resource Allocation as a Function of Study Status

The results of Experiment 1 provided support for Berlyne's (1950; 1960) relative novelty proposal, such that participants were more likely to use their limited resources to discover the corresponding information if the stimulus contained experimentally familiarized features compared to if it did not. A similar pattern was also found in the current experiment, such that, regardless of whether participants successfully identified the isolated tonal sequence or not, they were significantly more likely to indicated "Yes, use limited resources" if the isolated tonal test sequence corresponded to a whole, unaltered song presented at study (M = .24, SD = .08) compared to if it did not (M = .16, SD = .09), t(135) = 7.82, SE = .01, p < .001, d = .85, BF_{10} = 5.52 x 10^9. Further, when participants failed to identify the isolated tonal test sequence, they were significantly more likely to indicate "Yes, use limited resources" if the isolated tonal test sequence corresponded to a whole, unaltered song presented at study (M = .23, SD = .09)

compared to if it did not (M = .16, SD = .09), $t(134) = 7.79$, SE = .01, $p < .001$, d = .78, $BF_{10} = 4.49 \times 10^9$. These patterns of results provide insight into the potential mechanisms underlying information-seeking behaviors, specifically in that people may be more likely to use their limited resources to discover information about a puzzling stimulus that is otherwise novel aside from a pattern of familiarized elements.

Resource Allocation as a Function of Exposure Condition

The patterns of results examining the influence of exposure condition on participants' subjective curiosity ratings suggested that participants, during both identification success and identification failure, provided significantly higher curiosity ratings as the isolated tonal test sequence increasingly overlapped in features with whole, unaltered songs presented during the study phase. To determine whether this would also affect participants' information-seeking behaviors, the proportions of trials on which participants indicated "Yes, use limited resources" were computed as a function of exposure condition, regardless of identification success. A one-way repeated-measures ANOVA revealed a significant main effect of exposure condition on the probability that participants would indicted "Yes, use limited resources", $F(2, 270) = 43.37$, MSE = .01, $p < .001$, n_p^2 = .24, $BF_{10} = 2.85 \times 10^{14}$ (see Figure 32 below). When presented with an isolated tonal test sequence that corresponded to a whole, unaltered song presented once at study, participants were significantly more likely to use their resources (M = .21, SD = .09) than if the isolated tonal test sequence did not correspond to anything at study (M = .16, SD = .09), $t(135) = 4.85$, SE = .01, $p < .001$, d = .56, $BF_{10} = 4.04 \times 10^3$. When exposure to the whole, unaltered song was increased to three times at study, as opposed to only once, participants were significantly more likely

to indicate “Yes, use limited resources” (M = .26, SD = .11), $t(135) = 4.92$, SE = .01, $p <$ .001, d = .50, $BF_{10} = 5.23 \times 10^3$.

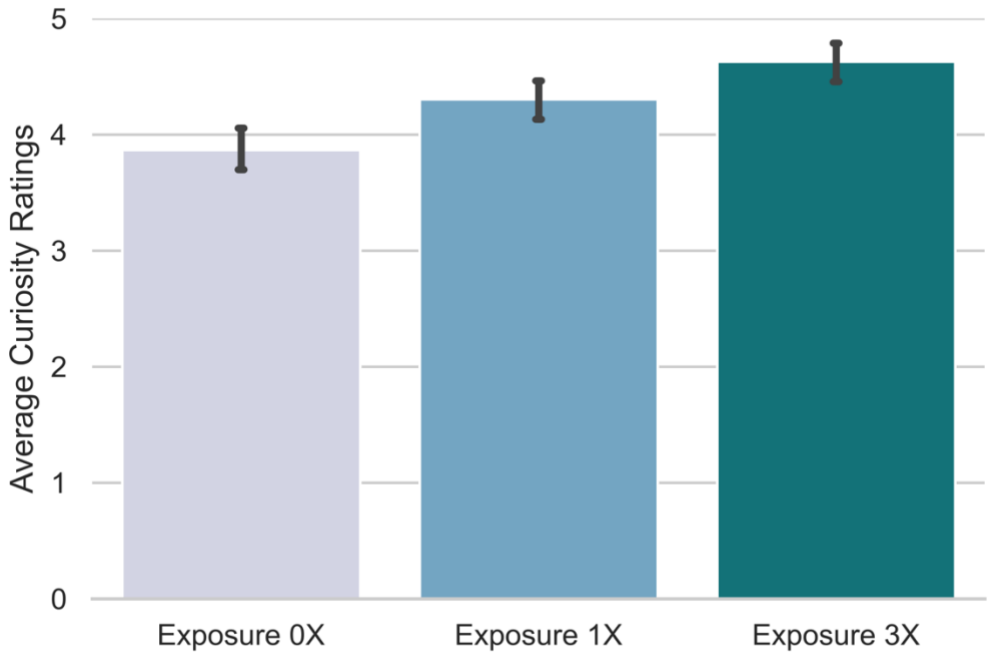

Figure 32. Probability of participants using limited resources as a function of exposure condition, regardless of identification status. Error bars represent standard error of the mean.

The analyses reported above were conducted again, but this time only on trials associated with identification *failure*. A one-way repeated-measures ANOVA revealed a significant main effect of exposure condition on the probability that participants experiencing retrieval failure would indicate “Yes, use limited resources”, $F(2, 268) = 38.27$, MSE = .01, $p <$ .001, η_p^2 = .22, $BF_{10} = 4.17 \times 10^{12}$. As can be seen in Figure 33 below, participants were significantly more likely to indicate “Yes, use limited resources” for unidentified isolated tonal test sequences that corresponded to a whole, unaltered song presented once during the study phase (M = .21, SD = .10) compared to those that did not correspond to any whole, unaltered study songs (M = .16, SD = .09), t(134)

= 4.94, *SE* = .01, $p < .001$, *d* = .53, $BF_{10} = 5.62 \times 10^3$. When exposure to the whole, unaltered song was increased from one instance to three separate instances during the study phase, participants were even more likely to indicate "Yes, use limited resources" to receive information concerning the unidentified isolated tonal test sequence (*M* = .26, *SD* = .13), *t*(134) = 4.19, *SE* = .01, $p < .001$, *d* = .43, $BF_{10} = 314.51$.

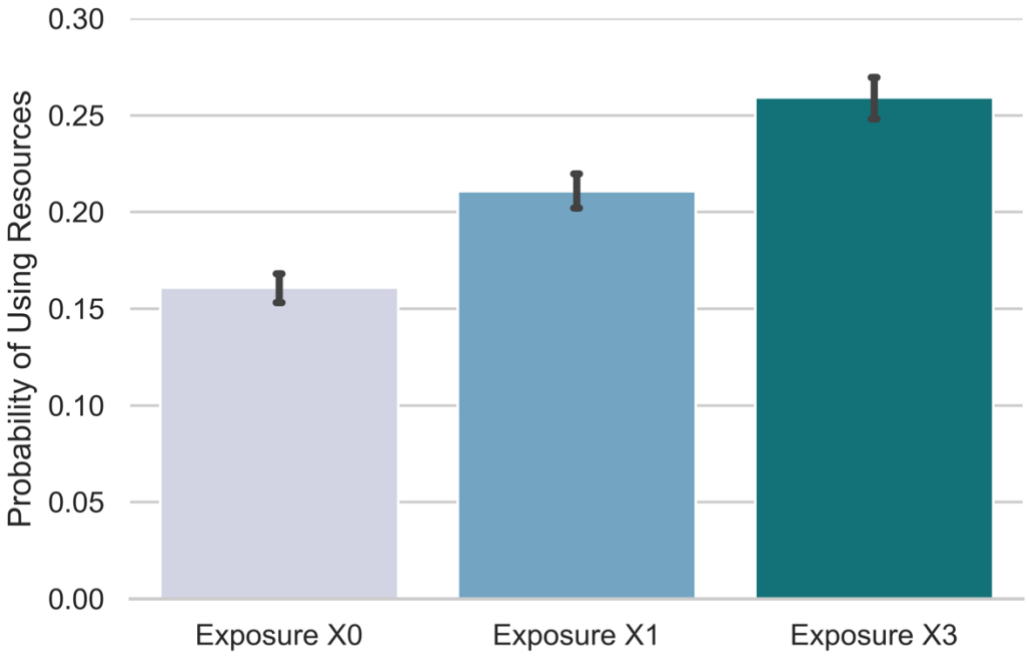

Figure 33. Probability of participants using limited resources as a function of exposure condition, during retrieval failure. Error bars represent standard error of the mean.

