Metacognitive Aspects of Face Identification

Nicholas Watier

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Ph.D degree in Psychology

School of Psychology
Faculty of Social Sciences
University of Ottawa

Acknowledgements

I could not have completed this project, or graduate school for that matter, without the encouragement, guidance, and care from a core group of people. First and foremost, my parents have provided tremendous support throughout my graduate studies. From baking pancakes in bulk to providing methods of transportation to and from Ottawa, words do little to convey how thankful I am. My beloved Ashley and our marvelous bull terrier Gus have consistently warmed by heart, and taught me to appreciate the little things. I am deeply indebted to my supervisor Charles for giving me the freedom to explore and the safety net when I fall. He is the best supervisor a student could ever ask for. Finally, I'd like to thank Claude, for introducing me to the problem of induction and giving me the opportunity to become aware of my infinite ignorance.

Statement of Candidate Contribution

The following journal articles and conference presentation have been adapted from the experiments reported in this dissertation:

Watier, N., & Collin, C. (2012). The effects of distinctiveness on memory and metamemory for face-name associations. *Memory*, *20*(1), 73-88.

Watier, N.N., & Collin, C.A. (2011). Metamemory for Faces, Names, and Common Nouns. *Acta Psychologica*, *138*(1), 143-154.

Watier, N.N., & Collin, C.A. (June, 2010). Metamemory for Faces, Names, and Words. Poster Presented at the annual meeting of the *Canadian Society for Brain, Behaviour and Cognitive Science*, Halifax, Nova Scotia.

Watier solely wrote the text of the dissertation. For all of the experiments, Watier developed the hypotheses and experimental design, programmed the experiments, and collected and analyzed the data. Collin aided in a general and supervisory manner, offering feedback and assistance in experimental design and manuscript preparation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AFC: alternative forced-choice – A format for a memory test in which a target is placed among a set of foils and participants are forced to select one of the items, even if they have to guess.

EOL: ease-of-learning judgement – A prospective memory monitoring judgement in which participants are instructed to rate the difficulty of learning an item.

FOK: feeling-of-knowing – A prospective memory monitoring judgement in which participants are instructed to rate their confidence in their ability to recognize an item. FOK judgements are typically solicited for items that could not be recalled during a cued-recall test.

FRU: face recognition unit – A theoretical component of models of person identification. The unit stores the unique visual properties of a face, which enables the face to be distinguished from others.

IAC: interactive activation and competition – A connectionist network that is used to model person identification.

JOL: judgement-of-learning – A prospective memory monitoring judgement in which participants are instructed to rate their confidence in their ability to retrieve a recently learned item at some point in the future.

NRU: name recognition unit – A theoretical component of models of person identification. Analogous to a FRU, the unit stores inputs that are not tied to any font or size, and allows for a name to be distinguished from others.

PIN: personal identity node – A theoretical component of models of person identification. The unit controls access to semantic information and signals the identity of an individual.

RCJ: retrospective confidence judgement – A retrospective memory monitoring judgement in which participants are instructed to rate their confidence in the accuracy of their recognition response.

SIU: semantic information unit – A theoretical component of models of person identification. The unit contains identity-related semantic information.

TOT: tip-of-tongue – A subjective state where a participant feels as though a currently non-recallable item is on the verge of imminent retrieval.

Abstract

To date, relatively little research has investigated participants' ability to monitor their memory for faces and names. Four experiments were conducted with aim of developing a comprehensive profile of memory monitoring performance during face identification tasks. In each experiment, memory monitoring judgements were solicited during encoding and/or retrieval of unfamiliar face-name pairs. In general, subjective estimates of future and past memory performance were valid predictors of objective memory performance, regardless of whether a face or name was the item to be retrieved from memory. As a test of the stability of memory monitoring accuracy across different categories of stimuli, memory monitoring for face-name pairs was compared with noun-noun pairs. The predictive validity of estimates of future memory performance was similar across the categories of stimuli, but the predictive validity of estimates of past memory performance was superior for nouns compared with names.

A subset of the studies examined the influence of face and name distinctiveness on memory and memory monitoring for face-name associations. This was done in an attempt to identify sources of information that individuals might use to monitor their memory during face-name learning. The beneficial effects of distinctiveness on associative memory were symmetrical between faces and names, such that relative to their typical counterparts, distinct faces enhanced memory for names, and distinct names enhanced memory for faces. These effects were also apparent in memory monitoring. Estimates of future and past memory performance were greater for face-name associations that contained a distinct face or name compared with a typical face or name, regardless of whether the distinct item was a cue or target. Moreover, the predictive

validity of prospective monitoring improved with name distinctiveness, whereas the predictive validity of retrospective monitoring improved with facial distinctiveness. Altogether, the results of the dissertation indicate that participants can monitor their memory for faces and names at a level above chance, that retrospective metamemory is more accurate for nouns compared with names, and that distinctiveness not only affects the strength of the association between a face and a name, but also the ability to monitor that association.

Introduction

A remarkable aspect of cognition is that its content can extend onto itself – that is, individuals are capable of thinking about their own thoughts. Metacognition is the term that is used to refer to cognitions about cognition. Researchers in metacognition investigate how individuals reflect on their mental processes. Metamemory is the subfield of metacognition that is concerned with how individuals evaluate their memory. This dissertation is situated in the problem space of metamemory.

There are two fundamental aspects of metamemory: monitoring, which involves reflecting on the content and processes of memory, and control, which alters the content or processes of memory depending on the output of monitoring (Koriat, Ma'ayan, & Nussinson, 2006; Nelson & Narens, 1990). By and large, metamemory studies rely on general knowledge questions or noun-noun paired-associates to probe monitoring and control; consequently, their findings might not generalize to tasks that involve learning and retrieving items outside of the semantic realm. Indeed, metamemory researchers have acknowledged the need to extend the field into other areas of cognition (Diana & Reder, 2004; Nelson & Narens, 1994).

One area that could potentially benefit from exploring the contributions of metamemory research is face identification. Face identification is an important perceptual and cognitive capacity for humans, and as such it has received a great deal of attention in the literature. While numerous studies have investigated how face recognition is influenced by encoding and retrieval operations, (for a review see Coin & Tiberghien, 1997) spatial frequency content, (for a review, see Ruiz-Soler & Beltran, 2006), stimulus orientation, (for a review, see Valentine, 1988), configural/holistic processing, (for a review, see Maurer, Le Grand, & Mondloch, 2002), familiarity, (for a review, see Burton, Bruce & Hancock, 1999), and verbalization, (for a review, see Schooler, 2002) relatively little attention has been given to participants' knowledge of their face identification abilities. For instance, individuals colloquially describe their face identification self-efficacy as "good with faces, bad with names" or vice-versa, but it remains to be objectively determined whether individuals can accurately monitor their memory for faces and names or if monitoring performance differs between faces and names.

The purpose of this dissertation is to address this gap in the literature and examine the metacognitive aspects of face identification. In particular, the primary aim is to determine whether participants are capable of accurately monitoring their memory for face-name associations at various stages of learning and whether metamemory findings obtained with traditional stimuli (i.e. noun-noun paired-associates) generalize to faces and names. A secondary aim of this dissertation is to identify sources of information that individuals might use to monitor their memory for face-name associations. The relative distinctiveness of a face or name is likely to be one such source. While the memory enhancing effects of distinctiveness are commonly reported in the face identification literature (e.g. Brandt, Gardiner, & Macrae, 2006; Brown & Lloyd-Jones, 2006; Cohen & Carr, 1975; Courtois & Mueller, 1981; Ellis, Shepherd, Giblin, & Shepherd, 1988; Light, Kayra-Stuart, & Hollander, 1979; Semmler & Brewer, 2006; Stanhope & Cohen, 1993; Valentine & Bruce, 1986; Valentine & Moore, 1995; Winograd, 1981), it is not yet known, however, if individuals are aware of these effects or if they consider these effects when reflecting on their memory for face-name associations. Thus this dissertation also examines the influence of distinctiveness on metamemory for faces and names.

There are practical and theoretical reasons for investigating the metacognitive aspects of face identification. The results would provide insight into the strategies that individuals use to identify a face or a name and could be used to constrain theoretical models of person identification. This in turn has practical implications for eyewitness testimony or localized memory disorders such as prosopagnosia. For example, if healthy adults can accurately predict their face identification performance, then it might be the case that the spectrum of person identification disorders (e.g. prosopagnosia, Capgras syndrome, Fregoli syndrome) is related to impairments in monitoring and control. Ultimately, a richer understanding of face identification can be established by examining the metacognitive processes that accompany the task.

An outline of the dissertation is as follows. The first chapter reviews core concepts in metacognitive research. Various operational definitions of metacognition are included in order to provide a broad context of the research described in the dissertation. Monitoring and control mechanisms are then briefly discussed, followed by a theoretical framework of metamemory. The chapter concludes with a rationale for applying the framework to face identification tasks.

The second chapter provides an overview of four monitoring judgements that are commonly used in metamemory research. The chapter is organized into four sections, with each section focusing exclusively on one type of judgement. Within each section, the typical procedure for soliciting the judgement is reviewed, followed by the relevant factors that can influence the judgement, and the possible mechanisms underlying the judgement. Each section concludes with a review of the studies that solicited the monitoring judgement using faces or face-name pairs as stimuli.

The third chapter reviews pertinent findings in the face identification literature, with an emphasis on models of person identification, distinctiveness effects in faces and names, and errors in personal naming. Distinctiveness effects are given special emphasis throughout the chapter because the potential influence of facial distinctiveness on metamemory accuracy has theoretical and applied significance, especially in the realm of eyewitness identification.

Chapter four describes the general procedure that is shared among all of the experiments reported in the dissertation. This chapter also explicates the measures of monitoring accuracy that are used to assess metamemory performance. Chapters five and six outline the rationale, methodological details, and results of four experiments that examine metamemory performance during face identification tasks. A brief discussion follows each experiment.

The seventh and final chapter is a general discussion of the findings of the dissertation. The first half of the chapter re-iterates the rationale and relates the major results of the experiments to theoretical and applied issues in face identification and metamemory research. The second half of the chapter adopts a narrower focus, and discusses the idiosyncratic findings of each experiment in more detail. The chapter ends with suggestions for future research and summarizes the conclusions of the dissertation.

Chapter 1 – Metacognition

The ability to reflect on our cognitive processes is a remarkable phenomenon, one that has fascinated psychologists for over a century (e.g. William James (1890) remarked that introspection is the cornerstone for measuring psychological events). However, systemic research on the subjective awareness of cognitive processes only gained prominence in 1970s, following the paradigmatic revolution from behaviorism to cognitivism (for a history of metacognition, see Hacker, 1998). It was during this period that Flavell (1979) coined the term metacognition to describe the development of children's knowledge about the mind. Specifically, metacognition was defined as "knowledge and cognitions about cognitive phenomena" (pg. 906). The term has since been applied to encompass a variety of cognitive processes, including, metacomprehension (Maki & Berry, 1984), metaemotion (Gottman, Katz, & Hooven, 1996) metamemory (Nelson & Narens, 1990), visual metacognition (Levin, 2004), and spatial metamemory (Schwartz, 2006).

Needless to say, metacognition is a broad category and has been operationally defined in many ways. For instance, Kluwe (1982) argues that metacognition not only involves a person's reflection of their own thinking but the thinking of others as well, whereas Schooler (2002) considers metacognition to be one aspect of metaconsciousness, which is a re-representation of conscious events. Others have distinguished between awareness of on-going cognitions and awareness of how a cognitive process operates. Flavell (1979) refers to the latter as metacognitive knowledge and the former as metacognitive experience. For example, knowing that our memory for names is poor would be considered metacognitive knowledge, whereas feeling that a face is familiar and that it warrants an effortful search for a corresponding name would be considered metacognitive experience. Likewise, Cavanaugh (1989) distinguishes between three subtypes of metacognitive awareness: systemic awareness, which involves knowledge about of how a cognitive system operates; epistemic awareness, which involves knowledge about the contents of our memory; and on-line awareness, which involves knowledge about an on-going cognitive process.

Along similar lines, Koriat and colleagues (Koriat & Levy-Sadot, 1999; Koriat, 2007) distinguish between experienced-based and theory-based metacognition. In the former, metacognitive judgements are thought to reflect automatic and unconscious applications of heuristics such as encoding fluency (e.g. Benjamin & Bjork, 1996; Koriat & Ma'ayan, 2005), retrieval fluency (e.g. Benjamin & Bjork, 1996; Kelley & Lindsay, 1993), cue familiarity (e.g. Koriat & Levy-Sadot, 2001; Metcalfe & Finn, 2008) or accessibility of information relating to the item to be retrieved from memory (Koriat, 1993; 1995; Koriat & Levy-Sadot, 2001), whereas in the latter, metacognitive judgements are thought to reflect deliberate and conscious inferences based on specific a-priori knowledge of a cognitive process (e.g. psychology students know that working memory is limited to 7 ± 2 chunks), intrinsic properties of the stimuli, such as item difficulty (Dunlosky & Matvey, 2001), extrinsic properties of the learning conditions, such as practice or massed repetition (Koriat, Sheffer & Ma'ayan, 2002) or an effortful consideration of selected answers against possible alternatives in memory (Koriat & Goldsmith, 1996b; Nelson & Narens, 1990). Each of these heuristic is discussed in more detail in Chapter 2. The distinction between theory-based and experienced-based metacognition can be crudely summarized as the distinction between an educated guess derived from empirical facts and a gut feeling derived from sheer subjective experience, respectively.

Although these operational definitions differ with respect to the content of metacognition, they share several assumptions (Nelson & Narens, 1990; Nelson, 1996). First, it is assumed that metacognitive thoughts arise from an internal representation rather than an external stimulus and one role of a metacognitive system is to monitor that internal representation. Second, metacognitive judgements are presumed to indirectly measure cognitive processes and are fallible. An objective measure of participants' introspections can be obtained by comparing the accuracy of their judgements to objective performance measures. For example, a participant might think that learning a series of faces is easy and judges that he has effectively committed the faces to memory, but by comparing his metacognitive judgements with his performance on a retrieval test, the accuracy of his introspections can be determined.

A Metamemory Framework

According to the metacognitive model proposed by Nelson and Narens (1990), cognitive processes are split between two interrelated levels: an object-level, which is the cognitive process itself, and a meta-level, which contains an ever-changing representation of the object-level. Communication between levels is accomplished through monitoring and control processes. Monitoring refers to the flow of information from the object-level to the meta-level. The purpose of monitoring is to update the meta-level with regards to the state of the object-level. Control refers to flow of information from the meta-level to the object-level. The purpose of control is to modify the state of the object-level. In short, a metacognitive model contains at least two interrelated levels that communicate via monitoring and control mechanisms. Monitoring provides information about the current state of a cognitive process whereas control alters the state of a cognitive process based on the output of monitoring. A metacognitive model is presented in Figure 1.

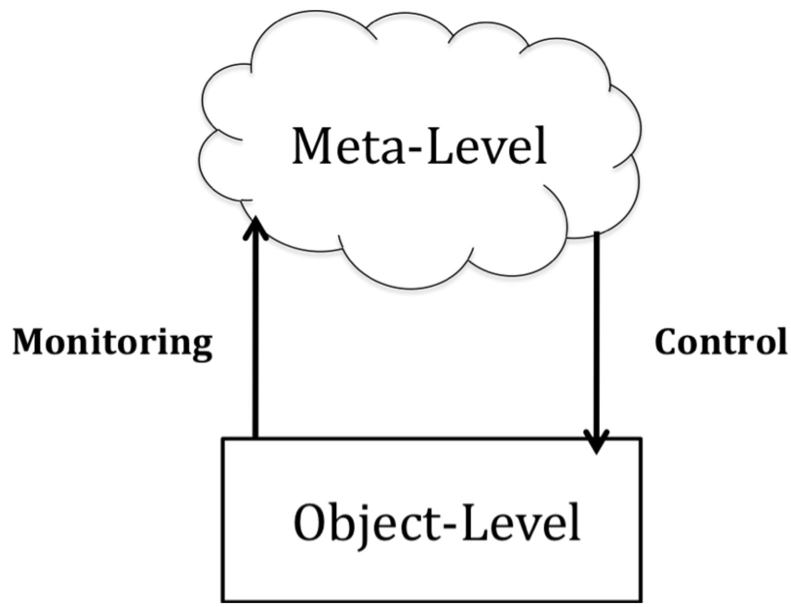

Figure 1. Formulation of Nelson and Narens (1990) metacognitive system. The meta-level develops a model of the object-level through monitoring and can alter the state of the object-level through control. The cloud represents the imperfect representation of the object-level at the meta-level. Adpated from Nelson and Narens (1990).

Nelson and Narens' (1990) model allows for an empirical evaluation of monitoring and control processes during various stages of memory. An overview of the monitoring and control processes is presented in Figure 2. Control processes can be evaluated by measuring study time, strategy selection, and duration of memory search, whereas monitoring is typically examined by soliciting prospective and retrospective confidence judgements. In particular, researchers have relied on the output of ease-of-learning judgements (e.g. Leonesio & Nelson, 1990; Underwood, 1966), judgements-of-learning (e.g. Koriat, 1997; Nelson & Dunlosky, 1991), feeling-of-knowing judgements (e.g. Koriat & Levit-Sadot, 2001; Krinsky & Nelson, 1985; Reder, 1987; 1988), and retrospective confidence judgements (e.g. Koriat & Goldmsith, 1994; 1996b) to examine the monitoring processes that occur prior to learning, immediately after learning, after a recall attempt, and after a recognition response, respectively (for reviews, see Nelson & Narens, 1990; 1994; Schwartz, 1994). These judgements are briefly described below (Chapter 2 provides a comprehensive review).