Resource Allocation and Curiosity Ratings

The patterns of results in Experiment 1 suggested a significant relationship between participants' resource allocation behaviors and subjective curiosity ratings, such that they provided significantly higher curiosity ratings for trials on which they decided to use their limited resources to discover information about the test scene while experience retrieval failure. To determine whether a similar pattern of behavior would emerge in the current experiment, a paired-samples *t*-test was conducted on

participants' subjective curiosity ratings provided during retrieval failure as a function of whether they decided to use or save their limited resources. Indeed, a similar pattern did emerge. Participants provided significantly higher curiosity ratings on trials they decided to expend their limited resources ($M = 6.62$, $SD = 1.92$) compared to trials on which they decided to save their limited resources ($M = 3.74$, $SD = 2.06$), $t(134) = 16.92$, $SE = .17$, $p < .001$, $d = 1.44$, $BF_{10} = 6.18 \times 10^{31}$ (see Figure 34 below).

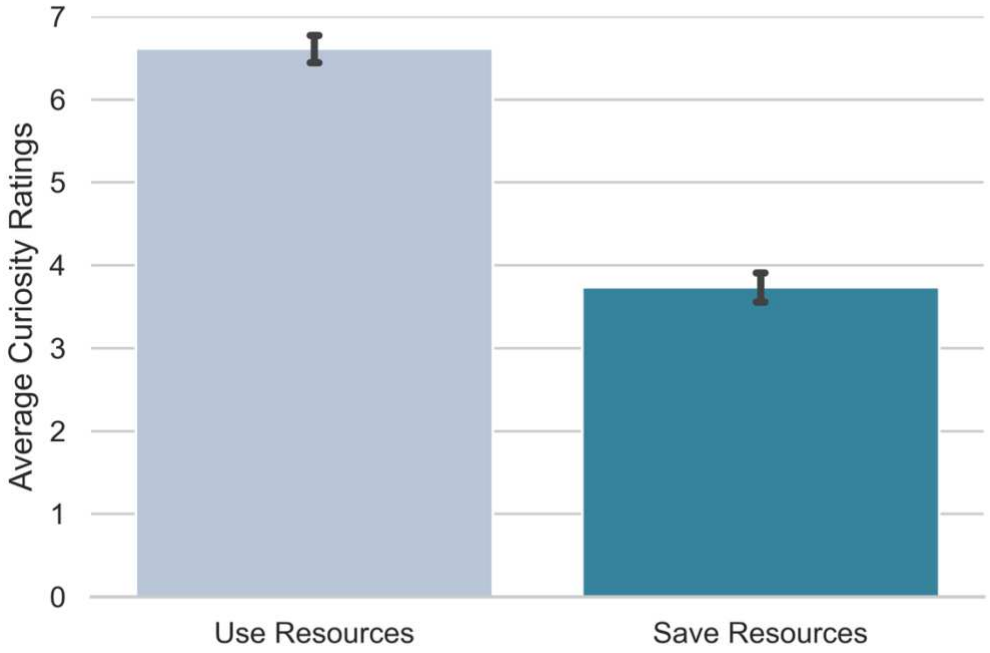

Figure 34. Participants subjective curiosity ratings provided during retrieval failure as a function of their decisions to use or save their limited resources. Error bars represent standard error of the mean.

Resources Allocation and Feelings of Familiarity

To assess how subjective feelings of familiarity correlated with participants' decisions to either use or save their limited resources, a paired-samples *t*-test was conducted. Overall, as can be seen in Figure 35 below, when participants failed to identify the isolated tonal test sequence, they provided significantly higher familiarity

ratings when they also decided to use their resources (M = 6.47, SD =1.94) compared to when they decided to save their resources (M = 3.55, SD = 1.61), $t(134) = 20.30$, SE = .14, $p < .001$, d = 1.62, $BF_{10} = 2.16 \times 10^{39}$.

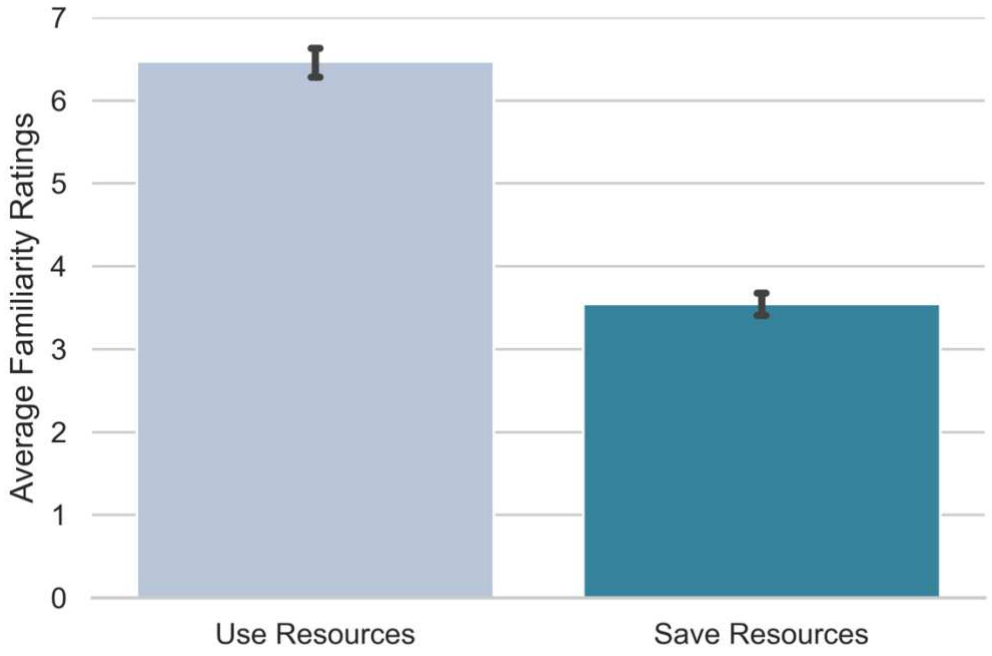

Figure 35. Participants subjective familiarity ratings provided during retrieval failure as a function of their decisions to use or save their limited resources. Error bars represent standard error of the mean.

Resource Allocation as a Function of Déjà Entendu Report

Based on the patterns of results found in the current experiment, specifically those demonstrating the significant association between participants' feelings of curiosity and déjà entendu reports, there is reason to expect that participants experiencing retrieval failure would be more likely to use their limited resources during reported déjà entendu states than non-déjà entendu states, as this would be an outward manifestation of their internal metacognitive feelings. As was shown in Experiment 1, participants were indeed more likely to use their limited resources while experiencing

déjà vu than non-déjà vu. To determine whether such an effect would emerge in the current experiment assessing curiosity for musical stimuli, a paired-samples *t*-test was conducted, comparing the probability that participants experiencing déjà entendu would indicate “Yes, use limited resources” against the probability of them saying “Yes, use limited resources” while experiencing non-déjà entendu. Indeed, a significant effect emerged, such that the probability of indicating “Yes, use limited resources” was significantly higher on trials associated with déjà entendu ($M = .39$, $SD = .16$) compared to non-déjà entendu ($M = .07$, $SD = .08$), $t(134) = 19.71$, $SE = .02$, $p < .001$, $d = 2.47$, $BF_{10} = 1.11 \times 10^{38}$ (see Figure 36 below).

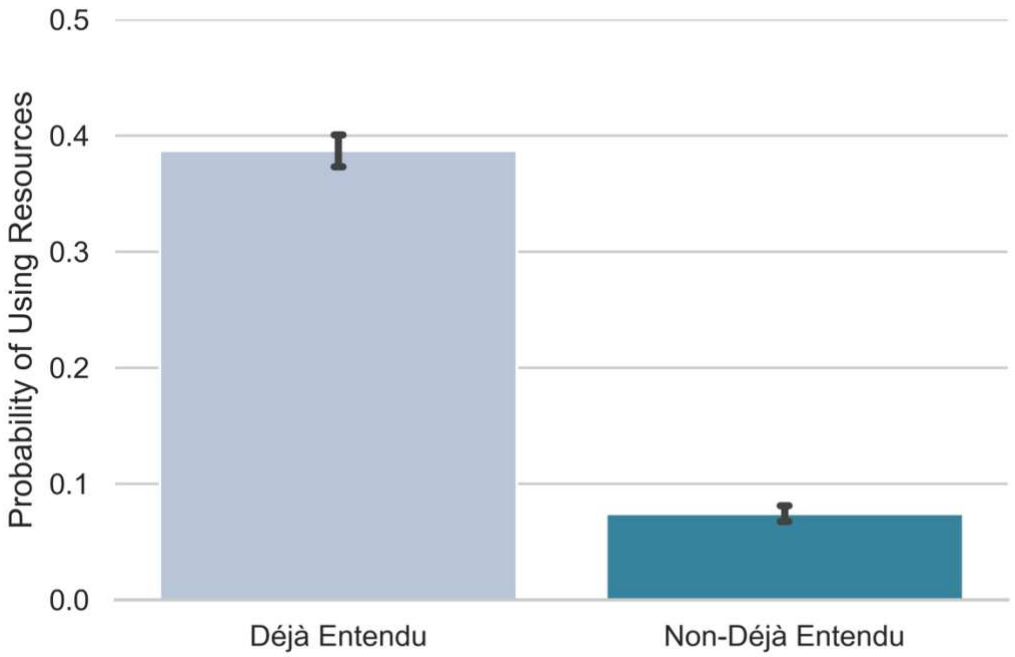

Figure 36. The average probability of participants choosing to use their limited resources as a function of reported déjà entendu state during retrieval failure. Participants were significantly more likely to use their resources during déjà entendu states than non-déjà entendu states. Error bars represent standard error of the mean.

Additionally, although not found in Experiment 1, a significant difference was found in the reaction time data. When participants were in a reported déjà entendu

state, they spent significantly less time on the Use Resources prompt (M = 2136.91 ms, SD = 979.80 ms) compared to when they were in a non-déjà entendu state (M = 2472.67 ms, SD = 1313.98 ms), $t(102) = -2.34$, SE = 143.77, p = .02, d = -.29, BF_{10} = 1.45. These patterns of results suggest that the metacognitive sensation of déjà entendu may be adaptive in that it signals to the experiencer that, although they are currently failing to retrieve relevant information concerning the current situation, they should continue searching, at least outwardly, for clues about why it feels strangely familiar.

Expanding on this proposal, the probabilities that participants would use their limited resources while experiencing déjà entendu versus non-déjà entendu were again computed, but this time limiting to instances of high curiosity (i.e., participants provided a curiosity rating of 6 or higher). Indeed, participants were significantly more likely to use their limited resources while experiencing intense levels of curiosity if they were also experiencing déjà entendu (M = .51, SD = .23) compared to non-déjà entendu (M = .24, SD = .30), $t(101) = 8.78$, SE = .03, $p < .001$, d = 1.00, $BF_{10} = 1.65 \times 10^{11}$ (note that there was no significant difference in the reaction time data). Collectively, these patterns of results support the hypothesis that internal metacognitive signals of déjà entendu and curiosity motivate the experiencer to engage in information-seeking behaviors in order to resolve the metacognitive sensations and potentially further complete their knowledge structures concerning the current situation.

Chapter 4 – General Discussion

For decades, scientific studies of metacognition and curiosity have existed, with researchers putting forth numerous theories on the mechanisms of each and providing empirical evidence intended to support those theories. However, despite hints throughout the two literatures, the metacognition and curiosity domains have remained largely separate. The current study was designed as an attempt to establish connections between these two domains, as there was reason to hypothesize that metacognitive and curiosity processes might be closely intertwined.

In one of the few formal theories proposing a connection between metacognition and curiosity, Litman (2019) proposed that curiosity might be a metacognitive signal that arises when the experiencer perceives a gap between one's current knowledge needs and the current accessible knowledge state, which in turn motivates them to resolve the gap. From a metacognitive perspective, these two states (perceiving a gap and resolving the gap) might be viewed as monitoring and control processes. However, empirical research testing this proposal had yet to be conducted. Therefore, the present study sought to further integrate the metacognition and curiosity domains by investigating whether retrieval-failure-based metacognitive states, specifically forms of familiarity-detection such as déjà vu and déjà entendu, might be associated with increased feelings of curiosity and alterations in information-seeking behaviors. The methodologies used in prior research examining déjà vu (Cleary et al., 2012; 2018) and déjà entendu (McNeely-White & Cleary, 2019; in progress) were used in the current

experiments to examine whether the mechanisms underpinning these metacognitive states might similarly drive feelings of curiosity.

The rationale behind the current experiments' hypotheses and methodologies comes from some of the current theories of curiosity (e.g., Litman, 2009; Litman & Jimerson, 2004; Lowenstein, 1994; Silvia, 2012) and empirical reports on metacognitive states (Cleary et al., 2012; 2018; 2021; McNeely-White & Cleary, 2019; Metcalfe et al., 2017; Rousseau & Kashur, 2021), providing hints that retrieval-failure-based metacognitive states may be associated with increased curiosity as a feeling-of-closeness and changes in information-seeking behaviors. Specifically, Metcalfe et al. (2017) demonstrated that TOT states, during which the experiencer is failing to retrieve the target information, are associated with increased inclinations to use limited opportunities to resolve the perceived gap in knowledge. Although the participants were experiencing momentary retrieval failure, the presence of the metacognitive signal of TOT motivated them to continue seeking information externally. However, these results, although establishing a clear connection between retrieval-failure-based metacognitive states and changes in information-seeking behaviors, do not speak to the potential mechanisms underlying feelings of curiosity, which in and of itself is a poorly understood aspect in the curiosity domain.