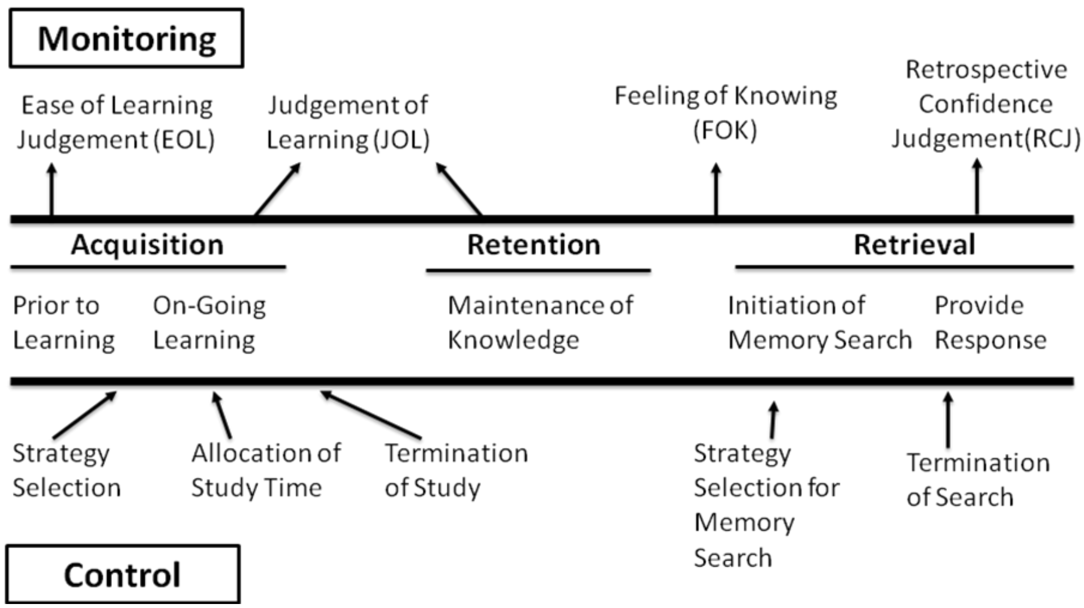

Figure 2. Overview of Nelson and Narens' (1990) framework. Monitoring components are shown above the bold horizontal lines, whereas control components are shown below. Adapted from Nelson and Narens (1990).

Ease-of-learning (EOL) judgements reflect an a-priori estimate of item difficulty. They are correlated with study time (Nelson & Leonesio, 1988), and are a significant predictor of an individual's rate of learning (Leonesio & Nelson, 1990; Underwood, 1966). Judgements-of-learning (JOL) are estimates of future retrieval performance and provide insight into the current level of mastery or whether additional study time is warranted. JOLs are a significant predictor of recall and recognition performance, especially after a delay following study (e.g. Connor, Dunlosky & Hertzog, 1997;

Dunlosky & Nelson, 1992; 1994; Keleman & Weaver, 1997; Nelson & Dunlosky, 1991; Nelson, Narens, & Dunlosky, 2004; Thiede & Dunlosky, 1994). If recall is unsuccessful, but participants nonetheless feel as though they can recognize an item, initiation and duration of a memory search is based on a feeling-of-knowing (FOK). FOK judgements are estimates of future recognition performance and are typically solicited after a recall failure. They are a significant predictor of recognition performance for items that cannot be recalled (Koriat & Levit-Sadot, 2001; Reder, 1987; 1988). Finally, after a recognition decision, a subjective assessment of the accuracy of the response is made with a retrospective confidence judgement (RCJ).

Correlations among the monitoring judgements are weak (Kelemen, Frost, & Weaver, 2000; Leonesio & Nelson, 1990), and certain neurological impairments affect performance on some monitoring tasks but not others (Modirrousta & Fellows, 2008; for a review, see Pannu & Kaszniack, 2005). Moreover, each monitoring judgement relies on different sources of information (discussed in more detail in Chapter 2). Thus, each of the judgements is considered to be distinct and taps into a different metamemory construct.

The framework has been applied to a variety of learning situations (for reviews see, Nelson & Narens, 1990; Son & Metcalfe, 2000; Van Overschelde, 2008), including those with non-verbal stimuli such as olfactory recognition (Jonsson & Olsson, 2003) and spatial navigation (Schwartz, 2006). Importantly, the framework allows for a comprehensive examination of metamemory performance during various stages of learning. For this reason, the experiments reported in this dissertation followed Nelson and Narens (1990) framework to examine metamemory for face-name associations.

Assessing the Accuracy of Monitoring Judgements

As mentioned above, one of the assumptions in metacognitive research is that introspections are imperfect measurements of internal states. Of course, this does not mean that introspection is not useful, as an imperfect monitoring device can still provide insight into behaviour, knowledge, or experience (Nelson, 1996). It is the role of the researcher to characterize the conditions in which monitoring is or is not accurate. The accuracy of monitoring is typically assessed in one of two ways. First, by calculating a correlation between the magnitude of a monitoring judgement and performance on an objective memory test (for a reviews, see Benjamin & Diaz, 2008; Nelson, 1984; Schraw, 2009), or by calculating the difference between the magnitude of a monitoring judgement and performance on an objective memory test (for a review, see Keren, 1991). The former is referred to as relative accuracy or resolution, whereas the latter is referred to as absolute accuracy or calibration.

Measures of relative accuracy are useful for examining how participants monitor items relative to each other. For example, a participant with perfect relative accuracy would always assign a high JOL (e.g. 100% confident that he will recall the item) to items that were subsequently correctly recalled and would always assign a low JOL (e.g. 20% confident that he will recall the item) to items that were subsequently incorrectly recalled. In this case, the participant is accurately monitoring his degree of mastery and can perfectly predict whether one item will be recalled relative to another item.

In contrast, measures of absolute accuracy are useful for determining whether monitoring judgements reflect actual memory performance. For example, perfect absolute accuracy occurs when items that were given a JOL of 60% have a mean recall accuracy of 60%. In this case, the observed recall performance corresponds exactly with the predicted recall performance. Importantly, absolute accuracy can be used to identify over or under confidence. For example, of the items that were given a JOL of 100%, recall accuracy might only be 70%.

The distinction between absolute and relative accuracy is not trivial. Relative accuracy establishes whether participants can distinguish between items that will or will not be remembered at some point in the future. Conversely, absolute accuracy establishes whether estimates of memory performance are consistent with actual behaviour. Absolute accuracy is often measured using either calibration curves, which depict the mean observed performance on a memory test as a function of the mean estimated performance on a memory test, collapsing across all participants in a given condition, or by calculating the signed difference between the mean estimated memory performance and the mean observed memory performance in a condition. Thus, absolute accuracy is a between-participant measure of monitoring accuracy. Interpreting absolute accuracy requires that certain assumptions be taken into consideration, some of which may be untenable (see Keren, 1991). For example, one must assume that all participants use the rating scale in an identical fashion or that all participants have the same overall level of confidence or self-efficacy in memory ability. On the other hand, relative accuracy is a within-participant measure of monitoring accuracy because the correlation between a monitoring judgement and memory performance is calculated for each participant in a given condition. Thus, for relative accuracy, it is not necessary to assume that all participants use the scale in the same manner.

It is possible for a participant in a given experimental condition to obtain perfect relative accuracy, but the absolute accuracy of the condition as a whole might reveal marked over or under confidence. Indeed, researchers have raised concerns over possible tradeoffs between absolute and relative accuracy, where maximizing one might come at the cost of attenuating the other (Keren, 1991, for examples of tradeoffs, see Koriat et al., 2002; van Overschelde & Nelson, 2006). Moreover, others have suggested that each measure reflects a different underlying metacognitive construct (Maki, Shields, Wheeler, & Zacchilli, 2005; Mengelkamp & Bannert, 2010). Ultimately, it is important to keep in mind that measures of relative and absolute accuracy are not equivalent, and should not be interpreted as such. Both types of monitoring accuracy are reported in the dissertation in order to provide a comprehensive profile of metamemory performance.

Rationale for Applying a Metamemory Framework to Face Identification Tasks

While it is interesting to study the metacognitive aspects of face identification in their own right, there are several practical and theoretical advantages. From a theoretical perspective, the subjective experience of learning a person's identity should be taken into consideration in order to have a complete account of face identification. Tulving (1989) notes that researchers in memory often make the tacit assumption that knowledge, behaviour, and experience are in perfect agreement. Unfortunately, the assumption limits the independent study of subjective experience because behaviour, knowledge and experience are erroneously equated with each other. A similar assumption is made in face identification research, where the role of subjective experience in identifying faces is largely neglected. To illustrate, discovering that distinct faces are recognized more accurately than typical faces does not provide any insight into the experience that participants have or strategies that participants use to make recognition decisions. However, by comparing the magnitude and accuracy of monitoring judgments between typical and distinct faces, one can determine whether individuals use facial distinctiveness as a cue when estimating their memory performance.

In accordance with Tulving (1989), face identification research should include traditional quantity-based measures of performance (e.g. the number of items retrieved) as well as accuracy-based measures of performance (e.g. the correspondence between our memory and actual past events). A beneficial attribute of accuracy-based measures is that they incorporate the subjective characteristics of memory retrieval and are evaluated based on the degree of correspondence between the target item and the item retrieved from memory (see Koriat & Goldsmith, 1996a). For example, in a courtroom setting the number of items that are recalled is not as significant as the accuracy and confidence in the memory for those items. By including accuracy-based measures in face identification research, a broader account of real-world behaviour can be established.

Within Nelson and Narens' (1990) framework, monitoring and control mechanisms are considered to be essential for an effective storage and retrieval system. The system must be able to monitor whether an item has been committed to memory, whether a memory search is warranted, and whether a retrieved item should be volunteered for an answer. Consequently, examining how monitoring and control mechanisms operate during face identification tasks would shed light on the conditions in which identification errors occur, which has obvious practical advantages. For instance, the accuracy of prospective monitoring judgements can be used to determine if certain task demands (e.g. a delay between studying a face and soliciting a prospective confidence judgement) or stimulus variables (e.g. a distinctive name or face) lead participants to erroneously believe that a face or name is committed to memory.

Another theoretically driven reason to examine metamemory for face-name associations is that general knowledge questions, verbal paired-associates and letter strings are predominantly used as stimuli in metamemory studies, and as a result, previous findings might not generalize to face identification paradigms. Several lines of research support this possibility.

First, several researchers have posited separate memory systems or codes for verbal and non-verbal stimuli. Tulving and Schacter's (1990) perceptual representational system (PRS) posits that perceptual memories can be stored separately from semantic memories. Likewise, Paivio's (1991) dual-coding theory (DCT) contends that visual representations are functionally and structurally independent of verbal representations. Thus, a metamemory system for words and faces could be structurally similar but functionally different – that is, the relative efficiency of monitoring and control mechanisms might vary depending on the category of stimuli.

Second, dissociations between face recognition and word recognition are not uncommon. For example, Woodhead and Baddeley (1981) found that participants who were superior at recognizing faces were average in regards to recognizing words. Moreover, faces exhibit less forgetting over a two-week interval and are less affected by retroactive interference compared to words (Deffenbacher, Car & Leu, 1981). In addition to behavioural evidence, the neural substrates underlying word and face recognition are anatomically distinct (for a recent review, see Dien, 2009) and reveal markedly different patterns of cerebral activation (Kim et al., 1999). In general, hemispheric activity is asymmetric in regards to each type stimulus, with the left and right hemispheres predominantly involved in word and face recognition, respectively. Furthermore, selective impairments in face recognition (i.e. prosopagnosia) or word recognition (i.e. alexia) can arise without a general impairment in object recognition (Farah, 2004), which suggests that separate cognitive processes underlie recognition of faces and words.

Third, it is yet to be empirically determined if metamemory performance for face-name associations can be deduced from studies with verbal stimuli. In their commentary on visual versus verbal metacognition, Diana and Reder (2004) point out:

> "Given that most research on metacognition has involved verbal and semantic tasks and stimuli, it would be ideal to find evidence to support the hypothesis that these conclusions generalize into the visual realm." (pg. 193).

Only a handful of studies have compared monitoring performance between different categories of stimuli within the same experiment (Bornstein & Zickafoose, 1999; Perfect, Watson, & Wagstaff, 1993; Perfect & Hollins, 1996). These have reported conflicting results. On the one hand, Perfect and Hollins (1996) have found that the relative accuracy of FOKs and RCJs is superior for general knowledge questions compared with eyewitness memory. On the other hand, Bornstein and Zickafoose (1999) have found that the relative accuracy of RCJs is similar across both general knowledge questions and eyewitness events. Thus, there is some ambiguity in the literature with regards to the stability of monitoring accuracy across different categories of stimuli, especially those relating to episodic (i.e. eyewitness events) and semantic (i.e. general knowledge) information. To anticipate, Experiments 1 and 2 shed some light on this issue.

As a final point, examining metamemory during face-name learning would address a gap in the literature and provide some insight into participants' knowledge of their face identification capabilities. In a review of encoding operations during face recognition, Coin and Tiberghien (1997) cite five studies that measured the encoding strategy believed to be most effective from participants. Of the five studies, only two gathered data in controlled conditions. Nonetheless, Coin and Tiberghien conclude that the relationship between the encoding strategy selected by participants and the encoding strategy that yields optimal performance in experimental conditions is far from perfect. However, they do not provide any hypotheses as to why this is the case. Likewise, seminal models of person identification either delegate metacognitive processes to a comprehensive but unspecified cognitive system (e.g. Bruce & Young, 1986) or do not include them at all in the architecture (e.g. Burton & Bruce 1992; Burton, Bruce, & Johnston, 1990). The two studies that have measured self-rated face recognition ability were conducted in the context of an eyewitness paradigm with three trials or less per participant (Olsson & Juslin, 1999; Perfect, 2004). Thus, the literature in regards to self-rated memory skill for faces is sparse, which is one motivation for conducting the present research.

Chapter 2 – Mechanisms Underlying Metacognitive Judgements

The purpose of this chapter is to provide a detailed account of the four monitoring judgements in Nelson and Narens' (1990) framework. Numerous studies have reported dissociations among judgements (e.g. Busey, Tunnicliff, Loftus, & Loftus, 2000; Dougherty, Scheck, Nelson, & Narens, 2005; Modirrousta & Fellows, 2008; Pannu, Kaszniak, & Rapcsak, 2005) or have found that the correlations among judgements are weak (e.g. Kelemen et al., 2000; Leonesio & Nelson, 1990). Consequently, each judgement will be reviewed in isolation of others. It should be noted that a greater emphasis was placed on JOLs and FOKs because of their relative weight in the literature.

Ease-Of-Learning Judgements

Purpose.

EOLs are estimates of the degree of difficulty of learning an item to a desired level of mastery. The output of an EOL determines how much study time is allocated to a particular item (Nelson & Narens, 1990). For example, if a participant believes that English-Swahili word pairs are harder to learn than English-English word pairs, he will strategically allocate more study time to the English-Swahili word pairs in order to compensate for the perceived level of difficulty. Several lines of evidence support this learning strategy. First, as the estimated degree of difficulty increases, the time spent studying items also increases (Nelson & Leonesio, 1988). Second, participants spend fewer trials memorizing easy-to-learn word pairs during self-paced study conditions (Leonesio & Nelson, 1990). However, if participants are not aware of an impending memory test, they will allocate more time to studying easier material (Son & Metcalfe, 2000). Thus, EOLs are an effective monitoring process because they enable study time to be efficiently allocated. Consider the situation where EOLs are inaccurate – individuals would spend their time studying items that can be easily recalled and ignore items that warrant effortful encoding.

Procedure.

EOLs can be solicited in a variety of ways. Typically, items are presented individually and participants rate the perceived difficulty of each item on a numerical scale (e.g. de Carvahlo & Yuzawa, 2001). For example, if item A receives an EOL rating of 8 and item B receives an EOL of 2, item A is subjectively easier to learn than item B. EOLs have also been gathered by instructing participants to draw a line beside each item, where line length corresponds to the perceived level of difficulty (Underwood, 1966). Occasionally, all of the items are presented simultaneously and are ranked based on ease of learning (e.g. Leonesio & Nelson, 1990; Son & Metcalfe, 2000). Even though there are variations in the procedure, EOLs ultimately involve estimating the difficulty of memorizing a given item.

Based on the procedures described above, EOLs have been found to be significant predictors of study time allocation (for a brief review, see Son & Metcalfe, 2000), rate of acquisition (Leonesio & Nelson, 1990), and retrieval accuracy (Son & Metcalfe, 2000). Obviously, normative item difficulty has the strongest influence on EOL magnitude. For example, Underwood (1966) found that trigrams such as ‘BUG’ were given greater ease-of-learning judgements that items such as ‘XFH’. Moreover, EOL magnitudes form a bi-modal distribution if word pairs are grouped according to easy versus hard difficulty levels (de Carvahlo & Yuzawa, 2001) or into semantically related versus unrelated categories (Dunlosky & Matvey, 2001).

Mechanisms underlying EOLs.

As Schwartz (1994) points out, there is little research that has explicitly examined the mechanisms underlying EOLs. However, considering that participants have yet to memorize the to-be-remembered items, EOLs are presumably based on, to use Koriat's (2007) terminology, theory-based metacognition, such as a-priori knowledge of memory ability. For example, if participants believe that their memory for certain items (e.g. semantically related word pairs) will be poorer relative to others (e.g. semantically un-related word-pairs), EOLs will reflect those beliefs (Dunlosky & Matvey, 2001). Additionally, extrinsic properties of the learning situation may also contribute to EOL judgements. When participants are deceptively informed that certain word pairs are considered easy to memorize, they bias their EOLs to correspond with the social cue (de Carvahlo & Yuzawa, 2001).

EOLs for faces.

Although the predictive validity of EOLs for face-name associations has yet to be established, some researchers have examined self-efficacy in face recognition tasks. Olsson and Juslin (1999) set out to determine if self-reported face recognition skill is a valid predictor of recognition performance. Under incidental learning conditions, participants were shown a videotape of a staged robbery and were later asked to identify the perpetrators in a lineup. Prior to viewing the videotape, participants rated their face recognition ability on an 11-point bipolar scale. The mean point-biserial correlation between self-reported skill and face recognition performance revealed a weak (r_{pb} =.12) but significant relationship.

Along similar lines, Perfect (2004) examined the relationship between self-reported memory ability for faces and actual recognition performances under explicit learning conditions. Participants were asked to rate their face recognition ability on a ten-point scale. Afterwards, they studied six faces for five seconds each and then immediately attempted to recognize a previously studied face among five distracters. The recognition test was repeated for three separate sets of distracters. The correlation between self-reported ability for recognizing faces and objective recognition performance revealed a non-significant relationship (r = .06). Taken together, these two studies suggest that self-rated ability in face recognition is, at best, weakly related to actual performance.