Indeed, as discussed by Litman (2005), there is little consensus among the curiosity researchers concerning the mechanisms underpinning feelings of curiosity. Questions such as why some stimuli prompt stronger feelings of curiosity than others, and the circumstances under which feelings of curiosity drive participants to change their behaviors and act on those feelings are poorly understood, despite the study of

curiosity having been around since the early 1900s. Although theorists have developed frameworks for explaining the phenomenology of curiosity, such as curiosity-drive theory (see Berlyne, 1966; Litman, 2005), optimal-arousal theory (see Litman, 2005; Silvia, 2012 for reviews), information-gap theory (Lowenstein, 1994), and interest-deprivation theory of curiosity (Litman & Jimerson, 2004), the empirical evidence supporting such theories has either been inconclusive or sparse. The need to understand the mechanisms and circumstances under which curiosity emerges, though, is great, with a clear example relating to education and classroom environments, as knowing how to elicit feelings of curiosity, and in turn increase students' engagement and information-seeking behaviors, could have large impacts on learning outcomes (e.g., Gruber et al., 2014; Kang et al., 2009; Murphy et al., 2021; Wade & Kidd, 2019).

Despite the large lack of understanding concerning the mechanisms of curiosity, there are hints throughout the literature suggesting that the metacognitive domain may offer a means by which to better explain feelings of curiosity. Specifically, Berlyne's (1950; 1960) relative novelty proposal that some forms of curiosity may emerge when one is presented with a novel stimulus that contains a familiar pattern of elements suggests the possibility that internal metacognitive signals may be related to curiosity. In the metacognitive domain, research has shown that, when participants are presented with a novel stimulus that contains experimentally familiarized elements, yet they fail to recall the original context in which those experimentally familiarized elements were encountered, they are more likely to find that novel stimulus familiar compared to if there were no experimentally familiarized elements. This is the recognition without identification (RWI) phenomenon (e.g., Cleary & Greene, 2000; 2001; Cleary, 2004;

Cleary et al., 2012; McNeely-White et al., 2021) and has been repeatedly demonstrated using different modalities. The RWI phenomenon is thought to occur due to feature-matching processes, such that the features or elements of the current stimulus are compared against those that are stored within memory traces representing prior experiences. As the match between the current stimulus' features and those stored within memory traces increases, so does the perceived intensity of the familiarity signal. This behaviorally manifests as participants providing higher familiarity ratings to the current stimulus, despite failing to recall details about why it feels familiar.

Based on the RWI phenomenon, Berlyne's relative novelty proposal, and the few empirical links between the curiosity and metacognitive domains (e.g., Metcalfe et al., 2017), the current study was designed to examine how feature-overlap might induce increased feelings of curiosity during retrieval failure. Further, the relationships between familiarity-detection during retrieval failure, specifically the déjà vu and déjà entendu phenomena, and curiosity as a feeling-of-closeness were examined, as this would be informative of the potential adaptive use of the déjà vu and déjà entendu phenomena, as such a relationship would suggest that these metacognitive states encourage increased curiosity and information-seeking behaviors to resolve the strange internal signals that the current situation is familiar yet simultaneously novel.

In Experiment 1, participants were presented with test lists consisting of virtual tours of novel environments, such as a reception area of a bowling alley. Half of these novel environments spatially mapped onto environments presented during the study phases while the other half did not. In addition to asking participants the typical questions on a trial-by-trial basis that have been used in prior research examining the

déjà vu phenomenon (e.g., Cleary et al., 2012; 2018), participants were also asked to rate the intensity of their perceived levels of curiosity and also whether they wanted to use their limited opportunities to potentially receive information concerning the current test stimulus.

In Experiment 2, a similar approach was taken, but as the focus was on déjà entendu states, musical stimuli were instead used, following from the methodologies of Kostic and Cleary (2009), McNeely-White and Cleary (2019; in progress), and McNeely-White et al. (2021). Specifically, participants were presented with study lists consisting of whole, unaltered well-known songs, such as *Copacabana* or *Heigh-Ho*, which were presented either zero, one, or three separate times throughout the study phase. This followed from the findings of McNeely-White et al., who demonstrated that, as exposure to features increases (and therefore the number of stored memory traces), the perceived level of familiarity during retrieval failure with the test stimulus also increases. Therefore, might this similarly affect the perceived level of curiosity? Indeed, as discussed below, as the level of experimental feature familiarization increased, so did participants' perceived levels of curiosity.

Curiosity and Featural Overlap

As proposed by Berlyne (1950; 1960), some forms of curiosity may emerge due to relative novelty, such that a novel stimulus containing familiar elements might feel strange to the experiencer, prompting them to spend more time examining the stimulus to better understand the source of the familiarity and further expand their internal knowledge structures about it. Despite this proposal emerging decades ago, little empirical work attempting to provide support for the proposal had been done until the

current study. In Experiment 1, it was found that, when participants were presented with a novel test scene that contained experimentally familiarized spatial features, regardless of identification status, they provided significantly higher curiosity ratings compared to when they were presented with a novel test scene that did not contain experimentally familiarized spatial features. It was originally hypothesized that this relationship would be dependent upon retrieval failure, with the rationale being that the failure to recall the source of the familiarity would in turn trigger more intense, perhaps frustrating signal of curiosity; however, that was not the case. During retrieval failure, there were no significant differences in the curiosity ratings provided during retrieval failure as a function of spatial similarity. The effect was only significant when considering *all* trials, during which identification may have succeeded or failed. However, it may be that there is indeed an influence of spatial similarity on curiosity ratings during retrieval failure, but this effect is masked by the way in which data are recorded. Specifically, during retrieval failure, participants might feel more curious about scenes that spatially correspond to studied scenes, but the heightened feelings of curiosity and familiarity might prompt them to search through memory more, resulting in them retrieving the target name of the corresponding scene. As discussed, participants experiencing heightened levels of curiosity and familiarity tend to make more commission errors than omission errors, suggesting that they are attempting to retrieve more often. Thus, it may be that participants initially experiencing retrieval failure for a spatially similar test scene might indeed feel heightened curiosity but are driven to internally search memory because of this signal and therefore retrieve the answer, thus resulting in the trial being an instance of retrieval success, despite retrieval initialing failing. Future research should attempt to

investigate the temporal dynamics of when familiarity and curiosity emerge in relation to when participants start attempting to retrieve the answer, as doing so will allow the separation of trials in which retrieval succeeds early on versus trials in which retrieval fails but curiosity and/or familiarity are heightened, thus motivating search and successful retrieval. Collectively, though, the finding that participants did provide higher curiosity ratings for novel test scenes containing experimentally familiarized features has large implications for the curiosity domain. Specifically, it provides a potential mechanism by which curiosity emerges, in support of Berlyne's relative novelty proposal.

In further support of Berlyne's (1950; 1960) relative novelty proposal, the findings of Experiment 2 suggest that, when participants hear an isolated tonal test sequence that corresponded to the tonal features of a whole, unaltered song presented during the study phase, they provide significantly higher curiosity ratings compared to when there was no featural overlap. When examining how increasing feature exposure (Exposure0X, Exposure1X, Exposure3X) played a role, a corresponding increase in curiosity ratings was found. As featural overlap to previously stored tonal information increased, so did participants' subjective curiosity ratings with the current isolated tonal test sequence. These patterns of results were found when examining all trials, regardless of identification status, and when examining only instances of retrieval failure.

Collectively, these patterns of results provide evidence for the potential mechanisms underpinning curiosity, specifically that it can emerge due to feature-matching processes thought to also underly familiarity-detection. When presented with a

test stimulus that contains experimentally familiarized elements, participants are able to detect these familiarized elements, even during retrieval failure, which in turn leads them to feel increased levels of curiosity.

These findings are not only in support of Berlyne's (1950; 1960) relative novelty proposal, but also of Litman's (2019) argument that we should be viewing curiosity as a form of metacognition. In Litman and Jimerson's (2004) original interest-deprivation theory of curiosity, participants experience different types of metacognitive experiences which in turn affect whether or not they experience curiosity, and if so, what type of curiosity they experience and how this affects their decision-making processes. They proposed that the presentation of a stimulus motivates the experiencer to search memory and attempt to retrieve relevant information concerning the stimulus. Based on this search, participants must then make a metacognitive judgement based on their current knowledge state. This can be either an "I know information and can retrieve it" response (*K* response), an "I don't know any information" response (*DK* response), or an "I know information but momentarily cannot retrieve it" response (*FOK* or *TOT* response). These different types of metacognitive experiences, which reflect internal processes, dictate whether participants will experience merely a desire to receive confirmatory performance feedback (*K* response), I-Type curiosity, which is curiosity as a feeling-of-interest and is a mild, pleasant form of curiosity that participants experience when they are generally interested to potentially acquire new information (*DK* response), or D-type curiosity, which Litman and Jimerson proposed is the most intense type of curiosity that emerges when the experiencer detects that they currently are deprived of access to the needed information, but their metacognitive monitoring

indicates that they *should* be able to access it, such as a TOT state. This type of curiosity motivates the most information-seeking behaviors, both internally, such as the participant using the metacognitive judgements and curiosity feelings to motivate more internal searches of memory, or externally, such as the participant turning to external knowledge sources, such as the internet.

Based on the patterns of results found in the current study, though, in addition to the work done by Rousseau and Kashur (2021) suggesting that TOT states are associated with feelings-of-closeness, and also the theory proposed by Lowenstein (1994), in which he suggested that perceived knowledge gaps vary and subsequently affect the intensity of the curiosity, I propose a modified framework (see Figure 38 below) in which these retrieval-failure-based metacognitive states do not necessarily invoke D-Type curiosity, but instead curiosity as a feeling-of-closeness (C-Type curiosity). When participants experience retrieval failure-based metacognitive sensations, such as déjà vu, déjà entendu, or TOT, they have been shown to experience the most intense levels of curiosity which in turn leads them to change their information-seeking behaviors (e.g., Metcalfe et al., 2017). When participants engage in the initial memory search and sense, perhaps through feature-matching processes, that they *do* contain relevant memory traces, but that these memory traces are momentarily inaccessible, they may still feel on the verge of access, hence the curiosity as a feeling-of-closeness. As suggested by Lowenstein (1994), when the perceived knowledge gap is very small, as in the current state of access is very close to matching the needed state of access, this is when curiosity should be the most intense. Therefore, I propose that the retrieval-failure-based metacognitive states in which one feels as if they are on

the verge of accessing relevant information, such as TOT states or déjà vu states, might be better characterized as being associated with C-Type curiosity as opposed to D-Type.

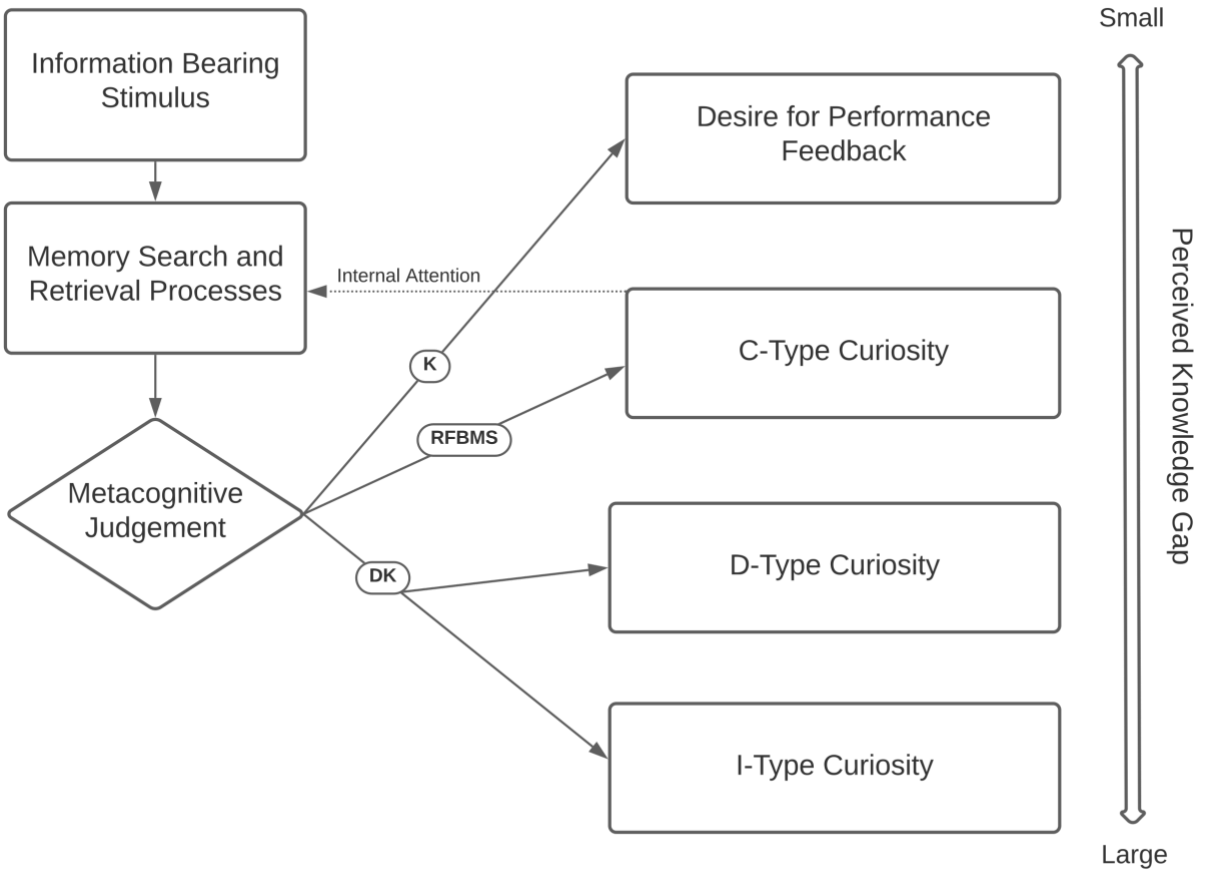

Figure 38. The Curiosity Continuum Framework. As with Litman and Jimerson's (2004) I-D model, participants' metacognitive experiences dictate their perceived curiosity experience, which in turn has an effect on their behavioral control processes. *K* = Know; *RFBMS* = Retrieval-failure-based metacognitive sensations; *DK* = Don't know.