Judgements-of-Learning

Purpose.

A JOL is an estimation of the likelihood of future recall. For example, when students reflect on whether they can successfully recall recently studied material for an upcoming exam, they are essentially conducting a JOL. Thus, JOLs are thought to monitor the current level of mastery of an item. In Nelson and Narens' (1990) framework, the control function of JOLs serves to determine whether additional study time is required or if a new learning strategy is warranted. For instance, participants allocate additional study time to items with low JOLs under self-paced study condition (Nelson & Leonesio, 1988).

Dunlosky and Hertzog (1998) proposed the discrepancy-reduction hypothesis to account for how the output of JOLs control study time. Briefly, the hypothesis assumes that participants can monitor the strength of an item's memory and compare it to a desired level of mastery. If the strength of the memory is less than the desired level of mastery, additional study time is allocated to that item. The process continues until the item is sufficiently committed to memory. However, a discrepancy-reduction mechanism is not necessarily the optimal strategy for all learning situations. Under time pressure, participants will allocate additional study time to easier items than more difficult ones (Thiede & Dunlosky, 1999). This strategy is ideal because in situations with limited time to learn items, focusing on the easiest material would maximize the amount of items that can be correctly retrieved during a memory test.

Procedure.

A JOL experiment consists of three phases. First, a list of items, usually noun-noun paired associates is presented to participants for study. Immediately after the offset of a paired-associate (i.e. immediately after studying the item), a cue is presented (i.e. one of the nouns) and participants are instructed to estimate the likelihood of remembering the target (i.e. the corresponding noun). This phase continues until JOLs for all of the studied paired-associates are obtained. The third phase involves a cued-recall test, in which the same cues are presented again and participants are instructed to recall the appropriate target.

Based on the procedure described above, JOLs have been found to be moderately accurate predictors of future recall performance. Across eleven experiments that used paired-associates, the average Goodman-Kruskal gamma (γ) correlation between recall accuracy and JOL magnitude is approximately 0.56 (Koriat et al., 2002). In terms of calibration, overconfidence is found when judgements are solicited on an item-by-item basis, but underconfidence if found when judgements are solicited for aggregates of items (Mazzoni & Nelson, 1995). Nonetheless, JOLs are considered to be predictive of future recall performance. Interestingly, when a significant delay occurs between study and solicitation of the JOL, the relative and absolute accuracy of JOLs significantly improves (e.g. Nelson & Dunlosky, 1991). This is referred to as the delayed-JOL effect, and is discussed in more detail below.

Factors affecting JOLs.

Several studies have reported that encoding and retrieval fluency can exert disproportionate effects on JOL magnitude compared with objective memory performance (e.g. Begg et al., 1989; Benjamin, 2003; Benjamin & Bjork, 1996; Koriat & Ma'ayan, 2005). For example, manipulating word frequency can result in dissociations between JOLs and recognition accuracy. Although recognition accuracy is greater for low frequency words compared with high frequency words (e.g. Glanzer & Adams, 1990), estimates of future retrieval performance are greater for high frequency words compared with low frequency words (Begg et al., 1989; Benjamin, 2003). This could be due to the fact that high frequency words are presumably encoded more fluently, and participants use encoding fluency as a heuristic when monitoring their memory (Koriat & Ma'ayan, 2002). Along similar lines, Matvey, Dunlosky and Guttentag (2001) found a negative correlation between response latency and JOL magnitude. Thus, the longer it takes for an item to come to mind (i.e. low retrieval fluency), the less confident a participant will be in his ability to recall that item.

In addition to encoding and retrieval fluency, JOLs are also sensitive to practice (Lovelace, 1984) and over-learning (Leonesio & Nelson, 1990). One surprising effect of practice is that it decreases absolute accuracy while improving relative accuracy (Koriat, 1997; Koriat et al., 2002). For example, in Koriat (1997) four study-test blocks contained either the same paired associates across all blocks or different paired associates in each block, allowing for participants in one condition to receive repeated practice for studying and retrieving a list of paired-associates. Surprisingly, participants in the repeated practice condition demonstrated underconfidence after the first block, but became increasingly accurate at discriminating between items that will or will not be recalled in later blocks. In other words, practice decreased absolute accuracy while improving relative accuracy. This effect is known as the underconfidence-with-practice effect (Koriat et al., 2002; Finn & Metcalfe, 2007).

There is some evidence to indicate that mental imagery contributes to the accuracy of JOLs at predicting memory performance. For example, both JOLs and recognition accuracy are greater for concrete words compared with abstract words (Begg et al., 1989). This suggests that imagery is a useful source of information when monitoring memory because JOLs remain concordant with recognition accuracy as imageability varies.

Perhaps the most influential factor on relative JOL accuracy is the time at which JOLs are solicited (Connor, Dunlosky, & Hertzog, 1997; Dunlosky & Nelson, 1992; 1994; Nelson & Dunlosky, 1991; Nelson, Narens, & Dunlosky, 2004; Thiede & Dunlosky, 1994). Although JOLs are significant predictors of recall performance, they are far from perfect. This could be due to the fact that JOLs solicited immediately after study monitor the contents of short-term memory, but a future recall test involves retrieving information from long-term memory. Thus, monitoring the contents of short-term memory is not predictive of long-term memory performance. This is referred to as the Monitoring Dual Memories (MDM) hypothesis. Nelson and Dunlosky (1991) tested this hypothesis by comparing JOLs solicited immediately after study or after a 30-second delay. The results revealed a clear advantage for delayed-JOLs: the gamma correlation between recall accuracy and JOL magnitude was +.90 in the delayed condition compared to +.36 in the immediate condition.

Others have argued that the delayed-JOL effect arises from a covert retrieval attempt of the target in the delayed condition (Spellman & Bjork, 1992) or on an increased reliance on retrieval fluency (Koriat & Ma'ayan, 2005) rather than from monitoring long-term memory as opposed to transient short-term memory. In support of this account, Koriat and Ma'ayan found that as the delay between study and JOL increases, retrieval latency becomes increasingly predictive of JOL accuracy. Although the mechanisms underlying the effect have yet to be resolved (Nelson et al., 2004; Metcalfe & Finn, 2008; Spellman & Bjork, 1992) the advantageous effects of delay on JOL accuracy have been reported in numerous studies (for a brief review, see Metcalfe & Finn, 2008), under various learning conditions (Dunlosky & Nelson, 1994; Thiede & Dunlosky, 1994), and in different age populations (Connor et al., 1997).

Mechanisms underlying JOLs.

How can participants reliably and accurately predict future recall for a given item? There are two contrasting viewpoints that attempt to provide an answer to this question. According to the direct-access view, JOLs monitor the strength of an item's memory trace during and after learning. Over the course of study, the strength of the memory trace fluctuates and the magnitude of JOLs reflects those fluctuations (e.g.

Dunlosky & Hertzog, 1998; King, Zechmeister & Shaughnessy, 1980, for a brief review of trace-access views, see Metcalfe, 2000). Accordingly, the strength of an item in memory is proportional to the magnitude of its JOL. The predictive accuracy of JOLs stems from the fact that items with stronger memory traces are more likely to be retrieved from memory.

However, as previously mentioned certain factors can dissociate JOL magnitude from object memory performance (e.g. Begg et al., 1989; Benjamin, 2003), which should not occur according to the trace-access view. As a result, an alternative view has been proposed which considers JOLs, and monitoring judgements in general, to be inferential in nature (Koriat, 1997; 2007). According to the cue-utilization hypothesis, JOLs are based upon a variety of sources of information, not just a memory trace. In particular, Koriat (1997) distinguishes between three classes of cues that participants can use to monitor their memory: 1) Intrinsic cues that refer to the inherent characteristics of the stimuli (e.g. concrete versus abstract nouns); 2) Extrinsic cues that refer to the learning conditions and strategies used by the learner (e.g. practice or imagery); and 3) Mnemonic cues that refer to the subjective experience of encoding and retrieval (e.g. perceptual fluency or retrieval fluency). The degree to which JOLs are accurate at predicting retrieval depends upon the validity of the cues that are present during JOL solicitation. As Koriat (1997) argues:

> “JOLs are based on inferences and heuristics, their accuracy is not guaranteed but depends on the empirical correlation between the cues used and the criterion memory test. In general, JOLs are accurate as long as the cues used at the time of

making the judgments are consistent with the factors that affect subsequent performance on the criterion memory test." (pg. 350).

It should be noted that some cues are more diagnostic of performance than others. In particular, intrinsic cues are given a greater weight than extrinsic cues because they reflect the relative memorability of items (however, see Dunlosky & Matvey, 2001 for an opposing view). The cue-utilization hypothesis can account for the findings (e.g. Begg et al., 1989) that were difficult to explain with direct-access theories. Dissociations between JOL magnitude and memory performance are expected whenever participants base their JOL on an erroneous cue. In situations where cues lack predictive validity, JOLs will be inaccurate.

JOLs during face recognition and identification.

Research on estimates of future memory performance for face-name associations is limited. Two studies have solicited JOLs for faces (Busey et al., 2000; Sommer, Heinz, Leuthold, Matt, & Schweinberger 1995) and two studies have solicited JOLs for face-name pairs (Modirrousta & Fellows, 2008; Tauber & Rhodes, 2010). These studies will be reviewed in chronological order.

In Sommer et al. (1995), 10 face images were sequentially presented to participants. Immediately following each face, half of the participants either rated the distinctiveness of the face or rated whether they would be able to recognize the face sometime in the future (i.e. a JOL). Afterwards, five of the study faces were randomly mixed with five novel distracters, and all participants made old-new recognition ratings for each face. All ratings were made on a 4-point scale, and this process was repeated for 17 blocks, with a new set of faces in each block. The mean gamma correlations between JOLs and recognition ratings ($\gamma = +.44$) and JOLs and distinctiveness ratings ($\gamma = +.75$) were significantly greater than zero. However, when controlling for the effect of distinctiveness, the strength of the relationship between JOLs and recognition ratings decreases ($\gamma = +.26$) and is no longer significant. These results indicate that subjects can monitor whether a face has been sufficiently committed to memory, but variations in facial distinctiveness can have a significant influence on the predictive validity of JOLs.

The second study that measured JOLs for faces also followed an old-new recognition paradigm (Busey et al., 2000). During the study session, 30 face images were sequentially presented at different durations (Experiment 1) or at different luminance values between study and test (Experiments 2 and 3). Following the offset of each face, participants were instructed to rehearse the face for 15 seconds or complete a filler task for 15 seconds. JOLs were solicited after the rehearsal or filler tasks. Once JOLs were obtained for all items, 30 novel faces were randomly mixed with the studied faces in an old-new recognition test. Following each recognition decision, retrospective confidence ratings were solicited. All ratings were made on a 5-point probabilistic scale (e.g. 25% likely I remember the picture sometime in the future). In regards to absolute accuracy, JOLs were disproportionately inflated following rehearsal and RCJs were disproportionately inflated if the luminance of the face image increased from study to test. Thus, participants overestimated their ability to recognize a face if they rehearsed that face or were overconfident in their recognition decisions if the face was brighter at test than at study. The gamma correlations[1] between JOL magnitude and recognition

---

[1] The correlations were not reported in the published version of the manuscript. However, they are reported in an unpublished version, which is available on-line at: http://cognitrn.psych.indiana.edu/busey/WWWPubs/buseyPubs.html

accuracy, on the other hand, did not significantly vary between rehearsal, presentation duration and luminance conditions. Nonetheless, across three experiments the relative accuracy of JOLs was significantly greater than zero and was comparable in magnitude with the findings from Sommer et al. (1995).

Turning to face-name learning, Modirrousta and Fellows (2008) compared metamemory ability between patients with frontal lobe damage and healthy controls. Four blocks of six unfamiliar face-name pairs were presented to each group of participants. Following the learning phase of each block, participants were instructed to estimate the proportion of faces they would be able to correctly name during an upcoming memory test. Thus, in contrast to previous studies, JOLs were operationalized to reflect prospective confidence for a block of paired-associates rather than for items on a trial-by-trial basis. After a one-minute delay, a cued-recall task was presented with the previously studied faces acting as cues. After every recall response, participants estimated their subjective confidence in their answer (i.e. a RCJ) and the likelihood of correctly recognizing the name from a list of three alternatives (i.e. an FOK). Once the FOK ratings were completed for every item in the block, participants completed a three-alternative-forced-choice (3-AFC) recognition test. This process was repeated for the remaining blocks. The authors found that for both healthy controls ($r = 0.24$) and brain damaged patients ($r = 0.27$), JOL ratings were positively correlated with name recall and were significantly greater than zero.

Recently, Tauber and Rhodes (2010) solicited JOLs during face-name and face-occupation learning to determine if participants are aware that their memory for names is generally poorer compared with other types of identity-related semantic information (see

McWeeny, Young, Hay, & Ellis, 1987). Importantly, the same word was used as both a name and an occupation (e.g. Mr. Baker versus a baker), which enabled control over possible confounds such as frequency, familiarity or imageability. Following the typical procedure for soliciting JOLs, the authors found that cued-recall accuracy was greater for words when their status was an occupation compared to when their status was a name, but the magnitude of JOLs did not differ between names and occupations. This pattern of results indicates that participants are not initially aware of the difficulty of learning names compared with occupations. However, in a follow-up experiment, when participants were given the opportunity to practice studying and recalling the paired associates, JOLs for names were lower compared with occupations. Thus, it appears that participants can learn from experience when monitoring their memory during face identification tasks, and errors in name recall could be due to inaccurate monitoring during study.

In summary, three studies have reported that JOLs are valid predictors of face recognition (Busey et al., 2000) or face identification (Modirrousta & Fellows, 2008; Tauber & Rhodes, 2010) performance, and one study found that the predictive validity of JOLs on face recognition is dependent on facial distinctiveness (Sommer et al., 1995). Thus there is some conflicting evidence regarding individuals' ability to monitor their memory for faces. Additional research is needed to delineate the effects of distinctiveness on JOLs for face-name associations and whether monitoring performance varies depending on the whether a face or name must be retrieved from memory.

Feeling-of-Knowing Judgements

Purpose.

An FOK judgement is an estimation of the likelihood of recognizing a currently non-recallable item. Subjectively, it is the state where an individual feels that an item can be recognized even though it cannot be recalled. The control function of FOKs determines the initiation and duration of a memory search. A low FOK signals that a memory search is not warranted or the duration of the memory search will be brief, whereas a high FOK signals that a memory search should continue until an item is retrieved or until the amount of retrieval attempts surpasses a certain threshold (e.g. Nelson, et al., 1984; Reder, 1987; 1988; Schwartz, 2001). Indeed, a typical finding is that the latency of a memory search is correlated with FOK magnitude (Costermans et al., 1992; Nelson & Narens, 1990).

A distinction should be drawn between FOKs and tip-of-the-tongue (TOT) states (for a review, see Schwartz, 2008). A TOT state refers to the feeling of immanent retrieval for an inaccessible item in memory (Brown & McNeill, 1966), whereas FOKs refer to a prediction regarding the probability that an inaccessible item can be recognized among alternatives sometime in the future (Koriat, 1993). TOTs are generally described as all-or-none states, whereas FOKs are graded along a probabilistic continuum. Furthermore, working memory load or social cues can dissociate TOTs from FOKs (e.g. Schwartz, 2008; Widner, Smith, & Graziano, 1996)

Procedure.

FOK studies generally follow the Recall-Judge-Recognize (RJR) procedure, which was first introduced by Hart (1965; 1967) in his pioneering studies on metamemory. In the first stage, participants are presented with cues that are intended to prompt recall of a target. If participants cannot accurately recall the target, they judge the likelihood of recognizing the target in an upcoming recognition test. The judgements are typically made on a 5-point probabilistic scale, much like JOLs. Finally, the third stage involves a recognition test for items that could not be recalled. Following this procedure, FOKs have been found to be moderate predictors of memory performance (e.g. Koriat, 1993; Nelson & Leonesio, 1990; Nelson, Gerler, & Narens, 1984; Nelson & Narens, 1990; for a review, see Schwartz, 1994), with a mean γ of 0.31 between FOK magnitude and recognition accuracy (for a meta-analysis of FOK studies, see Schwartz & Metcalfe, 1994). This indicates that items can be monitored without actually being retrieved.

Factors affecting FOKs.

In comparison to other metacognitive judgements, the literature on FOKs is extensive (for reviews, see Schwartz, 1994; Metcalfe, 2000, Schwartz & Metcalfe, 1994; Koriat, 2007). Only the most important factors that affect FOKs will be reviewed here. For a wider range of factors that influence FOKs, interested readers are directed to the reviews mentioned above.

The magnitude of FOKs varies depending on the type of retrieval error. Krinsky and Nelson (1985) used normative general knowledge questions to control for the number of commission errors (i.e. an incorrect answer) and omission errors (i.e. cannot recall the answer) during recall. They found that FOKs were significantly lower for omission errors compared with commission errors. In other words, participants predict a lower likelihood of recognition when they cannot initially recall the answer compared to when the answer is incorrect. A theoretical explanation of this finding is presented below.

Several studies have reported reliable effects of cue priming on FOK magnitude (for a review, see Narens, Jameson, & Lee, 1994). When cues are overtly primed, recall performance is unaffected but the magnitude of FOKs increases (Schwartz & Metcalfe, 1992; Reder, 1987; 1988). Conversely, when answers to general knowledge questions are tachiscopically presented at sub-threshold rates, recognition performance improves but FOK magnitude is unaffected. Thus, cue priming influences the magnitude of FOKs but target priming does not.

The predictive validity of FOKs is influenced by the format of the recognition test. For example, the number of alternatives on the recognition test is positively correlated with the relative accuracy of FOKs (see Schwartz & Metcalfe, 1994). As Leonesio and Nelson (1990) point out, the likelihood of correct guessing on the recognition test is lower as the number of response alternatives increase. Consequently, FOKs are more accurate estimates of recognition performance as the number of alternatives increases.

One factor that bears significant relevance to the dissertation is the effects of target set size. Koriat and Lieblich (1977) contend that FOK judgements are optimal predictors of memory when the cue intended to prompt an FOK distinguishes between several possible targets in memory. A cue that is associated to a unique target will serve as the best indicator for an accurate FOK judgement because unique targets enable the memory search to hone in on one item without added noise from other possible targets. Thus, cues associated with distinct targets should result in superior FOK accuracy.