In this modified framework, coined Curiosity Continuum Framework (CCF), one's perceived knowledge gap can range from very large to very small. Like with Litman and Jimerson's (2004) original framework, the presentation of a stimulus motivates a search of memory which in turn prompts a metacognitive judgement. By incorporating Lowenstein's (1994) perceived knowledge gap theory, I propose that the metacognitive

judgement reflects, to some extent, the participant's perceived feelings of closeness to accessing relevant information. If the perceived knowledge gap is very small, this would prompt a *K* response, in which the participant feels that they do indeed know relevant information about the current stimulus and can potentially retrieve it. While these responses also result in the participant experiencing a desire for performance feedback, unlike in the original I-D model, these responses actually reflect the highest levels of curiosity. The rationale behind this comes from the empirical findings of the current experiment. Specifically, without considering featural overlap or déjà vu state, participants in Experiment 1 provided the highest curiosity ratings when they successfully recalled the corresponding information ($M = 5.18$, $SD = 2.84$) compared to when they failed to recall the corresponding information ($M = 3.39$, $SD = 2.21$), $t(68) = 4.99$, $SE = .36$, $p < .001$, $d = .70$, $BF_{10} = 3.78 \times 10^3$. This pattern does not support the original I-D model, as Litman and Jimerson proposed that when participants provide a *K* response, these instances should be associated with the lowest intensity; however, the results of Experiment 1 do not necessarily support this (although note that the current participants were not providing *K* or *DK* responses, and prior research suggests that participants' metacognitive judgements of what they currently have access to does not always correlate with what they *can* produce from that access; Huebert et al., 2021). Further, when looking at participants' information-seeking behaviors in Experiment 1, they were significantly more likely to use their limited resources when identification succeeded than when it failed, suggesting that they are motivated to receive confirmatory feedback. Therefore, I propose that *K* responses accompanied by successful retrieval should be associated with the strongest levels of curiosity, as

participants who attempt to retrieve and produce information want to confirm that their memory systems and searches have supported their endeavors – that they are correct.

Other aspects of the CCF include *DK* responses, which can result in one of two forms of curiosity, I-Type or D-Type. I-Type curiosity is associated with the largest knowledge gaps and represent instances in which the metacognitive judgement reflects a lack of knowledge *and* a lack of need to acquire the knowledge. Instead, the experiencer might feel a general interest in learning something new, but this is not crucial for their success or survival. However, if the *DK* response is associated with a slightly smaller knowledge gap *and* a need to acquire the knowledge, then the experiencer will feel D-Type curiosity. For example, a student studying for an exam might feel D-Type curiosity for material that they are not confident about but know that they must be able to acquire the information in order to succeed on the exam.

Finally, the other aspect of the CCF encompasses retrieval-failure-based metacognitive sensations, including TOT, FOK, déjà vu, and déjà entendu. Based on the empirical findings of the current experiments in addition to prior work (e.g., Metcalfe et al., 2017; Rousseau & Kashur, 2021), retrieval-failure-based metacognitive states involve the participant feeling as if they are very close to accessing relevant information about the current situation. Although they cannot currently recall why the current stimulus feels familiar or is prompting these strange internal sensations, the presence of the metacognitive state itself is enough to indicate that they should continue searching. C-Type curiosity can in turn motivate both external searches for relevant information and internal memory searches. As discussed below, my empirical findings support such a proposal, as déjà vu and déjà entendu states are associated with significantly higher

levels of curiosity and information-seeking behaviors than non-déjà vu and non-déjà entendu states.

In consideration of the current state of the curiosity domain, my new proposed framework and empirical findings have implications for future research, specifically in that it provides a new avenue for investigating the circumstances under which curiosity emerges. Might the feature repetition effect found in Experiment 2 also affect curiosity ratings for spatial features, such that exposing participants to study scenes multiple times subsequently affect the perceived level of curiosity for the novel test scene? The patterns of results found in the current experiments also pave the way for exploring other types of features that might prompt increased levels of curiosity. The types of stimuli used to empirically examine feelings of curiosity have largely been general knowledge questions (e.g., Kang et al., 2009; Litman et al., 2005; Metcalfe et al., 2017), which are not completely expressive of curiosity in everyday experiences (see Litman, 2009). In breaking out of the general knowledge questions paradigm, the current study used visual and auditory stimuli, demonstrating rich curiosity effects. Future research should continue to use stimuli within other domains and subdomains, such as visual facial features. For example, future research should assess whether a face containing experimentally familiarized features, such as the eyes of the nose, might prompt intense levels of curiosity during retrieval failure. Additionally, how might levels of confidence in the provided recall attempts affect the form of curiosity experienced? Prior research done by Huebert et al. (2021) has shown that confidence in knowing details about the target information correspondingly increase as experimentally familiarization increases, despite the participants providing incorrect information. Might these instances also be

accompanied by the highest levels of curiosity and inclinations to receive confirmatory feedback?

Curiosity and Déjà vu Experiences

Prior research has shown that participants are significantly more likely to experience déjà vu for a novel test scene if that test scene contains spatial features that were previously familiarized during the study phase (e.g., Cleary et al., 2012; 2018). A similar mechanism has also been found to underpin the déjà entendu phenomenon, such that participants are more likely to experience déjà entendu for a song sequence if it musically corresponds to songs heard at study, either tonally or rhythmically (e.g., McNeely-White & Cleary, 2019; in progress). The current experiments also found these basic effects. Of interest to the current study, though, was whether there would be a relationship between increased levels of curiosity and déjà vu or déjà entendu states, as this would suggest a potentially adaptive function of these retrieval-failure-based metacognitive experiences (e.g., Schwartz & Cleary, 2016). In focusing on Experiment 1, which examined déjà vu states, such a relationship was found, with participants providing significantly higher curiosity ratings on trials associated with déjà vu reports compared to non-déjà vu reports. When participants reported that they felt a sense of déjà vu for the current novel test scene, they also reported experiencing more intense levels of curiosity for the novel test scene, which was a large effect (d = .89). A similar phenomenon was also found in Experiment 2, which examined déjà entendu states. When participants heard an isolated tonal test sequence, they provided significantly higher curiosity ratings if they simultaneously reported having a sense of déjà entendu compared to if they did not report having a sense of déjà entendu (d = 1.44).

In addition to the higher levels of perceived curiosity during déjà vu and déjà entendu states, there was also a relationship between these metacognitive phenomena and alterations in decision making behaviors. When experiencing retrieval failure, participants were significantly more likely to expend their limited resources on trials associated with déjà vu reports compared to non-déjà vu reports. This further supports the proposal that déjà vu may be adaptive in its function, such that it motivates the experiencer to continue searching for information when internal searches fail. Indeed, this effect is in line with what Schwartz and Cleary (2016) proposed. Retrieval failure may often be viewed or felt as a frustrating, negative experience. However, based on Experiment 1, these states may actually serve a useful purpose in one's memory system. Instead of failed memory searches being the final step, retrieval-failure-based metacognitive signals serve as a basis by which to encourage the experiencer to keep searching, as these states are indeed more likely to occur when the current situation does indeed correspond to something in memory despite it being momentarily unretrievable. The déjà vu experience may serve as an extra push to keep searching, either inwardly or externally. Indeed, an addition piece of evidence to this point can be found in the novel results pertaining to commission errors. Participants were significantly more likely to make commission errors while experiencing déjà vu compared to non-déjà vu, suggesting that they are indeed more motivated to continue internal search efforts while being in déjà vu states, albeit failed searches.

Similar patterns were also found in Experiment 2. When participants experienced retrieval failure for an isolated tonal test sequence, they were significantly more likely to use their limited resources while concurrently experiencing déjà entendu compared to

non-déjà entendu, again suggesting an adaptive function of this metacognitive state. Despite participants failing to provide accurate details retrieved from memory concerning the current test song, they were still able to strategically use their limited resources to discover information while experiencing déjà entendu compared to non-déjà entendu. Like with Experiment 1, participants were significantly more likely to make a commission error while experiencing déjà entendu than non-déjà entendu, suggesting an adaptive search-boost function.