Izaute, Chambers, and Larochelle (2002) set out to test this claim with proper names as stimuli. Proper names are an ideal class of stimuli because the semantic information associated with a name can only refer to one item in memory, and therefore the target will be relatively distinct. For example, the description “American astronaut.

The first man to set foot on the moon," can only refer to Neil Armstrong. If proper names are unique targets in memory, and if FOKs are more accurate for distinctive targets, then FOKs should be more accurate for proper names than common words.

Following this reasoning, Izaute et al. instructed participants to first read a brief description of a proper name (e.g. "20th century philosopher of science. Advanced a criterion of falsifiability for the demarcation between science and metaphysics")[2] or a common word (e.g. "Sort of metal ring hanging from either side of the saddle to support the rider's feet")[3], then estimate the probability of recalling the corresponding name or word on a 6-point scale, and then attempt to recall the name or word. After every recall attempt, an FOK was solicited. Afterwards, a cued-recall test was administered, with the first letter of the target name or word acting as a cue. The difference between proper names ($\gamma = 0.51$) and common words ($\gamma = 0.35$) on the predictive validity of FOKs on cued-recall accuracy was significant. As with JOLs (e.g. Sommer et al., 1995), these results indicate that FOKs are also sensitive to manipulations of target distinctiveness. However in the case of Izaute et al. (2000), distinctiveness was defined in terms of target set size rather than perceptual features.

Mechanisms underlying FOKs.

Like JOLs, the mechanisms thought to underlie FOKs can be categorized into trace-access theories and inferential theories. Trace-access theories assume that participants can directly monitor some aspect of the target in memory, whereas inferential

[2] Karl Popper
[3] Stirrup

theories posit that a variety of sources of information underlie the monitoring judgement. These two categories of theories are described in greater detail below.

Several trace-access theories have been proposed to explain how FOKs monitor unrecallable items (see Nelson et al., 1984). In general, these theories posit that FOKs reflect the output of an internal monitor that detects the strength of an item in memory. For items that are not in memory or are rarely accessed, the strength of their memory trace will be relatively low, which subsequently results in a low feeling-of-knowing. Thus, FOKs reflect access to a memory trace of sought after information without actually having to access that information.

Inferential theories, on the other hand, do not assume that FOKs reflect the ouput of an internal monitor. According to the cue-familiarity theory (Reder, 1987; 1988; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992), the familiarity of the cue determines the magnitude of FOKs. In this theory, FOKs are considered to be a rapid pre-retrieval process that guides the memory search depending on the properties of the cue (Miner & Reder, 1994). High FOKs indicate that the cue is familiar and a memory search is warranted, whereas low FOKs indicate that a memory search is not warranted.

Several findings support the cue-familiarity theory (for a review, see Metcalfe, 2000). Reder's (1987; 1988) work is typical of this body of research. Her paradigm for assessing the role of cue-familiarity involved two stages. First, participants were primed with the words from general knowledge questions through an incidental learning task, such as rating the frequency-of-occurrence. The second stage involved a cued-recall test that was modeled after trivia game shows. General knowledge questions were presented and participants were encouraged to respond as quickly as possible. Some of the general knowledge questions contained words that were primed by the ratings task. The key manipulation was the type of response that participants could provide. In one condition, participants were instructed to rapidly estimate whether they could provide an answer to the general knowledge question even if they could not actually recall it. In another condition, participants were instructed to immediately provide the answer. Thus, the estimation condition involved an FOK response, whereas the answer condition involved retrieving an item from memory.

Reder (1987) found that response latencies were quicker for the estimate condition compared with the answer condition, which is consistent with the idea that FOKs involve a rapid pre-retrieval process. Importantly, the magnitude of FOKs was greater for primed compared with unprimed questions, even though retrieval performance was not affected. Thus, the familiarity of the cue primarily affects FOKs rather than the strength of the item in memory. Finally, participants in the answer condition spent more time searching for the answer to primed rather than unprimed questions, which should occur since priming increases FOK magnitude. As mentioned above, others have found that advanced priming of cues in paired-associate paradigms increases the magnitude of FOKs without affecting recall (Schwartz & Metcalfe, 1992), whereas advanced priming of targets increases recall without affect FOK ratings (Jameson, Narens, Goldfarb, & Nelson, 1990), which is consistent with the cue-familiarity account.

An alternative account of FOKs has been proposed that emphasizes the role of the target as opposed to the cue. The accessibility hypothesis contends that FOKs are based on the quantity and quality of information retrieved during a memory search (Koriat, 1993; 1995; Koriat & Levy-Sadot, 2001). Importantly, any partial information that is accessed throughout the course of a memory search, regardless of objective accuracy, will contribute to the magnitude of an FOK. For example, when searching for an individual's name, partial information about the person's identity often comes to mind, such as occupation or the context in which the person was last seen, which increases the subjective feeling that a name will be retrieved.

The accessibility hypothesis assumes that participants cannot assess the validity of information related to the target and therefore, objectively inaccurate information may also contribute to an FOK. According to this account, when a retrieval attempt is unsuccessful, the by-products of the memory search act en masse to determine the likelihood that the target will be retrieved sometime in the future. Thus, the magnitude of FOK reflects the quantity of partial information accessed during the memory search, whereas the quality of information that comes to mind will affect the accuracy of FOKs. The fact that FOKs are valid predictors of recognition performance stems from the fact that information that comes to mind is more likely to be correct than incorrect.

There are several findings that support the accessibility hypothesis. First, the number of letters that can be recalled from studying nonsense letter strings, regardless of their objective accuracy, is positively correlated with FOK magnitude (Koriat, 1993). Second, the number of correct letters that are retrieved is a significant predictor of recognition performance (Koriat, 1993). Thus, the quality of retrieved partial information is related to FOK accuracy. Third, manipulations of outbound-accuracy (i.e. the veracity of partial information that comes to mind) affect the accuracy of FOKs, whereas manipulations accessibility (i.e. the amount of partial information that can be retrieved) affect the magnitude of FOKs (Koriat, 1995). Finally, FOKs are greater for commission errors compared with omission errors (Krinsky & Nelson, 1985). This finding is consistent with the predictions of the accessibility hypothesis because the quantity of information that is retrieved, regardless of accuracy, contributes to the subjective feeling that an answer can be retrieved.

Although the cue-familiarity and accessibility hypotheses differ in some respects, they are not mutually exclusive. Recently, Koriat and Levy-Sadot (2001) attempted to unify both accounts into one comprehensive explanation of FOK. They propose that cue-familiarity and accessibility are two separate mechanisms that contribute independently to FOK. Consequently, FOKs are based on the pre-retrieval familiarity of the cue and the post-retrieval quantity of information related to the target. Three predictions derive from this conceptualization. First, if the cue is not familiar, then a memory search for additional information is not warranted and FOK ratings will be low. Second, if the cue is familiar, but the quantity of accessible information is low, then FOK ratings will be intermediate. Finally, if the cue is familiar and the quantity of accessible information is high, then FOK ratings will be high. Thus, the amount of accessible information should only affect FOK magnitude when the cue is relatively familiar.

To test their theory, Koriat and Levy-Sadot (2001) selected general knowledge questions that varied along normative measures of accessibility and familiarity. They also solicited FOKs after every recall attempt in order to prevent participants from receiving feedback on the accuracy of their responses during the cued-recall test. They found that the effects of accessibility on FOK magnitude were stronger when the familiarity of the cue was high compared to when it was low. As predicted, participants only searched their memory for partial information related to the target when the cue was familiar. Indeed, the greatest FOKs occurred for questions that were highly accessible and highly familiar.

FOKs during face identification.

Considering that feelings-of-knowing often occur when a face is familiar but a name cannot be recalled, it is surprising that FOKs are not often investigated in face identification research. Three studies have solicited FOKs after name recall errors, two of which examined the neuropsychological impact of frontal lobe damage on monitoring accuracy (Hosey, Peynircioğlu, & Rabinovitz, 2009; Modirrousta & Fellows, 2008; Pannu et al., 2005). In Hosey et al.'s (2009) second experiment, participants learned to associate names and three pieces of semantic information with sixty unfamiliar faces. Afterwards, a cued-recall test was presented with the previously studied faces acting as cues and names acting as targets. For every unsuccessful retrieval attempt, an FOK judgement was solicited and participants were asked to indicate whether they based their FOK on a sense of familiarity of the face (i.e. a cue-familiarity strategy) or on recollection of specific semantic information associated with the face (i.e. an accessibility strategy). During the final stage of the experiment, participants completed a 4-AFC recognition test with previously studied names as distracters. The correlation between the magnitude of FOKs and name recognition accuracy was above chance ($\gamma = 0.22$), which indicates that FOKs are predictive of name recognition when cued with an image of a face. Furthermore, the magnitude of FOKs was greater for items that were based on an accessibility strategy compared with a cue-familiarity strategy, even though the cue-familiarity strategy was reported more often.

Modirrousta and Fellows (2008) and Pannu et al. (2005) compared the accuracy of FOKs for names between healthy controls and patients that sustained frontal lobe damage. The former used unfamiliar faces as cues whereas the latter used famous faces. Both studies reported that controls were above chance at predicting name recognition performance, which is consistent with the findings of Hosey et al. (2009). Interestingly, Pannu et al. (2005) found that RCJs were impaired in frontal patients compared with healthy controls, whereas Modirrousta and Fellows (2008) found that FOKs were impaired in frontal patients compared with healthy controls.

In summary, three studies suggest that when name recall fails in the presence of a face cue, estimates of future name recognition performance are valid predictors of actual name recognition performance. However, none of the aforementioned studies examined if FOKs are predictive of face recognition accuracy or if the distinctiveness of the target affects the accuracy of feeling-of-knowing judgements in face identification tasks.

Retrospective Confidence Judgements

Purpose.

An RCJ is an estimation of the likelihood that an item retrieved from memory is the correct target. According to Nelson and Narens (1990), RCJs monitor the outcome of retrieval by comparing items to a confidence threshold. If the retrieved item supersedes the confidence threshold, then a decision is made to provide the answer as a response and the RCJ reflects the value of the confidence threshold. In the case where a candidate answer is below the threshold, a new answer is selected or the threshold is ultimately lowered. The control function of RCJs is to determine whether a candidate answer should be volunteered or withheld. In comparison to EOLs, JOLs, and FOKs, which are prospective monitoring judgements, RCJs are solicited after recognition decisions, and reflect retrospective memory monitoring.

Procedure.

Retrospective confidence judgements are solicited immediately after a recognition response. Participants are asked to estimate the probability that the answer they provided is correct. Estimates are typically made on a five-point probabilistic scale, like JOLs and FOKs. The standard finding with general knowledge questions is overconfidence (for a review, see Lichtenstein, Fischhoff, & Phillips, 1982), which possibly stems from a bias to emphasize the supporting evidence of a retrieved answer without acknowledging contradictory evidence (Koriat, Lichtenstein, & Fischhoff, 1980). Interestingly, overconfidence is usually observed for difficult items, whereas underconfidence is usually observed for easy items. This is often referred to as the hard-easy effect (Juslin, Winman, & Olsson, 2000).

In regards to face identification, the relationship between retrospective confidence and memory accuracy is far from perfect. The between-subject point-biserial correlation between retrospective confidence and identification accuracy during eyewitness lineup tasks varies between zero and .30 (for reviews and meta-analyses of the retrospective confidence-accuracy (CA) relationship, see Sporer, Penrod, Read, & Cutler, 1995; Wells & Murray, 1984). However, when the encoding conditions are optimal (Deffenbacher, 1980) or when there is considerable variability across participants and the testing conditions are naturalistic (Lindsay, Read, & Sharma, 1998), the correlation between confidence and accuracy increases. In regards to calibration, some studies have found that retrospective confidence is concordant with the probability of correct identification in eyewitness tasks (Brewer & Wells, 2006; Juslin, Olsson, & Winman, 1996), but others have found that participants are overconfident (Olsson & Juslin, 1999). In general, it appears as though the retrospective confidence-accuracy relationship is variable and susceptible to many factors.

Factors affecting RCJs for faces.

Encoding specificity has been found to influence RCJ accuracy for faces. Busey et al. (2000) varied the degree to which luminance values for faces at the time of study match those at the time of test. When faces were studied in a dim condition but tested in a bright condition, the magnitude of RCJs increased but recognition accuracy decreased. Apparently, participants erroneously believe that brighter test stimuli improve face recognition accuracy.

The results are mixed in regards to the effects of facial distinctiveness on retrospective confidence. On the one hand, the magnitude of RCJs is greater for distinct faces compared to typical faces (Semmler & Brewer, 2006). On the other hand, the magnitude of RCJs is approximately identical when participants are encouraged to either study the differences between faces, and therefore emphasizing distinctiveness, or study the similarities between faces, and therefore emphasizing the relations between faces (Mantyla, 1997, Experiment 2). Unfortunately, neither of the aforementioned studies compared the relatively accuracy of RCJs between typical and distinct conditions.

Aside from manipulations of the properties of the stimuli, RCJs for faces also vary according to task demands. For instance, the accuracy of RCJs decreases as exposure duration of the face image during study decreases (Busey et al., 2000; Weber & Brewer, 2004). In a similar vain, asking participants to rehearse a face for 15 seconds immediately after study improves the relative accuracy of RCJs compared to a control condition (Busey et al., 2000). Ultimately, the evidence suggests that RCJs are sensitive to the parameters of the task and of the stimuli.

Mechanisms underlying RCJs.

There is a wide range of explanations concerning how individuals develop retrospective confidence in their answers (e.g. Clark, 1997; Gigerenzer, Hoffrage, & Kleinbölting, 1991; for a brief review, see Harvey, 1997). Two prominent approaches to explaining the mechanism underlying RCJs reflect the distinction between theory-based and experienced-based metacognition: One involves an effortful and deliberate consideration of selected answers against possible alternatives (Nelson & Narens, 1990; Koriat et al., 1980; Koriat & Goldsmith, 1996b), whereas the other involves an automatic and unconscious application of an ease of retrieval heuristic (Dougherty et al., 2005; Kelley & Lindsay, 1993). These two approaches are discussed in more detail below.

According to the model developed by Koriat and Goldsmith (1996b), confidence judgements reflect a deliberate signal-detection like mechanism. When faced with a situation where an answer must be retrieved from memory, individuals monitor the probability that the retrieved answer is correct and compare that probability to a preset threshold, which can be modified by situational factors such as accuracy incentives (e.g. penalties versus rewards) or report options (e.g. free report or forced recognition). If the retrieved answer has a higher subjective probability of being correct than the pre-determined threshold, the answer is volunteered. If the subjective probability is below the threshold, the answer is withheld and a new cycle commences with an alternative answer.

Retrospective confidence is the subjective probability of the volunteered answer. If no answer can be provided, retrospective confidence is zero.

Alternatively, Kelley and Lindsay (1993) speculated that the ease with which information comes to mind serves as a heuristic underlying retrospective confidence judgements. Accordingly, prior exposure to answers of general knowledge questions should increase the speed to which the answers come to mind, which in turn should lead participants to increase their confidence in the accuracy of their answer. The results of four experiments support this prediction. Kelley and Lindsay found that confidence varied as a function of retrieval fluency whereas accuracy did not. Specifically, prior exposure to either correct or incorrect answers decreased response latency and increased confidence. Thus, if an answer comes to mind quickly, regardless of its objective accuracy, participants will be more confident in that answer. In support of this finding, Nelson et al. (1990) reported inverse correlations between retrieval latency and RCJ magnitude for correct and incorrect responses. Indeed, the aforementioned affects of encoding specificity, stimulus duration, and distinctiveness on RCJs can be interpreted in to support a retrieval fluency explanation of RCJs, as each of these factors is likely to affect the ease at which items come to mind.

Chapter 3 – Face Identification

Compared to the identification of everyday objects, faces present a unique challenge to a perceptual memory system. Not only do faces contain the same features but they also share a common configuration, which makes individuation an impressive feat. Precisely how this is accomplished has yet to be determined (cf. Pelli, 1999). The purpose of this chapter is to review research on face-name learning and highlight relevant factors that influence face and name recognition. This chapter begins with a review of two seminal models of person identification, followed by reviews of distinctiveness effects in faces and names, and concludes with a discussion of errors in face identification.

Models of Person Identification

Although several models of face recognition have been proposed (e.g. Schwaninger, Wallraven, & Bulthoff, 2004; Turk & Pentland, 1991; Valentine, 1991), Bruce and Young's (1986) functional model is arguably one of the most influential. An overview of their model is presented in Figure 3. According to their model, face identification is accomplished through a series of discrete and sequential stages. At the first stage a face is encoded into a structural-code, which contains the unique visual cues that belong to a given face. With frequent exposure, the structural code becomes increasingly detailed and its information becomes stored in a viewpoint and expression invariant face recognition unit (FRU). If there is a match between an encoded representation of a face and a structural code, then the corresponding FRU signals to a higher level cognitive system to indicate that the face is familiar. The strength of the signal is dependent on the degree of similarity. Faces can be recognized as familiar at the level of the FRU, but identity-related semantic information associated with the face is only retrieved following the activation of a Person Identity Node (PIN). Thus, FRUs signal that a face is familiar, whereas PINs signal the identity of the person associated with the face. PINs receive signals from FRUs, but also from other inputs such as voices or body-shape. These inputs are provided by a higher level cognitive system, which is bi-directionally connected to PINs. The bi-directionality also allows for tangential autobiographical information to be accessed from semantic memory (e.g. The unabomber's academic pedigree traces back to the logical positivists and members of the Vienna circle, of which Karl Popper was sometimes associated with. According to the model, my knowledge of this fact is due to the Karl Popper PIN having access to semantic memory). Name codes are stored in a separate unit and can only be accessed via PINs. Thus, the name associated with a face can only be recalled once semantic information is retrieved.

Bruce and Young acknowledge that some components of their model require additional research in order to be fully fleshed out. For instance, the executive functions of the model are designated to a comprehensive but unspecified cognitive system. Consequently, monitoring and control processes are incorporated into the model, but are not explicitly described. For example, the cognitive system selects various recognition strategies depending on the context of the situation or directs resources to specific components of the system such as expression analysis or speech analysis. Moreover, the cognitive system is presumably responsible for allocating study time, determining whether a name has been sufficiently associated with a face, or whether the correct FRU and PIN have been activated.