Collectively, these patterns of results provide several theoretical implications for both the metacognitive and curiosity domains. First, as suggested by Schwartz and Cleary (2016), one adaptive function of retrieval-failure-based metacognitive states might be to encourage further information-seeking behaviors. When failing to conjure up details concerning the current stimulus, the person may still experience a cognitive signal or sensation, such as TOT or déjà vu, which indicates to them that *something* relevant is held within memory and that they should continue searching, as why would this signal emerge for truly novel, never-before-encountered stimuli? Indeed, the current results suggest that déjà vu and déjà entendu experiences, although strange-seeming in nature, are actually adaptive. While the time course information is yet unknown, it may be that déjà vu experiences prompt one to feel a sense of curiosity which in turn motivates them to continue searching for relevant information concerning the current situation in order to explain why they are experiencing these signals. Future research should endeavor to examine time course information to potentially disentangle the order of events.

Additionally, the results of the current experiment provide a support for the newly proposed CCF, such that retrieval-failure-based metacognitive states are associated with increased levels of curiosity, specifically C-Type curiosity. Despite the initial memory search failing to produce concrete, accurate details concerning the current event, the fact that there is a featural match between the stored information and current information is enough to trigger a retrieval-failure-based metacognitive sensation (i.e., déjà vu and déjà entendu). Those metacognitive signals then encourage additional internal searches of memory as to *why* one is experiencing these sensations, which is supported by the patterns of results found in both Experiments 1 and 2 concerning longer time spent on the Recall prompts when participants reported experiencing a sense of déjà vu or déjà entendu as opposed to not. This supports the proposal of the CCF, such that C-Type curiosity can prompt both internal and external searches to uncover relevant information.

Conclusions

The current experiments investigated the relationships between feelings of curiosity and metacognitive states, specifically déjà vu and déjà entendu. These experiments not only served as a means to connect these largely separated domains, but also to inform both curiosity theory concerning the mechanisms underlying curiosity and metacognition theory as to the adaptive function of retrieval-failure-based metacognitive sensations. When presented with a test stimulus, such as a novel visual environment (Experiment 1) or an isolated tonal sequence (Experiment 2), that corresponded to features presented during the study phase, participants provided significantly higher curiosity ratings, suggesting that curiosity can emerge due to featural

overlap. Additionally, participants provided the highest curiosity ratings while experiencing déjà vu or déjà entendu compared, which subsequently lead them to be more inclined to use limited resources to uncover relevant information. Overall, the results inform curiosity theory by suggesting a mechanism for curiosity and also metacognitive theory by suggesting that déjà vu and déjà entendu may be adaptive in that they encourage additional search efforts, both internal and external.

References

Arnone, M. P., & Small, R. V. (1995). Arousing and sustaining curiosity: Lessons from the ARCS model. *Proceedings of the 1995 Annual National Convention of the Association for Educational Communications and Technology (AECT), Anaheim, CA.*

Berlyne, D. E. (1950). Novelty and curiosity as determinants of exploratory behaviour. *British Journal of Psychology, 41*(1-2), 68-80.

Berlyne, D. E. (1958). The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology, 55(3)*, 289-296.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York: NY: McGraw-Hill.

Berlyne, D. E. (1962). Uncertainty and epistemic curiosity. *British Journal of Psychology, 53(1)*, 27-34.

Berlyne, D. E. (1966). Curiosity and exploration. *Science, 153*(3731), 25-33.

Brown, A. S. (2003). A review of the déjà vu experience. *Psychological Bulletin, 129,* 394-413.

Brown, A. S. (2004). *The déjà vu experience*. (Psychology Press, New York).

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*(1), 37–60.

Cleary, A. M. (2004). Orthography, phonology, and meaning: Word features that give rise to feelings of familiarity. *Psychonomic Bulletin & Review, 3*(11), 37–60. doi:10.3758/BF03196593

Cleary, A. M. (2008). Recognition memory, familiarity, and déjà vu experiences (2008). *Current Directions in Psychological Science, 17,* 353-357.

Cleary, A. M., & Brown, A. S. (2022). *The déjà vu experience* (2nd ed.). Routledge

Cleary, A. M., Brown, A. S., Sawyer, B. D., Nomi, J. S., Ajoku, A. C., & Ryals, A. J. (2012). Familiarity from the configuration of objects in 3-dimensional space and its relation to déjà vu: A virtual reality investigation. *Consciousness and Cognition, 21,* 969-975.

Cleary, A. M., & Claxton, A. B. (2018). Déjà vu: An illusion of prediction. *Psychological Science, 29,* 635-644.

Cleary, A. M., & Greene, R. L. (2000). Recognition without identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1063–1069.

Cleary, A. M., & Greene, R. L. (2001). Memory for unidentified items: Evidence for the use of letter information in familiarity processes. *Memory & Cognition*, 29, 540-545.

Cleary, A. M., Huebert, A. M., & McNeely-White, K. L. (2020). The tip-of-the-tongue state bias permeates unrelated concurrent decisions and behaviors. *Memory & Cognition, 48,* 596-606.

Cleary, A. M., Huebert, A. M., McNeely-White, K. L., & Spahr, K. S. (2019). A postdictive bias associated with déjà vu. *Psychonomic Bulletin & Review, 26*, 1433–1439. https:// doi.org/10.3758/s13423-019-01578-w.

Cleary, A. M., McNeely-White, K. L., Huebert, A. M., & Claxton, A. B. (2018). Déjà vu and the feeling of prediction: An association with familiarity strength. *Memory (Special Issue on Déjà vu*).

Cleary, A. M., McNeely-White, K. L., Russell, S. A., Huebert, A. M., & Hausman, H. (2021). The tip-of-the-tongue state as a form of access to information: Use of tip-

of-the-tongue states for strategic adaptive test-taking. *Journal of Applied Research in Memory and Cognition, 10*(1), 131-142.

Cleary, A. M., Langley, M. M., & Seiler, K. R. (2004). Recognition without picture identification: Geons as components of the pictorial memory trace. *Psychonomic Bulletin & Review, 11,* 903-908.

Cleary, A. M., Ryals, A. J., & Nomi, J. S. (2009). Can déjà vu result from similarity to a prior experience? Support for the similarity hypothesis of déjà vu. *Psychonomic Bulletin & Review, 16*, 1082–1088. doi:10.3758/PBR.16.6.1082

Cleary, A. M., Winfield, M. M., & Kostic, B. (2007). Auditory recognition without identification. *Memory & Cognition, 44(1),* 50-62.

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.

FitzGibbon, L., Lau, J. K. L., & Murayama, K. (2020). The seductive lure of curiosity: information as a motivationally salient reward. *Current Opinion in Behavioral Sciences, 35,* 21-27.

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron, 84,* 486-496.

Huebert, A. M., McNeely-White, K. L., & Cleary, A. M. (2021). Can cue familiarity during recall failure prompt illusory recollective experience? *Memory & Cognition*.

Huebert, A. M., McNeely-White, K. L., & Cleary, A. M. (under review). On the relationship between tip-of-the-tongue states and partial recollective experience:

Illusory partial recollective access during tip-of-the-tongue states. *Journal of Experimental Psychology: General.*

Kang, M. J., Hsu, M., Krajbich, I. M., Lowenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20,* 963-973.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289-325). New York, NY: Cambridge University Press.

Kostic, B., & Cleary, A. M. (2009). Song recognition without identification: When people cannot “Name That Tune” but can recognize it as familiar. *Journal of Experimental Psychology: General, 138(1),* 146-159.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126,* 159-179.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4:863*.

Lindholm, M. (2018). Promoting curiosity? Possibilities and pitfalls in science education. *Science & Education, 27*, 987-1002.

Litman, J. A. (2005). Curiosity and the pleasure of learning: Wanting and liking new information. *Cognition and Emotion, 19*, 793-814.