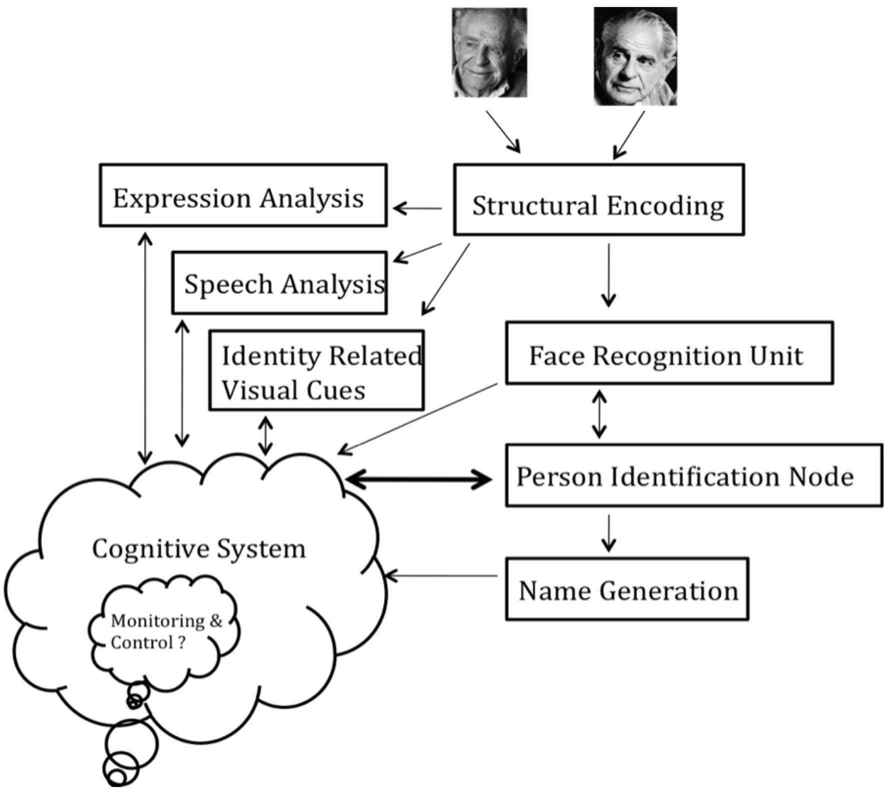

Figure 3. Depiction of Bruce and Young's model of person identification. Adapted from Bruce and Young (1986).

The interactive activation and competition (IAC) model of person identification (Burton, Bruce, & Johnston, 1990) is a computationally implemented extension of Bruce and Young’s (1986) model. Rather than discrete units and stages, the IAC model is based on connectionist architecture, where information is stored in pools of processing units. Within each pool, units are linked to each other in an inhibitory manner, whereas excitatory links form between pools. Activation cascades between pools and all links are bi-directional and equal strength. Three pools are particularly prominent in the model:

face recognition units (FRUs), semantic information units (SIUs) and person identity nodes (PINs). The FRU pool contains units of face representations (for how these representation are developed, see Burton, Bruce, & Hancock, 1999). Once a unit in the FRU pool is activated, it sends an excitatory signal to the relevant unit in the PIN pool, which is subsequently connected to several SIUs depending on what is known about the person. Names are considered to be a single unit in the pool of SIUs (Burton & Bruce, 1992). Compared to Bruce and Young (1986), familiarity is not signaled by the face per se, but by activation of a PIN. Once a PIN reaches a common threshold, a person is deemed familiar. Activation of a PIN allows for the relevant semantic information associated with that person to be accessed via SIUs. If the SIUs are blocked but the PIN and FRU is activated, then a face can generate a feeling-of-knowing without being able to recall a name or semantic information, or alternatively, if the FRU is blocked but the PIN and SIUs are activated, then a name can generate a feeling-of-knowing without being able to recall a face.

Implementations of the model have successfully accounted for the effects of semantic and identity priming (Burton, Bruce, & Johnston, 1990), covert recognition in prosopagnosia (Burton, Young, Bruce, Johnston, & Ellis, 1991), and name recall (Burton & Bruce, 1992). More importantly, Burton (1994) extended to the IAC model to account for how new faces are learned. However, he does not specify any mechanisms that would operate at a meta-level. In the IAC, face-name learning is an automatic process that does not make use of a feedback mechanism. If this model accurately reflects the underlying cognitive processes of face identification, then monitoring judgements for faces might not reflect actual performance.

The Effects of Distinctiveness in Faces

A robust finding in face recognition literature is the beneficial effects of facial distinctiveness on memory performance. Compared to typical faces, distinctive faces have lower response latencies (e.g. Ellis, Shepherd, Giblin, & Shepherd, 1988; Semmler & Brewer, 2006; Valentine & Bruce, 1986), more hits and fewer false alarms (e.g. Cohen & Carr, 1975; Light et al., 1979; Winograd, 1981), greater recognition accuracy (e.g. Brown & Lloyd-Jones, 2006; Courtois & Mueller, 1981; Semmler & Brewer, 2006) and greater retrospective confidence ratings (Semmler & Brewer, 2006). Thus, there is a consistent finding in the literature that memory for distinct faces is superior compared with typical faces.

"Face-space" models have been influential in explaining distinctiveness effects in faces (e.g. Lewis, 2004; Lewis & Johnston, 1999; Valentine, 1991; 2001). Briefly, these models posit that faces are represented as points in a Euclidean multi-dimensional space. The dimensions of the space reflect the properties of faces that allow them to be discriminated (e.g. nose length). The values of each dimension are normally distributed around some multivariate measure of central tendency, which serves as the origin in the space. Typical faces have values that are close to the measure of central tendency, whereas distinct faces have values that are outliers. Thus, distinct faces are located in regions of the space where the density of points is low, and therefore, recognition is easier because less noise is generated from neighboring points.

Bruce, Burton, and Dench (1994) provide strong support for "face-space" models. The authors measured a variety of physiognomic features of 175 faces that could serve as dimensions in face space (e.g. length of nose, ratio of pupil distance to eye-mouth distance). Based on these calculations, the mean value for a given physical measurement was obtained. An independent group of observers rated the 175 faces in terms of distinctiveness. The key finding was that distinctiveness ratings were positively correlated with standardized scores of the physiognomic measurements. In other words, distinct faces had greater physical deviations from the mean compared with typical faces, which is consistent with “face-space” models.

In addition to the memory-enhancing effects of distinctiveness, the phenomenological experience of face recognition also differs depending on the degree of typicality. Dual-process models of recognition memory assume that recognition judgements are based on either the explicit recollection of a specific episodic event or on an automatic feeling of familiarity that occurs in absence of explicit recollection (for a review, see Yonelinas, 2002). In order to measure the subjective experience accompanying recognition memory, Tulving (1985) developed the remember/know procedure, in which participants indicate whether their recognition decision was based on explicit recollection of the item (i.e. remember) or on a feeling of familiarity (i.e. know). By applying the remember/know procedure to face recognition tasks, researchers have found that distinct faces (Brandt, Macrae, Schloerscheidt, & Milne, 2003; Brown & Lloyd-Jones, 2006; Mantyla, 1997) are more likely to entail recognition decisions based on explicit recollection compared with typical faces. Conversely, typical faces (Brandt et al., 2003; Vokey & Read, 1992) are more likely to evoke a feeling of familiarity compared with distinct faces.

Similar findings were reported using principal components analysis (PCA) (Dewhurst, Hay, & Wichkham, 2005). Unfamiliar faces were rated on familiarity, distinctiveness, attractiveness, memorability, typicality, and resemblance to a familiar person. All ratings were made on a 7 point likert-type scale. Afterwards, the same set of faces was used in an old/new recognition test, and for every "old" recognition response, a remember or know judgement was solicited. Ratings on each of the aforementioned dimensions for remembered and known faces were then submitted to a PCA. The component that accounted for the most variance had positive loadings on distinctiveness and memorability for remember responses and a negative loading on typicality for remember responses. Hence, as the distinctiveness and memorability of a face increases and the typicality decreases, the face is more likely to be recognized on the basis of remembering rather than knowing.

Errors in Personal Naming

Diary studies reveal that recognizing a face without being able to recall a name is one of the most common identification errors (Cohen & Faulkner, 1986; Young, Hay, & Ellis, 1985). Furthermore, names appear to be more difficult to remember compared with other types of autobiographical information, such as occupations, birthplaces or hobbies (Young et al., 1985). Evidence from the laboratory supports this finding. For familiar faces, response latencies are quicker for occupation classification compared to name recall (Young, McWeeny, Ellis, & Hay, 1986), and identity-related semantic information is more likely to be recalled compared with names (Hanley & Cowell, 1988). For unfamiliar faces, several studies have reported that proper name recall is strikingly poorer compared with recalling autobiographical information (Cohen, 1990; Craigie & Hanley, 1997; McWeeny et al., 1987; Stanhope & Cohen, 1993). For example, when cued with a face, recognizing the appropriate name is contingent upon being able to remember an occupation, and recognition accuracy for occupations is significantly greater compared with names (Craigie & Hanley, 1997). These results suggest that face-name learning is mediated by additional biographical details, which is consistent with Bruce and Young’s (1986) model. What is perhaps even more surprising is that when controlling for the effects of contextual cues, frequency of usage, imageabiltiy, and meaningfulness of words used as names or occupations, names are still more difficult to recall than occupations. To illustrate, it is more difficult to recall that a person’s name is Mr. Baker compared to recalling that a person’s occupation is a baker (McWeeny et al., 1986).

Names also appear to be more difficult to recover compared with faces (Bahrick, 1984; Cohen & Faulkner, 1986; Young, Hay, & Ellis, 1985). For example, participants are better at matching face-name pairs when cued with names compared with faces, especially when the number of facial characteristics they can recall (i.e. vividness of the mental image of the face) increases from zero to four (Groninger, 2006). The same study also found that the vividness of mental images of faces is positively correlated with binary prospective confidence judgements in face-name matching. This could be due to the fact that the clarity of a mental image of a face is a useful cue for judging the likelihood of retrieving its corresponding name (Groninger, 2000). Indeed, the relative difficulty of retrieving names compared with faces could be due to the absence of imageable qualities in most names.

The relative difficulty of name retrieval is incorporated into models of person identification. According to Bruce and Young’s (1986) model, names can only be retrieved via PINs, thus autobiographical semantic information must be accessed beforehand. In support of the model, individuals in tip-of-tongue states can still recall occupations without access to names (Cohen & Faulkner, 1986). Furthermore, when cued with a face, response latencies are greater for naming compared with retrieving semantic information (Cohen, 1990; Craigie & Hanley, 1997; McWeeny et al., 1987; Stanhope & Cohen, 1993), presumably because there are no direct links between faces and names.

The IAC model also predicts that naming will be more difficult than retrieval of other types of autobiographical information (Burton & Bruce, 1992). Once a PIN is activated, an excitatory signal will cascade to SIUs connected to the PIN. The activation of these SIUs, in turn, will cascade to other PINs that are connected to them and the activation of these PINs will reverberate back to the SIUs. This has the effect of raising the level of activation of the SIUs that are shared among PINs. Proper names are a single unit in a pool of SIUs, and are usually connected to only one PIN. Occupations, on the other hand, are connected to multiple PINs. As a result, a name unit is less likely to reach asymptotic activation compared with other units in the SIU pool. To illustrate, multiple PINs can activate the 'philosopher' SIU unit (i.e. we know of many philosophers), but only one PIN will be connected to the 'Karl Popper' unit (i.e. we only know of one Karl Popper). Thus, occupations or other types of semantic information are more likely to be activated than name units, which results in a greater degree of difficulty in retrieving names.

In three experiments, Stanhope and Cohen (1993) tested several predictions of Bruce and Young's (1986) model and the IAC model (Burton & Bruce, 1992). According to former, name recall should only occur following access to semantic information. If names are paired with novel faces without additional autobiographical information, they should be harder to recall compared to when autobiographical information is present. The results of their first experiment do not support the model. Recall accuracy for names paired only with faces was significantly greater than names paired with occupations and faces. In contrast to the predictions of Bruce and Young's (1986) serial processing architecture, occupations did not facilitate access to names.

In their third experiment, Stanhope and Cohen (1993) manipulated the distinctiveness of names, such that a set of eight faces contained six common and two distinct names or six distinct and two common names. According to the IAC model, highly distinct names should be difficult to access because of the relatively low amount of connecting PINs compared with common names. In contrast to the predictions of the IAC model, cued-recall performance was better for distinct names compared to common names regardless of whether the majority or the minority of items were distinct. However, the difference between distinct and typical names on free recall accuracy was not significant. Nonetheless, the results of Stanhope and Cohens' (1993) experiments are difficult to reconcile with the assumptions of Bruce and Young's (1986) and Burton and Bruce's (1992) models of person identification.

One possible explanation as to why names are more difficulty to recall compared with identity-related semantic information is that they are arbitrary labels that do not convey any meaningful information. As a test of this hypothesis, Cohen (1990) compared recall performance for names and meaningless autobiographical information. Participants were instructed to learn names, occupations, and non-word semantic attributes to unfamiliar faces. Each face and its accompanying facts were studied for 30 seconds. Afterwards, a cued-recall test was presented with faces acting as cues. As predicted, recall accuracy did not significantly differ between names and non-word semantic attributes, and both were significantly below the level of performance for occupations. In support of the arbitrariness hypothesis, when personal names are manipulated to be meaningful, the likelihood of a retrieval error decreases (Bredart & Valentine, 1998) and recall accuracy improves relative to meaningless occupations (Cohen, 1990, Experiment 2).

Aside from meaningfulness, facial distinctiveness can influence the memorability of names. Valentine and Moore (1995) examined whether facial distinctiveness and surname frequency improves the rate of acquisition of face-name pairs. In the learning phase of their experiment, eight typical or distinct faces with a corresponding low-frequency or high-frequency surname were sequentially presented to participants. Thus, the experiment contained four conditions, typical faces with low-frequency surnames, typical faces with high-frequency surnames, distinct faces with low-frequency surnames, and distinct faces with high-frequency surnames. Surname frequency was manipulated by selecting the most frequent and infrequent surnames from a telephone directory. Participants learned the eight face-name pairs to a criterion of two correct name recalls per face. The process continued until all 32 face-name pairs were learned to the criterion. Afterwards, the test phase involved presenting a face for five seconds and asking participants to name the face as quickly as possible. Names paired with distinct faces required fewer trials to learn and were recalled more quickly compared with names paired with typical faces, regardless of surname frequency. These results point to the possibility that the distinctiveness of a face can influence recall of the associated name. Indeed, there is inverse relationship between the probability of correct name recall and the degree of similarity among faces in a study set (Pantelis et al., 2008).

If facial distinctiveness enhances memory for faces (e.g. Valentine & Bruce, 1986) and names (e.g. Valentine & Moore, 1995), then what effects does name distinctiveness have? Not surprisingly, distinct names are more likely to be recognized than typical names, and are more likely to be recognized on the basis of explicit recollection (Brandt, Gardiner, & Macrae, 2006). More importantly, name distinctiveness might facilitate memory for faces, in the same way the face distinctiveness enhances memory for names. The aforementioned findings of Stanhope and Cohen (1993) revealed that a distinctiveness effect in names is only evident on cued-recall. Thus, it is contingent on the presence of a face cue, which supports the idea that distinctiveness modulates the strength of the face-name association.

Summary of Chapters 1 – 3

Metacognition involves reflecting on one's own cognitive processes. Metamemory is the application of metacognition to memory. The framework developed by Nelson and Narens (1990) posits that monitoring and control are the fundamental components of metamemory. Montoring can be studied by soliciting prospective (i.e. EOLs, JOLs, and FOKs) and retrospective confidence judgements (RCJ) during various stages of learning. The accuracy of monitoring judgements can be determined by calculating a correlation coefficient between a monitoring judgement and performance on an objective memory test (i.e. relative accuracy) or by calculating the difference between a monitoring judgement and performance on an objective memory test (i.e. absolute accuracy).

The mechanisms underlying monitoring judgements are varied and are likely to be multiply determined. An underlying theme across all judgements is the distinction between trace-access and inferential mechanisms, and between implicit and explicit use of cues and heuristics. Interestingly, certain factors can dissociate memory performance and metacognitive monitoring, and therefore, cognitive illusions are likely to occur. This has significant relevance to understanding how faces can be falsely recognized or how faces can subjectively feel familiar even though a name does not come to mind.

Names appear to be more difficult to retrieve than faces or other types of identity-related information. Although several explanations for this have been proposed, none have gained widespread consensus. Seminal models of person identification posit than names can only be accessed after the identity of the face has been established. On the basis of these models, monitoring face-name associations might be very difficult, if not impossible. There is some evidence to suggest that memory for a name improves if it is distinct or if it is associated with a distinct face. However, it is yet to be determined whether the effects of distinctiveness are symmetrical between faces and names.

To date, the majority of metamemory studies have focused on verbal learning paradigms and relatively little attention has been given to face-name learning or comparing metamemory performance between categories of stimuli. The few studies that have solicited monitoring judgements during face-name learning reported that JOLs, FOKs, and RCJs are predictive of memory performance, but they are far from perfect. Also, these studies focused exclusively on name retrieval and neglected metamemory performance for faces in the presence of a name cue. While some evidence suggests that distinctiveness affects monitoring of faces and names, a full account of how distinctiveness impacts face-name learning is needed.

Chapter 4 – General Methods

While chapters one through three outlined the theoretical context and general rationale for conducting the present research, this chapter outlines the general procedure and the quantitative methods that were used to assess monitoring accuracy. Although most metacognitive studies usually solicit only one type of monitoring judgement, others have solicited multiple judgements during one (e.g. Busey et al., 2000; de Carvahlo & Yuzawa, 2001; Dougherty et al., 2005; Modirrousta & Fellows, 2008) or two testing sessions (Leonesio & Nelson, 1990). In order to gain a comprehensive profile of metamemory performance during face-name learning, the present experiments solicited up to four monitoring judgements. The procedure for soliciting each monitoring judgement is presented below.

General Procedure

The experiments reported in this dissertation were composed of six stages, which were adapted from Nelson and Narens' (1990) framework. An overview of the stages is presented in Figure 4. All of the experiments used paired-associates as stimuli. A paired-associate refers to a cue-target pair (e.g. a face-name pair or a noun-noun pair). Specific details concerning the paired-associates that were used in each experiment are provided in Chapters 5 and 6.

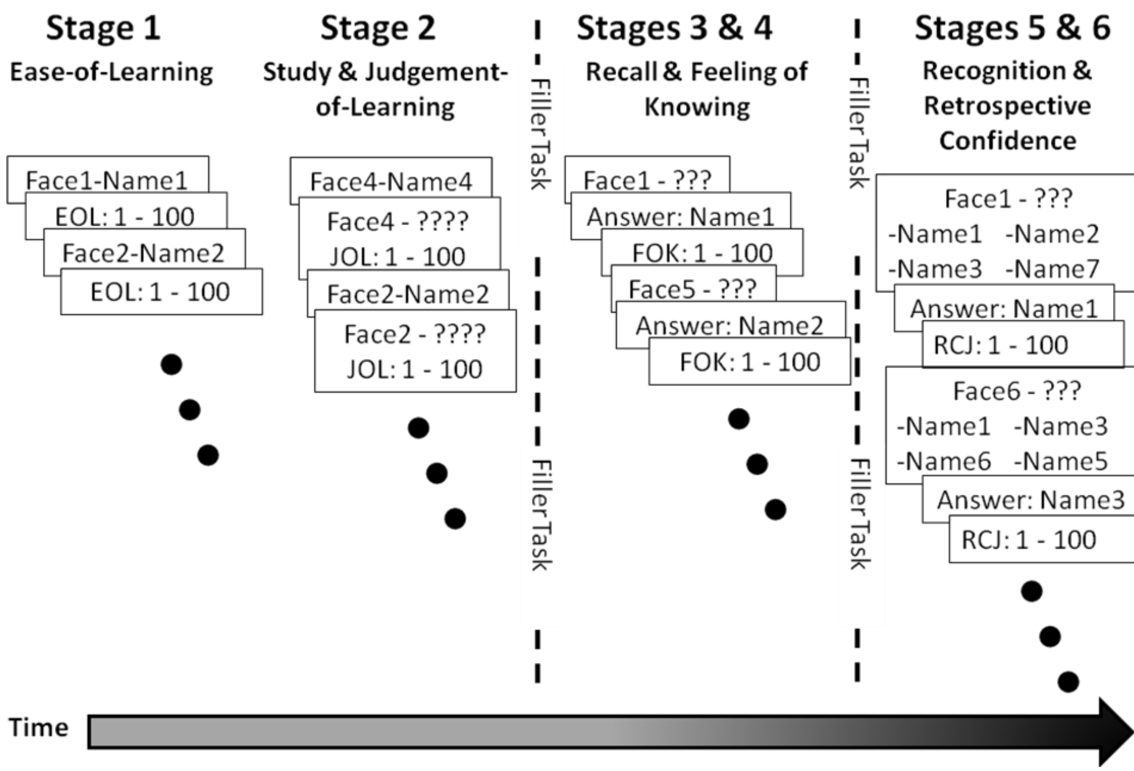

Figure 4. Overview of the general procedure. Adapted from Hunt and Ellis (2004).

For each experiment, participants were seated at a computer and presented with the general instructions. The general instructions informed participants that the experiment was composed of a series of stages that involve introspective judgements. The participants were encouraged to be as honest as possible when making their judgements. Each judgement was made on a 5-point 0-100 scale, using keys on a keyboard that were labeled 0, 20, 50, 80, and 100. Participants were given unlimited time to make their judgements and to provide responses on the recall and recognition tests. They were also encouraged to use the full range of the scale. Prior to beginning the experiment proper, participants were informed that they would complete a series of practice trials.

For the practice trials, participants completed three trials per stage and proceeded through each stage in a sequential manner, with the exception that they did not have to complete the filler tasks (described below). After the practice trials were complete, participants were given the opportunity to ask the researcher any questions. Otherwise, they could begin the experiment proper by pressing any key. During the experiment proper, the procedure was identical to the practice trials, with the exception that participants completed 36 trials per stage and the filler tasks were included in the procedure. Participants were informed that the experiment proper would contain more trials than the practice and would also include additional tasks (i.e. the filler tasks).

For name target conditions, faces were used as cues during the study-JOL, recall-FOK, and recognition stages. Participants were instructed to retrieve the name that was associated with the cued face. For face target conditions, names were used as cues during the study-JOL, recall-FOK, and recognition stages. Participants were instructed to retrieve the face that was associated with the cued name. For noun target conditions, the first word of a noun-noun pair was used as the cue during the study-JOL, recall-FOK, and recognition stages. Participants were instructed to retrieve the second noun that was associated with the cued noun. An overview of each stage is as follows.

EOL stage. After participants read the general instructions and pressed any key, a set of instructions for the EOL task appeared at the centre of the screen. Participants were instructed to rate the subjective difficulty of learning each paired-associate, so that when prompted with the cue, they would be able to remember the target sometime in the future. The ratings were made on a 0 – 100 scale, with 0 = very difficult, 20 = difficult, 50 = moderate, 80 = easy, and 100 = very easy. Participants were encouraged to make their judgement as quickly as possible. For each trial, a paired-associate was randomly selected and presented at the centre of the monitor for five seconds. After five seconds elapsed, an

EOL judgement was solicited at the bottom of the monitor. The paired-associate remained on the monitor until a response was made. Immediately following a response, a new paired-associate appeared on the monitor. The process continued until EOL judgements were made for all paired-associates.

Study-JOL stage. Immediately following the EOL stage, the order of the pairs was randomized anew and a set of instructions for the study-JOL stage appeared on the monitor. Participants were informed that the paired-associates from the previous stage would be presented to them again, but it in a different order. They were instructed to study each paired-associate so that they would be able to correctly remember the appropriate target when presented with the cue. They were also informed that following the study of each paired-associate, they would be immediately prompted to rate their confidence in their ability to correctly retrieve the target from memory in approximately five minutes. Thus, participants were not explicitly informed about the nature of the upcoming memory test (i.e. recall or recognition). Ratings were made on a scale from 0 to 100, with 0 = 0% confident, 20 = 20% confident, 50 = 50% confident, 80 = 80% confident, and 100 = 100% confident. On each trial, a paired-associate was presented on the monitor for five seconds. Immediately, after the offset of the item, a JOL was solicited. Only the cue was presented on the monitor when the judgement was solicited.

Recall stage. After the final JOL, participants completed an unrelated filler task for two minutes. The filler task involved counting backwards in intervals of three from a series of numbers presented at the centre of the screen. Following the filler task, the order of the paired-associates was randomized anew and a cued-recall test was presented. A cue was presented at the centre of the monitor and participants were required to recall the appropriate target. For name and noun target conditions, participants were instructed to type the corresponding name or noun and hit enter. If they could not recall the appropriate target, they were instructed to leave the response field blank and hit enter.

Recall accuracy for face targets cannot be objectively determined. However, considering that ratings of mental image clarity of faces are correlated with the strength of face-name associations in memory (Groninger, 2000; 2006), imagery ratings and catch trials were used to approximate face recall. For the presentation of every name cue during the cued-recall task, participants were asked to rate the clarity of their mental image of the corresponding face. Ratings were made on a scale from 0 to 100, with 0 = no mental image at all, 20 = poor mental image clarity, 50 = moderate mental image clarity, 80 = clear mental image, 100 = very clear mental image.

Following the imagery rating for every 3rd name cue, an image of a face was presented and participants were required to indicate whether the face that was presented on the monitor was the same face that they imaged. Thus, out of 36 trials, 12 involved assessing whether the imaged face was correct. The correct face was presented on one half of the catch trials whereas an incorrect face was presented on the other half. The incorrect faces were sampled from the study list. The order in which correct or incorrect faces were presented was randomly determined. If performance on the catch trials was below 25%, then the participant's data from the imagery task was not included in the analyses. However, none of the participants in any of the experiments performed below 25% on the catch trials.

FOK stage. Immediately after every recall or imagery response, an FOK was solicited. The instructions for the FOK task were presented alongside the instructions for the cued-recall test/imagery instructions. The instructions for the FOK task informed participants that after every recall or imagery response, they would be prompted to rate their confidence in their ability to recognize the appropriate target from a list of eight alternatives. Ratings were made on the same scale as the JOL task. The cue was presented on the monitor during the FOK judgement.

The recall-FOK process continued until FOK judgements were made for all items. Thus, in accordance with the procedure established by Koriat (1993; 1995), FOKs were solicited for all items rather than for a subset of items that were not recalled. This prevented participants from receiving feedback on the accuracy of their responses during the cued-recall test. Following the completion of the FOK stage, participants completed the same unrelated filler task for two minutes, but with a different series of numbers.

Recognition stage. After the filler task, an 8-AFC associative recognition test was administered. The order of the paired-associates was randomized anew. Prior to the beginning of the test, participants were instructed to select the appropriate target that was associated with the cue during the beginning stages of the experiment, even if they had to guess. For each trial, a cue, a target, and seven randomly selected distracters were presented. The distracters consisted of previously studied items. The cue was presented at the bottom of the monitor and the distracters were presented simultaneously at the top of the monitor, along with a number beside each distracter. Participants could select any of the alternatives by pressing its corresponding number on the keypad. The location of the target in the distracter array varied from trial to trial.

RCJ stage. Immediately after each recognition response, an RCJ was solicited. The instructions for the RCJ task were presented alongside the instructions for the recognition test. The instructions for the RCJ task informed participants that after every recognition response, they would be prompted to rate their confidence in the accuracy of their answer. The ratings were made on the same scale used in the FOK and JOL tasks. The recognize-RCJ process continued until recognition responses and RCJs were made for all paired-associates. After the final RCJ, participants were informed that the experiment was complete.

Measures of Monitoring Accuracy

Relative accuracy was operationalized as the Goodman-Kruskal gamma correlation (γ) between the magnitude of a monitoring judgement and accuracy on the cued-recall or associative recognition test. Gamma is a measure of association for variables on ordinal scales. It compares the difference between the number of concordant pairs[4] (e.g. Item A has a greater rank on variable X and variable Y compared with Item B) and disconcordant pairs (e.g. Item A has a greater rank on variable X but not variable Y compared with Item B) relative to the total number of concordant and disconcordant pairs. Formally, γ is defined as:

$$\gamma = \frac{c - d}{c + d},$$

where c and d represent concordant and disconcordant pairs, respectively. Ties are not included in the calculation of gamma. Gamma can range from -1 to 1, with 0 indicating no association, or in the context of metamemory, monitoring at chance level. A value of 1

[4] Concordant pairs are instances of accurate metamemory monitoring. A pair of items is concordant when Item A has a greater estimated likelihood of being remembered compared with Item B, and Item A is actually remembered, whereas item B is not. Disconcordant pairs are instances of inaccurate metamemory monitoring. A pair of items is disconcordant when item A has a greater estimated likelihood of being remembered compared with item B, but Item A is not remembered, whereas item B is remembered.

indicates that a participant can distinguish between items that will and will not be retrieved with almost perfect accuracy.

As Nelson (1984) points out, gamma is ideal for characterizing the relationship between monitoring judgements and objective memory performance. Unlike signal detection theory, it makes no assumptions regarding the underlying psychological distributions of monitoring judgements or memory strength. Furthermore, it is logically independent of memory performance, and has a straightforward psychological interpretation: the probability that Item A is recalled and Item B is not, given that Item A received a greater estimated likelihood of being recalled compared to Item B can be calculated through a linear transformation[5] of γ.

Recently, gamma has been criticized as a suitable statistic for evaluating hypotheses that require interval-level assumptions (e.g. whether relative accuracy is superior for one condition compared with another) because it is bounded between -1 and 1 (Benjamin & Diaz, 2008). As recommend by Benjamin and Diaz (2008), a monotonic transformation of each participant's γ was conducted to improve γ's metric qualities so that it becomes more suitable for between-group comparisons. The transformation is called G* and is calculated via the following equation:

$$G^* = \log\left(\frac{\gamma+1}{1-\gamma}\right).$$

For participants whose γ in a given condition is at ceiling, G* cannot be calculated.

In order to remain consistent with the majority of previous metamemory research, but also to account for the drawbacks of using γ, both γ and G* are reported throughout

[5] The transformation is called V, and is defined by: $V = .5\gamma+.5$

the experiments. For each monitoring judgement in each experimental condition, γ was used to determine if participants could monitor their memory above chance level. This was accomplished by comparing the mean magnitude of γ with 0 using a single sample t-test. For each monitoring judgement, a Bonferonni correction was used when testing if γ was significantly greater than zero. For comparisons of relative accuracy between experimental conditions, both γ and G* were used as dependent variables.

Absolute accuracy was operationalized as the linear parameters of a calibration line (see Nelson & Dunlosky, 1991) and the difference between the mean magnitude of a monitoring judgement and mean performance on an object memory test (see Tauber & Rhodes, 2010). As Figure 5 illustrates, a calibration line depicts the mean observed performance on a memory test as a function of each interval of estimated performance on a memory test, collapsing across all participants in a given condition. A least-squares linear equation was fit to the calibration line and the values of the slope and intercept were used to assess absolute accuracy. Perfect absolute accuracy occurs when the slope equals one and the intercept equals zero.

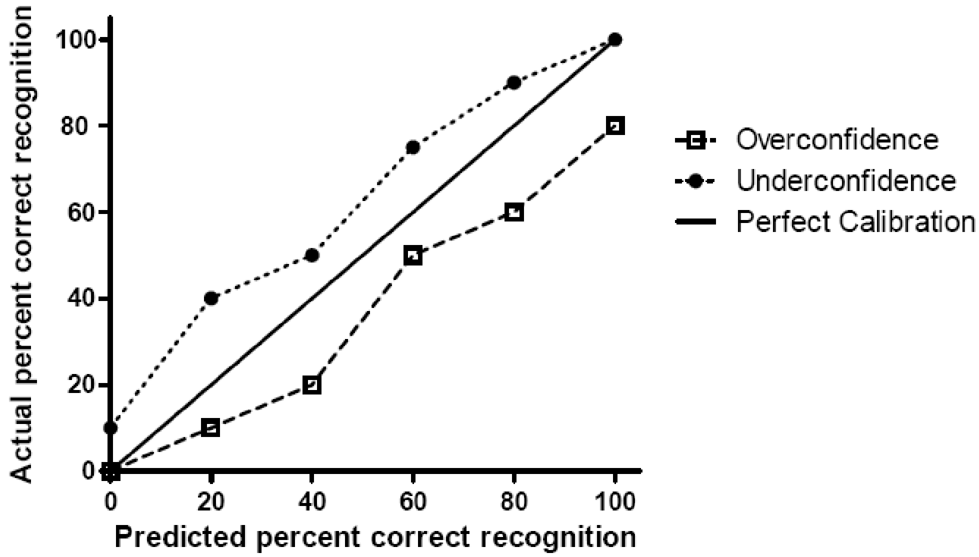

*Figure 5***.** Sample calibration curves demonstrating over and underconfidence.

Although calibration lines are useful for visually determining over and under confidence, they are particularly resistant to inferential statistics. As a result, the second measure of absolute accuracy tested the statistical significance of the difference between the mean magnitude of a monitoring judgement and mean performance on the objective memory tests in each experimental condition. If the difference is significant, then absolute accuracy is considered to be poor and the sign of the difference can be used to determine over or under confidence. For example, if the mean magnitude of JOLs is significantly greater than the mean cued-recall accuracy, then participants are overconfident in their estimates of future memory performance. Absolute accuracy was not calculated for EOL judgements and imagery ratings because these responses were not on a probabilistic scale (Keren, 1991).

Chapter 5 – Experiments 1 and 2: Metamemory for faces, names, and common nouns

Experiment 1

To date, the accompanying metamemory processes of face-name learning have largely gone unnoticed. Seminal models of person identification either relegate metacognitive processes to a comprehensive but unspecified cognitive system (Bruce & Young, 1986) or do not include them at all in the architecture (Burton et al., 1990). Yet, recent research reveals that difficulties with proper name learning can be attributed to inaccurate monitoring during study (Tauber & Rhodes, 2010), and false facial recognition in patients with frontal lobe damage might stem from impairments in monitoring (Rapcsak et al., 1999). Thus there is some evidence to suggest that metamemory is implicated during face-name learning, and this warrants additional research.

Although a handful of studies have solicited monitoring judgements using face-name pairs (Hosey et al., 2009; Modirrousta & Fellows, 2008; Pannu et al., 2005; Tauber & Rhodes, 2010), disparities in the selection of stimuli (i.e. familiar faces versus unfamiliar faces), in the type of retrieval task (i.e. cued-recall versus associative recognition), in the monitoring judgements that were solicited (i.e. JOLs versus FOKs versus RCJs), and in the measures of monitoring accuracy (i.e. relative accuracy versus absolute accuracy) among these studies make it difficult to elucidate metamemory performance during encoding and retrieval of novel face-name associations. To cite one example, the relative accuracy of JOLs reported in Modirrousta and Fellows (2008) was based on an aggregate of items rather than on an item-by-item basis, which limits the generalizability of their findings. Moreover, these studies focused on name retrieval and neglected the monitoring processes involved in retrieving a recently learned face when cued with a name. Thus, there are aspects of monitoring face-name learning that remain unclear.

One question that has yet to be addressed is whether monitoring accuracy for face-name associations differs depending on the type of target during retrieval. For instance, memory for names is typically poorer compared with faces (e.g. Bahrick, 1984; Cohen & Faulkner, 1986; Young, Hay, & Ellis, 1985), but participants might not be aware of this fact and they may place a greater degree of confidence in their ability to remember names than what is warranted, which would attenuate their metamemory accuracy and could ultimately prompt sub-optimal study and retrieval strategies for names compared with faces. Indeed, on the basis of inferential theories of monitoring judgements (Koriat, 1993; 1995; 1997, for a review see Koriat, 2007), monitoring accuracy should be superior in situations where a face must be retrieved from memory when prompted with a name compared with situations where a name must be retrieved when prompted with a face. Name cues can prompt a mental image of a face target, which contributes to the efficacy of JOLs (Begg et al., 1989) and strengthens the association between faces and names (Groninger, 2006). Furthermore, face targets have fewer competing alternatives in memory compared with name targets (i.e. individuals often share the same name but rarely share the same face), which could potentially increase the accuracy of monitoring judgements for face targets compared with name targets.