Litman, J. A. (2009). Curiosity and metacognition. In C. B. Larson (Ed.), *Metacognition: New Research Developments* (pp.105-116). Hauppauge, NY: Nova Science Publishers.

Litman, J. (2019). Curiosity: Nature, dimensionality, and determinants. In K. A. Reinninger & S. E. Hidi (Eds.), *The Cambridge handbook of motivation and learning* (pp. 418-442). Cambridge University Press.

Litman, J., Hutchins, T., & Russon, R. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition and Emotion, 19*, 559-582.

Litman, J. A., & Jimerson, T. L. (2004). The measurement of curiosity as a feeling-of-deprivation. *Journal of Personality Assessment, 82,* 147.

Lowenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116(1)*, 75-98.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4,* 333-369.

Mandler, G. (1980). Recognizing: the judgement of previous occurrence. *Psychological Review, 87*, 252-179.

Maw, W. H., & Maw, E. W. (1966). An attempt to measure curiosity in elementary school children. *American Educational Research Journal, 3(2)*, 147-156.

McNeely-White, K. L., & Cleary, A. M. (2019). Music recognition without identification and its relation to déjà entendu: A study using “Piano Puzzlers.” *New Ideas in Psychology, 55,* 50-57.

McNeely-White, K. L., & Cleary, A. M. (in progress). Illusory feelings of auditory prediction during déjà entendu.

McNeely-White, K. L., McNeely-White, D. G., & Cleary, A. M. (2021). Global matching in music familiarity: How musical features combine across memory traces to

increase familiarity with the whole in which they are embedded. *Journal of Memory and Language, 118,* 104217.

Metcalfe, J., Kennedy-Pyers, T., & Vuorre, M. (2021). Curiosity and the desire for agency: Wait, wait… don't tell me! *Cognitive Research: Principles and Implications, 6:69.*

Metcalfe, J., Schwartz, B. L., & Bloom, P. A. (2017). The tip-of-the-tongue state and curiosity. *Cognitive Research: Principles and Implications, 2*(31), 1-8. doi:10.1186/s41235-017-0065-4

Metcalfe, J., & Xu, J. (2017). Learning from one's own errors and those of others. *Psychonomic Bulletin & Review, 25,* 402-408.

Murphy, C., Dehmelt, V., Yonelinas, A. P., Ranganath, C., & Gruber, M. J. (2021). Temporal proximity to the elicitation of curiosity is key for enhancing memory for incidental information. *Learning & Memory, 28*(2), 34-39.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26*, 125-173.

Noordewier, M. K., & van Dijk, E. (2020). Deprivation and discovery motives determine how it feels to be curious. *Current Opinion in Behavioral Sciences, 35,* 71-76.

Peynircioğlu, Z. F. (1990). A feeling-of-recognition without identification. *Journal of Memory and Language, 29*, 493-500.

Pluck, G., & Johnson, H. (2011). Stimulating curiosity to enhance learning. *GESJ: Education Sciences and Psychology, 2*(19), 24-31.

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology, 46(2),* 168-175.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237.

Rousseau, L., & Kashur, N. (2021). Socially shared feelings of imminent recall: More Tip-of-the-Tongue states are experienced in small groups. *Frontiers in Psychology, 12:704433*, 1-15.

Ryals, A. J., & Cleary, A. M. (2012). The recognition without cued recall phenomenon: Support for a feature-matching theory over a partial recollection account. *Journal of Memory and Language, 66,* 747–762.

Schellenberg, E. G., Stalinski, S. M. & Marks, B. M. (2014). Memory for surface features of unfamiliar melodies: independent effects of changes in pitch and tempo. *Psychological Research, 78,* 84-95.

Schwartz, B. L. (2001). The relation of tip-of-the-tongue states and retrieval time. *Memory & Cognition, 29(1)*, 117-126.

Schwartz, B. L. (2002). The strategic control of retrieval during tip-of-the-tongue states. *The Korean Journal of Thinking & Problem Solving, 12(1),* 27-37.

Schwartz, B. L., & Cleary, A. M. (2016). Tip-of-the-tongue states, déjà vu experiences, and other odd metacognitive experiences. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 95–108). Oxford University Press.

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: retrieval, behavior, and experience. *Memory & Cognition, 39*, 737-749.

Schwartz, B. L., Travis, D. M., Castro, A. M., & Smith, S. M. (2000). The phenomenology of real and illusory tip-of-the-tongue states. *Memory & Cognition, 28(1)*, 18-27.

Silvia, P. J. (2012). Curiosity and motivation. In R. M. Ryan (Ed.), *The Oxford Handbook of Human Motivation* (pp. 157-166). New York, NY: Oxford University Press.

Tulving, E., & Bower, G. H. (1974). The logic of memory representations. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8). New York: Academic Press.

Wade, S., & Kidd, C. (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin & Review, 26,* 1377-1387.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*(5), 779-804.

Appendix A

In this experiment, you will see a series of video tours. During the study phase, while watching each video, you will hear a voice tell you what the scene is. For example, while viewing a golf course, the voice would say "This is a golf course. Golf course." Simply watch each scene and try to remember it. When the first set of video tours is over, you will be presented with the test phase.

During the test phase, you will view short virtual tours of scenes again, but this time, they will be new ones not seen in the previous study phase. Some of the virtual tours will resemble scenes from the study phase while others will not. After you view each tour, you will be asked several questions about it. First, you will be asked if the test scene prompted you to feel a sense of déjà vu (the feeling of having been someplace or done something before, without being able to pinpoint why, and despite knowing that the current situation is new). You will indicate Yes or No for this question. Next, you will be asked if the test scene feels familiar to you on a scale of 0 (Not familiar at all) to 10 (Extremely familiar). You will then be asked to rate how curious you are about the scene of a scale of 0 (Not at all curious) to 10 (Extremely curious). Following from this, you will be asked to indicate if you can think of a scene from the study phase that reminds you of the current test scene, and if so, type in the name of that scene. Finally, you will be asked to indicate whether or not you would like to use your limited opportunities to see information concerning the current test scene (Y=Yes, N=No). Specifically, you can request to see the study scene, if any, that corresponds to the current test scene in addition to seeing the name of the scene. However, know that you can only request to receive this information on 20% of the trials.

Appendix B

In this experiment, you will hear a list of full song clips. This is the study segment of the experiment. As you listen to the songs, you should try your best to identify and remember each of them. When the study segment is over, you will be presented with a memory test.

During the test segment, you will hear the isolated tonal sequences of songs. Some of these isolated tonal sequences will have had their whole, unaltered versions presented at study while others will not have. After each isolated tonal sequence is presented, you will be asked whether you are experiencing déjà entendu, the feeling of having heard something before despite knowing that you have not heard it before (Y=Yes, N=No). Afterwards, you will be asked to judge how familiar the isolated tonal sequence seems to you on a scale of 0 (Not at all familiar) to 10 (Extremely familiar), with the idea being that if you had just heard the whole, unaltered version of this test song clip presented at study, the test song clip will seem more familiar to you. Additionally, you will be asked to rate how curious you feel about the isolated tonal sequence on a scale of 0 (Not at all curious) to 10 (Extremely curious). You will then be asked if you can identify the name of the isolated tonal sequence, if you can. Finally, you will be asked to decide if you would like to use your limited opportunities to receive information concerning the current isolated tonal sequence. Specifically, you can request to hear the whole, unaltered version of the current isolated tonal sequence in addition to seeing the name of the song. However, you can only request to receive this information on 25% of the test trials.