In contrast to the aforementioned studies, which typically solicited monitoring judgements at only one or two stages of learning, and did not compare monitoring accuracy between faces and names, the primary aim of this experiment was to produce a comprehensive profile of monitoring performance for novel face-name associations at multiple stages of learning. In particular, participants learned a series of face-name pairs, and were solicited with a monitoring judgement prior to learning (i.e. EOLs), after learning (i.e. JOLs), after a recall attempt (i.e. FOKs) and after a recognition response (i.e. RCJs). In one condition, the study and retrieval tasks involved face cues and name targets, whereas in another condition, the tasks involved name cues and face targets. This was done in order to determine whether participants can monitor their memory for novel face-name associations at various stages of learning, and whether the accuracy of monitoring judgements differs between face and name targets.

A secondary aim of this experiment was to assess differences in monitoring performance between traditional stimuli (i.e. noun-noun paired-associates) and face-name paired-associates. To this end, a noun-noun[6] condition was included in the design of the experiment as a control. By doing so, monitoring performance between face-name pairs and noun-noun pairs could be compared, which would speak to the validity of generalizing metamemory findings from one category of stimuli to another. It should be noted that face-name pairs and noun-noun pairs vary on a number of dimensions (e.g. episodic versus semantic; visual versus verbal; spatial and visual characteristics), and the goal is not to single out a particular dimension as being diagnostic for differentiating monitoring performance between one category of stimuli with another; rather, to determine whether, at a gross level of comparison, there would be any differences in monitoring between conditions that used traditional stimuli and face-name paired-

[6] For the present purposes, noun refers to a common noun rather than a proper name.

associates. As mentioned in Chapter 1, only a few studies have compared monitoring performance between different categories of stimuli (Bornstein & Zickafoose, 1999; Perfect, Watson, & Wagstaff, 1993; Perfect & Hollins, 1996), and have reported conflicting results. Directly comparing metamemory performance between face-name pairs and noun-noun pairs could shed some light on this issue.

A second reason for including noun-noun pairs in the experimental design was to ensure that the delayed-JOL effect could be replicated with traditional stimuli prior to investigating the effect with face-name stimuli. JOLs made after a delay are more accurate predictors of retrieval performance than JOLs made immediately following study (Connor et al., 1997; Dunlosky & Nelson, 1992; 1994; Nelson & Dunlosky, 1991; Nelson et al., 2004; Thiede & Dunlosky, 1994). However, this effect has yet to be replicated with face-name pairs, or indeed with any stimuli outside the semantic realm. If the delayed-JOL effect transfers to face identification tasks, it would provide some insight into whether delayed reports of prospective confidence from an eyewitness should be considered trustworthy predictions of future memory performance, and would lend credence to the hypothesis that the same metamemory mechanisms operate over different categories of stimuli (Diana & Reder, 2004).

This experiment is designed to test five hypotheses. First, on the basis of previous findings from studies that used faces or names as targets (Busey et al., 2000; Hosey et al., 2009; Modirrousta & Fellows, 2008; Pannu et al., 2005; Sommer et al., 1995; Tauber & Rhodes, 2010), it is expected that participants can monitor their memory while learning novel face-name associations. In particular, the gamma correlation between EOLs, JOLs, FOKs, or RCJs and cued-recall or recognition accuracy for face and name targets are expected to be significantly greater than zero. Second, it is expected that relative accuracy will be superior for face targets compared with name targets. If memory for names is poorer compared with faces, as suggested by previous findings (e.g. Bahrick, 1984; Cohen & Faulkner, 1986; Young, Hay, & Ellis, 1985), and if participants are unaware of this deficit, then the magnitude of the gamma correlations should be greater for face targets compared with name targets.

Third, the magnitudes of the monitoring judgements are expected to be greater for nouns compared with faces and names. Nouns are thought to belong to an associative semantic network (Collins & Quillian, 1969; Quillian, 1967), which is likely to contribute to encoding fluency (Koriat & Ma'ayan, 2005), retrieval fluency (Kelley & Lindsay, 1993), accessibility of partial information relating to the target (Koriat, 1993; 1995), or cue familiarity (Reder, 1987), all of which have been found to increase the magnitude of monitoring judgements. In contrast, the episodic and idiosyncratic nature of unfamiliar faces, and the inherently meaningless (Cohen, 1990) and pure referring (Semenza & Zettin, 1989) nature of names are unlikely to provide similar contributions to the heuristics that underlie monitoring judgements.

Fourth, it is expected that monitoring accuracy will be superior for noun targets compared with face and name targets. The aforementioned findings of Perfect and colleagues (Perfect et al., 1993; Perfect & Hollins, 1996) suggest that the episodic aspects of faces and names might attenuate relative accuracy compared with nouns. Names in particular are expected to have the poorest monitoring accuracy. In general, names are more difficult to retrieve compared with nouns (e.g. Burke, MacKay, Worthley, & Wade, 1991; Reason & Lucas, 1984) or identity-related semantic information (e.g. Craigie &

Hanley, 1997), and participants do not seem to be aware of this deficit (Tauber & Rhodes, 2010), thus dissociations between monitoring judgements and memory performance are likely to be greater for name targets compared with noun targets.

Finally, in line with previous studies that used noun-noun paired-associates, (e.g Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991; Thiede & Dunlosky, 1994), delayed JOLs are expected to be more accurate predictors of memory performance than immediate JOLs for both face and name targets.

Method.

Participants.

The participants were 75 undergraduate students from the University of Ottawa (56 Females and 19 Males). The mean age of the participants was 18.9 years. Each participant received course credit for participating. All participants were tested individually.

Design.

A 3 × 2 mixed factorial design was employed, with target condition (noun vs. name vs. face) as a between-subject factor and JOL-type (immediate vs. delayed) as a within-subject factor. The participants were randomly assigned to the between-subject conditions, with twenty-five participants in each condition.

Materials and apparatus.

The experiment was run with Matlab 5.2 (Mathworks Inc., Natick, MA; http://www.mathworks.com) on a Macintosh Dual 2 GHz PowerPc G5. Stimuli were presented on an NEC AccuSync 95F 17” monitor.

In the face and name target conditions, 44 Caucasian male face images were used in the experiment. The images were obtained from an online database at the University of Stirling’s Psychology Department (http://pics.psych.stir.ac.uk), and all of the images were cropped to depict the neck, face and hairstyle. None of the faces contained particularly distinct features (e.g. scars, facial hair, glasses, etc). All of the faces appeared to be from the same age range (21-35 years) and depicted a neutral expression. All faces were presented in the frontal (0°) view. Each image was reduced to 8-bit grey-scale and scaled so the image dimensions were approximately 5 cm in width and 7 cm in height (visual angle of 5° x 7° at the viewing distance of 57 cm). All face images were imbedded in a 256 x 256 pixel uniform black background. Eight of the face images were used during practice trials, whereas the remaining 36 images were used during the experiment proper.

In the face and name target conditions, 44 male first names were used in the experiment. For the practice trials, eight male names were associated with the eight practice faces. For the experiment proper, the remaining 36 male names were randomly associated with the 36 faces. Of the 36 names, one third of the names were three letters long, another third were four letters long and the remaining third were five letters long. All of the names were taken at random from the US census bureau and reflect North American naming conventions (e.g. Ian, Tim, Mike, Dale). For each participant, each name was randomly paired with each face during the experiment proper, resulting in 36 unique face-name pairs per participant. All of the names were presented in a white Arial font, size 21.

In the noun target condition, 88 nouns were used as stimuli. For the practice trials, eight nouns were randomly paired with another unrelated eight nouns. For the experiment proper, 72 nouns were selected from Paivio, Yuille, and Madigan's (1968) classification of 925 nouns on imagery, meaningfulness, and concreteness. The 72 nouns were selected at random without replacement, which produced the following means and standard deviations on each of the aforementioned dimensions: imagery ($M=5.15$, $SD = 1.39$), meaningfulness ($M = 6.07$, $SD = 1.05$), and concreteness ($M = 5.12$, $SD = 1.19$). For each participant, each noun was randomly paired with another noun during the experiment proper, resulting in 36 unique noun-noun pairs per participant (e.g. fork – engine). All of the pairs were presented in a white Arial font, size 21.

Procedure.

The procedure of Experiment 1 was identical to the general procedure outlined in Chapter 4, with the exception that the study-JOL stage was revised to accommodate delayed JOLs. Aside from this, Experiment 1 did not deviate from the general procedure. A description of the study-JOL stage for Experiment 1 is as follows.

Study-JOL stage. Immediately following the EOL stage, the order of the pairs was randomized anew and instructions for the study-JOL stage appeared. Participants were informed that the paired-associates from the previous stage would be presented to them again, but it in a different order. They were instructed to study each paired-associate so that they would be able to correctly remember the appropriate target when presented with the cue. They were also informed that following the study of each paired-associate, they would be prompted to rate their confidence in their ability to correctly retrieve the target from memory in approximately five minutes. Thus, participants were not explicitly informed about the nature of the upcoming memory test (i.e. recall or recognition). Ratings were made on a scale from 0 to 100, with 0 = 0% confident, 20 = 20% confident, 50 = 50% confident, 80 = 80% confident, and 100 = 100% confident. The participants were informed that the confidence judgements would be solicited immediately after studying the item or after a slight delay.

During the practice, the first trial was always a delayed JOL. For the experiment proper, half of the 36 paired-associates were randomly assigned to receive immediate JOLs whereas the other half received delayed JOLs. On each trial, a paired-associate was presented on the monitor for five seconds. For items slated to receive immediate JOLs, judgements were solicited immediately after the offset of the item. For items assigned to the delayed JOL condition, judgements were solicited only if the following criteria were satisfied. First, all items assigned to the immediate JOL condition were studied and received a JOL. Second, all items assigned to the delayed JOL condition were studied. Thus, for delayed JOLs, the interval between study and solicitation of the JOL was filled with studying and completing JOLs for items assigned to the immediate JOL condition and studying items assigned to the delayed JOL condition. For both immediate and delayed JOLs, only the cue was presented on the monitor when the JOL was solicited. Of the 18 items in the delayed JOL condition, JOLs for the first six items that were studied were solicited in a random order, followed by JOLs for the next six items that were studied, followed by JOLs for the last six items that were studied.

Results.

The results of associative memory tests are presented first, followed by the magnitude of monitoring judgments, and then measures of relative accuracy. Analyses concerning absolute accuracy are presented last.

Descriptive statistics for memory and metamemory performance across target conditions are presented in Tables 1 and 2, respectively. Figure 3 depicts the mean magnitude of the monitoring judgements across target conditions. Any outliers that were detected based on the criteria developed by Van Selst and Jolicoeur (1994) were removed from the analysis[7]. All hypothesis tests concerning memory performance or the magnitude of monitoring judgements were conducted with alpha set at .05.

Table 1

Descriptive Statistics for Cued-Recall and Associative Recognition Accuracy across Target Conditions in Experiment 1

	Target Condition								
	Names			Faces			Nouns		
Memory Test	*M*	*SEM*	*N*	*M*	*SEM*	*N*	*M*	*SEM*	*N*
Cued-Recall	0.26	0.03	25	-	-	-	0.53	0.05	25
Associative Recognition	0.53	0.04	25	0.67	0.04	24	0.82	0.04	24

Note. Empty cells occur in conditions where accuracy on the memory test could not be objectively determined (see Chapter 4 for details).

Memory performance.

Bear in mind that cued-recall accuracy could not be objectively determined for face targets, thus the face target condition was omitted from all analyses involving cued-recall accuracy. The results for the mental imagery ratings for face targets are presented

[7] Tables 1 and 2 contain the sample sizes in each condition for each dependent variable after outliers were removed.

below. On average, participants accurately recalled more noun targets compared with name targets, $t(48) = 4.57$, $p < .001$, $r^2 = .30$. A finer-grained analysis of the type of errors that were committed in each target condition was conducted using the non-parametric Mann-Whitney test[8]. The number of commission errors in the name target (*Mdn* = 6.00) condition was significantly greater than in the noun (*Mdn* = 1.00) target condition, U = 63.500, $z = -4.22$, $p < .001$, $r^2 = 0.79$ whereas the difference between name (*Mdn* = 18.00) and noun (*Mdn* = 14.00) targets on the amount of omission errors was not significant, U = 221.50, $z = -.48$, $p > .05$. Hence, participants were more likely to recall an incorrect name compared with an incorrect noun. The mean response latency on the cued-recall test was significantly lower in the noun target condition compared with the name target condition, $t(46) = 3.21$, $p = .002$, $r^2 = .18$. Altogether, these results reveal that cued-recall performance was superior for noun-noun pairs compared with face-name pairs.

A one-way analysis of variance (ANOVA) revealed a significant effect of target condition on recognition accuracy, $F(2,70) = 15.35$, $p < .001$, $\omega^2 = 0.29$. A Scheffe post-hoc test revealed that recognition accuracy was greater in the noun target condition compared with the name ($p < .001$) or face ($p = .021$) target conditions. Furthermore, as anticipated, recognition accuracy was significantly greater for face targets compared with name targets ($p = .034$). Response latencies also varied among target conditions, $F(2,70) = 13.81$, $p < .001$, $\omega^2 = 0.26$. The results of a Scheffe post-hoc test indicate that noun targets were more rapidly recognized than face ($p < .001$) or name ($p = .002$) targets. Like cued-recall, associative recognition was superior for nouns compared with faces and names.

---

[8] Parametric statistics were unsuitable because of violations of the assumption of normality.

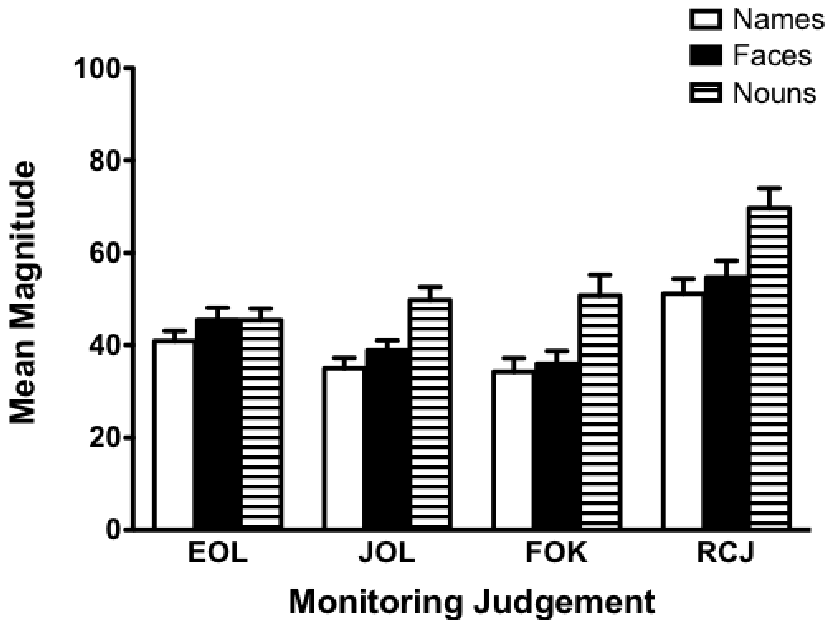

Figure 6. The mean magnitude of monitoring judgements across target conditions. EOL, ease-of-learning judgements; JOL, judgements-of-learning collapsing across immediate and delayed conditions; FOK, feeling-of-knowing judgements for all items; RCJ, retrospective confidence judgements. Error bars represent 1 SEM.

Magnitude of the monitoring judgements.

Ease-of-learning judgements. The difference between target conditions on the mean magnitude of EOL judgements was not significant, $F(2,71) = 1.20$, $p > .05$.

Judgements-of-learning. In contrast to EOLs, the mean magnitude of JOLs significantly differed among target conditions $F(2,71) = 9.78$, $p < .001$, $\omega^2 = 0.15$. A Scheffe post-hoc test revealed that the mean magnitude of JOLs in the noun target condition was significantly greater than the mean magnitude of JOLs in the name ($p <$ .001) and face target conditions ($p = .011$). A separate analysis was conducted to determine if the magnitude of immediate and delayed JOLs varied among target conditions. A 2 (JOL-type) × 3 (target condition) mixed ANOVA with JOL-type as a within-subject factor revealed that the main effect of target condition was significant, $F(2,70) = 9.07$, $p < .001$, partial $\eta^2 = 0.21$, whereas the main effect of JOL-type approached significance, $F(1,70) = 3.59$, $p = .062$, and the interaction between JOL-type and target condition was not significant, $F(2,70) = 1.86$, $p > .05$. Thus, as expected, participants are more confident in their ability to remember noun targets compared with face or name targets. This occurs regardless of the time at which the JOL was solicited.

Feeling-of-knowing judgements. There was a significant difference among target conditions on the mean magnitude of FOKs, $F(2,71) = 6.65$, $p = .002$, $\omega^2 = 0.11$. A Scheffe post-hoc test revealed that the mean magnitude of FOKs was greater in the noun target condition compared with the name ($p = .006$) or face ($p = .017$) target conditions. Thus, participants believe that noun targets are more likely to be recognized than face or name targets. In keeping with previous studies and the traditional definition of FOKs, I also analyzed the magnitude of FOKs for the subset of items that could not be recalled during the cued-recall test. The difference between nouns (M = 15.56, SE = 2.1) and names (M = 16.18, SE = 2.4) on the magnitude of FOKs for these items was not significant, $t(48) = .19$, $p > .05$.

Imagery ratings[9]. For the face target condition, the mean imagery rating (M = 33.96, SE = 2.73) was significantly greater than zero, $t(24) = 12.34$, $p < .001$, $r^2 = .86$. According to the imagery rating scale (see Chapter 4, Pg. 62), this indicates that on average, participants could image the target face and the clarity of the mental image was poor to moderate.

---

[9] All analyses in all experiments involving imagery ratings exclude data from catch trials. Thus, imagery ratings are based on 24 trials per participant rather than 36.

Retrospective confidence judgements. There was a significant effect of target condition on the magnitude of RCJs, $F(2,72) = 7.02$, p = .002, $\omega^2 = 0.14$. A Scheffe post-hoc test revealed that the mean magnitude of RCJs was significantly greater in the noun target condition compared with the name ($p = .003$) or face ($p = .020$) target conditions. Thus, participants are more confident in the accuracy of their recognition decisions for noun targets compared with face or name targets. Altogether, the pattern of the magnitude of the monitoring judgements suggests that participants are more confident in their memory for nouns compared with faces and names. For all monitoring judgments other than EOLs, magnitudes were greater for nouns than they were for face or name targets, which did not differ from one another.

Relative Accuracy.

The relative accuracy of a given monitoring judgement was calculated for each participant in each target condition. Three participants in the name target condition and one participant in the noun target condition scored below 5% on the cued-recall test. In addition, eight participants in the noun target condition scored above 95% on the recognition test. These participants were not included in the analyses of relative accuracy because of the limited variability in their data (see Connor et al., 1997). Descriptive statistics for the relative accuracy results are presented in Table 2.

Table 2

Descriptive Statistics for Relative Accuracy Across Target Conditions in Experiment 1

	Target Condition								
	Names			Faces			Nouns		
Relative Accuracy (γ)	*M*	*SEM*	*N*	*M*	*SEM*	*N*	*M*	*SEM*	*N*
EOL-recall	0.53	0.07	21	-	-	-	0.50	0.06	24
EOL-recognition	0.33	0.06	24	0.32	0.06	24	0.47	0.07	17
JOL-recall	0.82	0.03	21	-	-	-	0.74	0.05	23
JOL-recognition	0.55	0.04	24	0.49	0.06	25	0.48	0.09	17
FOK-recognition	0.66	0.04	22	0.64	0.06	25	0.76	0.05	17
FOKerr-recognition	0.31	0.06	20	-	-	-	0.33	0.13	16
Imagery-recognition	-	-	-	0.66	0.06	25	-	-	-
RCJ-recognition	0.78	0.03	24	0.84	0.01	23	0.91	0.02	15

Note. Empty cells occur in conditions where a monitoring judgement was not solicited or where accuracy on the memory test could not be objectively determined (see Chapter 4 for details). EOL, ease-of-learning judgements; JOL, judgements-of-learning collapsing across immediate and delayed conditions; FOK, feeling-of-knowing judgements for all items; FOKerr, feeling-of-knowing judgements for items that were not recalled; Imagery, mental image clarity ratings; RCJ, retrospective confidence judgements. The suffixes -recall and -recognition refer to the objective memory test that was used in the calculation of Goodman-Kruskal's gamma (γ).

Ease-of-learning judgements. Two one sample t-tests revealed that the mean gamma correlation between cued-recall accuracy and the magnitude of EOL judgements for name and noun target conditions were significantly above zero, $t(20) = 7.95$, $p < .001$, $r^2 = .76$, $t(23) = 8.30$, $p < .001$, $r^2 = 0.76$, respectively. Likewise, three one sample t-tests revealed that the mean gamma correlation between associative recognition accuracy and the magnitude of EOL judgements was significantly above zero for name, $t(23) = 5.65$, $p < .001$, $r^2 = 0.58$, face, $t(23) = 5.60$, $p < .001$, $r^2 = 0.57$, and noun, $t(16) = 6.35$, $p < .001$, $r^2 = 0.72$, target conditions. Thus, estimates of learning difficulty are predictive of future recall and recognition performance for all three categories of targets.

An independent sample t-test revealed that the mean gamma correlation between cued-recall accuracy and EOLs did not significantly differ between noun and name target conditions, $t(43) < 1$. Likewise, the mean G* between cued-recall accuracy and EOLs did not significantly differ between noun and name target conditions, $t(43) < 1$. Turning to the predictive validity of EOLs on associative recognition accuracy, two one-way ANOVAs revealed that the mean gamma, $F(2,62) = 1.52$, $p > .05$, or G*, $F(2,60) < 1$, did not significantly differ among target conditions.

Judgements-of-learning. Collapsing across immediate and delayed JOLs, two one sample t-tests revealed that the mean gamma correlation between cued-recall accuracy and JOL magnitude for name and noun target conditions was significantly above zero, $t(20) = 27.93$, $p < .001$, $r^2 = 0.96$, and $t(22) = 13.62$, $p < .001$, $r^2 = 0.88$, respectively. Similarly, collapsing across immediate and delayed JOLs, three one sample t-tests revealed that the mean gamma correlation between JOL magnitude and associative recognition accuracy was significantly above zero for name, $t(23) = 15.47$, $p < .001$, $r^2 = 0.90$, face, $t(24) = 8.37$, $p < .001$, $r^2 = 0.74$, and noun, $t(16) = 5.49$, $p < .001$, $r^2 = 0.64$, target conditions. Thus, JOLs are predictive of future memory performance, regardless of the category of stimuli.

The mean gamma correlation between JOLs and cued-recall accuracy, collapsing across immediate and delayed conditions, did not significantly differ between name and noun target condition, $t(41) < 1$. Using G*, the difference between name and noun target conditions on the relative accuracy of JOLs on cued-recall was not significant, $t(42) < 1$. The predictive validity of JOLs on recognition accuracy revealed a similar pattern. Two one-way ANOVAs revealed that the mean gamma, $F(2,63) < 1$, or the mean G*, $F(2,60) < 1$, did not significantly differ among target conditions.

The delayed-JOL effect. The mean gamma correlations for immediate and delayed JOLs with cued-recall and associative recognition accuracy across target conditions are presented in Figure 7. As can be seen in Figure 7, the correlation between delayed JOLs and cued-recall accuracy is greater than the correlation between immediate JOLs and cued-recall accuracy in both noun and name target conditions. A Kolmogorov-Smirnov test revealed that the distribution of gamma correlations between delayed JOLs and cued-recall accuracy in the noun, $D(20) = .33$, $p < .05$, and name, $D(19) = .24$, $p < .05$, target conditions were significantly non-normal. Thus, for both target conditions, non-parametric statistics were conducted to determine if delayed JOLs are more accurate at predicting cued-recall performance compared with immediate JOLs.

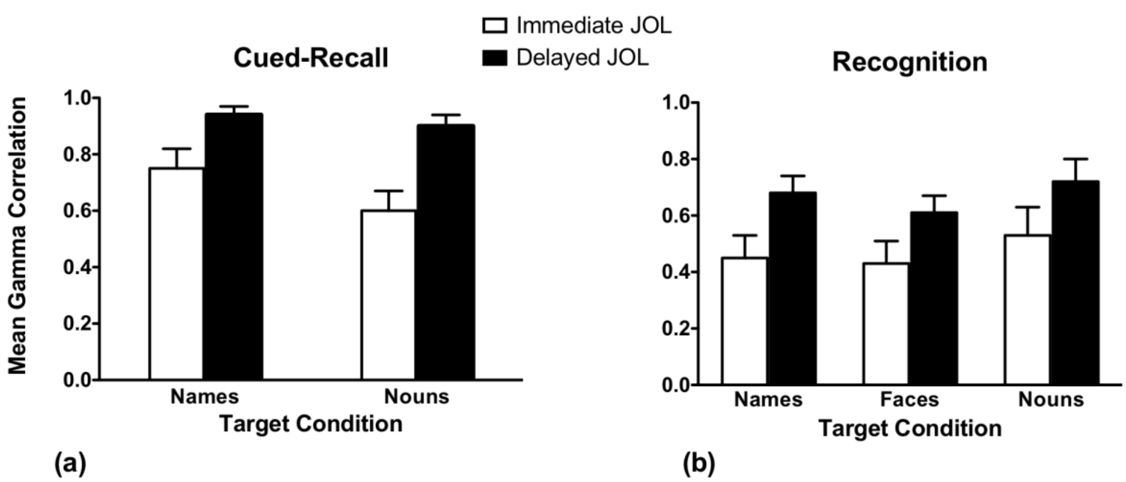

Figure 7. Mean Goodman-Kruskal Gamma correlations between judgements-of-learning (JOL) and (a) cued-recall accuracy or (b) recognition accuracy across target conditions. Error bars represent 1 SEM.

A Wilcoxon signed ranks test revealed that for name targets, the magnitude of the gamma correlation was significantly greater for delayed JOLs (*Mdn* = .97) than for immediate JOLs (*Mdn* = .85), z = 2.33, $p = .020$, $r^2 = 0.28$. Similarly, for noun targets, a Wilcoxon signed-ranks test revealed that the magnitude of the gamma correlation was significantly greater for delayed JOLs (*Mdn* = .94) than for immediate JOLs (*Mdn* = .65), z = 3.58, $p < .001$, $r^2 = 0.64$. Although smaller in magnitude compared to the noun target condition, the delayed JOL effect in cued-recall was replicated with name targets.

Using G*, a repeated measures t-test revealed that the delayed-JOL effect was replicated in the noun target condition. The mean G* was significantly greater for delayed JOLs compared with immediate JOLs, $t(8)$[10] = 7.27, p < .001, $r^2 = 0.86$. However, for the name target condition, the difference between immediate and delayed JOLs only showed a trend toward significance, $t(8) = 1.88$, p = .096.

The predictive validity of immediate versus delayed JOLs on associative recognition accuracy was assessed using two 2 (JOL-type) × 3 (target condition) mixed ANOVAs with JOL-type as a within-subject factor. For gamma, there was a significant main effect of JOL-type, $F(1,56) = 10.05$, p = .002, partial $\eta^2 = 0.15$, whereas the main effect of target condition, F(2,56) < 1, and the interaction between target condition and JOL-type, F(2,56) < 1, were not significant. Thus, as Figure 7 illustrates, delayed JOLs were more accurate predictors of associative recognition accuracy than immediate JOLs regardless of the target condition. For G*, the main effect of JOL-type approached significance, $F(1, 40) = 3.60$, p = .065, whereas the main effect of target condition, F(2,40) < 1, and the interaction between target condition and JOL-type, F(2,40) < 1, were not significant.

---

[10] The relatively low degrees of freedom for analyses involving G* is due to the fact that G* could not be calculated for gammas with a value of 1.

The disparity between G* and gamma on the delayed JOL effect is most likely due to the fact that G* cannot be calculated for gammas that have a value of 1. As Figure 8 illustrates, collapsing across target condition, the number of gamma correlations with a value of 1 was greater in the delayed condition compared with the immediate condition. Consequently, the difference between immediate and delayed JOLs on G* was attenuated because participants with perfect relative accuracy in the delayed condition were omitted from the analysis.

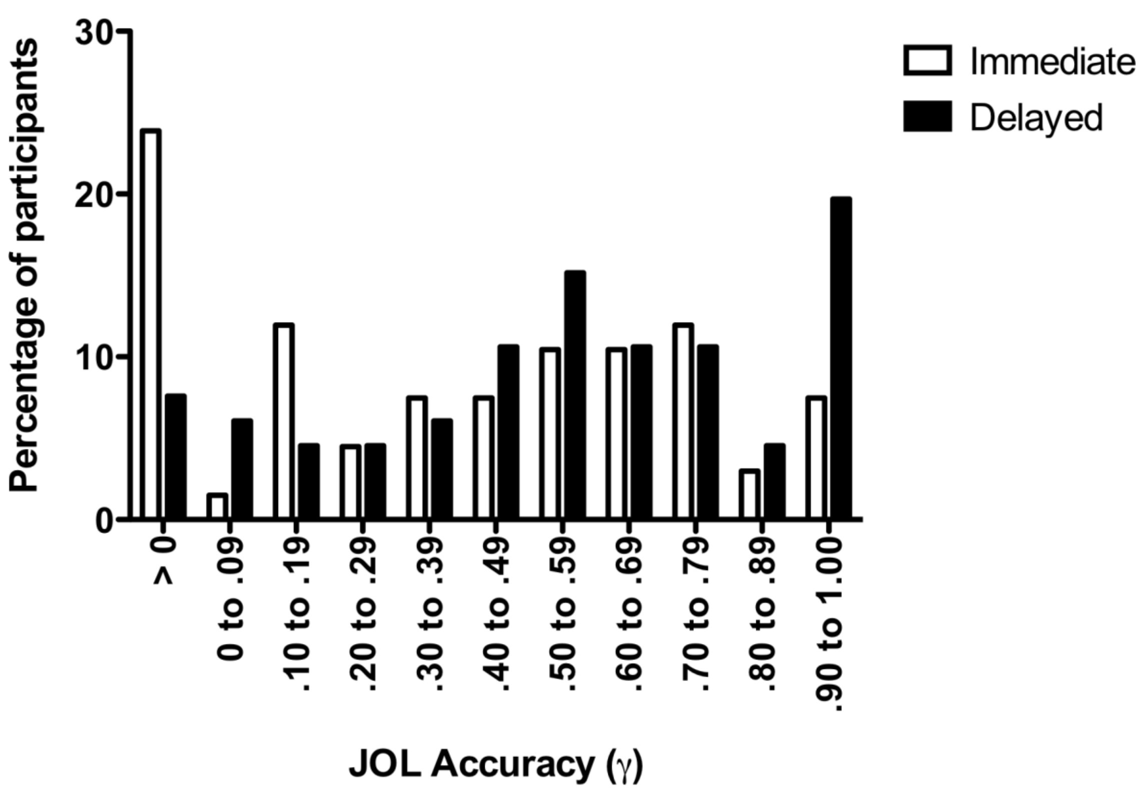

Figure 8. Distribution of gamma correlations between JOLs and recognition accuracy in immediate and delayed conditions.

Feeling-of-knowing judgements. Three one sample t-tests revealed that the mean gamma correlation between associative recognition accuracy and the magnitude of FOK judgements in name, $t(21) = 17.48$, $p < .001$, $r^2 = 0.92$, face, $t(24) = 11.63$, $p < .001$, $r^2 = 0.84$, and noun, $t(16) = 11.51$, $p < .001$, $r^2 = 0.88$, target conditions were significantly greater than zero. There were no significant differences among target conditions on the mean gamma correlation between FOK judgements and recognition accuracy, $F(2,62) < 1$, or on the mean G*, $F(2,55) < 1$. The relative accuracy of FOKs for the subset of items that could not be recalled during the cued-recall test was also examined. The mean gammas for both name and noun target conditions on these items were significantly greater than zero, $t(19) = 4.79$, $p < .001$, $r^2 = 0.54$, and $t(15) = 2.48$, $p < .001$, $r^2 = 0.30$, respectively. Thus, regardless of the category of stimuli, participants are capable of accurately predicting their recognition performance for items that could not be recalled. Two independent samples t-tests revealed that the difference between noun and name target conditions on the mean gamma, $t(34) < 1$, or G*, $t(28) < 1$, for the subset of items was not significant.

Imagery ratings. Participants in the face target condition were asked to make imagery ratings in order to approximate face recall. Considering that previous research has showed that the vividness of mental images of faces is related to the strength of face-name associations in memory (Groninger, 2000; 2006), I wanted to determine if imagery ratings were a significant predictor of face recognition accuracy. To this end, I calculated a gamma correlation between the magnitude of imagery ratings and recognition accuracy. A one-sample t-test revealed that the mean gamma correlation was significantly greater than zero, $t(24) = 10.66$, $p < .001$, $r^2 = 0.82$, and is similar in magnitude to the gamma correlation between FOKs and recognition accuracy (see Table 2). This led me to consider the possibility that imagery ratings were used as cues for FOK judgements. As a result, I calculated a gamma correlation between imagery ratings and FOK judgements. The mean gamma correlation (γ = .96, SE = .01), was significantly greater than zero, $t(24) = 85.82$, $p < .001$, $r^2 = 0.99$, which indicates that faces with clear mental images were more likely receive a higher FOK judgement than faces with poor mental images. Altogether, these results suggest that when cued with a name, the clarity of the mental image of the face can be used as an indicator of the likelihood of recognizing the face at some point in the future.

Retrospective confidence judgements. Three one sample t-tests revealed that the mean gamma correlation between the magnitude of RCJs and associative recognition accuracy for name, $t(23) = 26.15$, $p < .001$, $r^2 = .96$, face, $t(22) = 44.35$, $p < .001$, $r^2 = .98$, and noun, $t(15) = 46.51$, $p <.001$, $r^2 = .98$, target conditions were significantly greater than zero. A one-way ANOVA was conducted on the mean gamma correlation between RCJs and recognition accuracy to determine if relative accuracy of retrospective monitoring differed among target conditions. In contrast to prospective monitoring judgements, there was a significant difference among target conditions on gamma, $F(2,59) = 5.88$, $p = .005$, $\omega^2 = 0.13$. A Scheffe post-hoc test revealed that the mean gamma for the noun target condition was significantly greater compared with the name target condition ($p = .01$), but not the face target condition ($p > .05$). A similar pattern emerged using G*. A one-way ANOVA revealed a significant difference among target conditions on the predictive validity of RCJs, $F(2,57) = 4.12$, $p = .021$, $\omega^2 = 0.09$. A Scheffe post-hoc test revealed that the mean G* for the noun target condition was significantly greater than the name target condition ($p = .023$), but not the face target condition ($p > .05$). Thus, the relative accuracy of RCJs significantly differs between name and noun targets.

Absolute Accuracy.

Figure 9 contains plots of calibration lines for JOLs collapsing across immediate and delayed conditions, FOKs for all items, and RCJs across target conditions in Experiment 1. Beginning with JOLs, although the slopes of the calibration lines on cued-recall accuracy are similar between name (Y = .75X - .01) and noun (Y = .72X + 17) target conditions, the calibration line for the name target condition was consistently below the diagonal, which indicates overconfidence. This was statistically evaluated with a 2 (target condition: faces vs. names) × 2 (memory: cued-recall accuracy vs. JOL magnitude) mixed ANOVA. The main effect of memory was not significant, $F(1,48) = 1.34$, $p > .05$, whereas the main effect of target condition, $F(1,48) = 23.54$, $p < .001$, partial $\eta^2 = 0.33$, and the interaction between target condition and memory, $F(1,48) = 6.18$, $p = .016$, partial $\eta^2 = 0.14$ were significant. Simple effects tests revealed that for noun targets, the difference between mean JOL magnitude and mean cued-recall accuracy was not significant, $F(1,48) < 1$, whereas for name targets, the mean JOL magnitude was significantly greater than the mean cued-recall accuracy, $F(1,48) = 6.65$, $p = .013$. Thus, participants in the name target condition were overconfident in their future cued-recall ability compared with participants in the noun target condition.

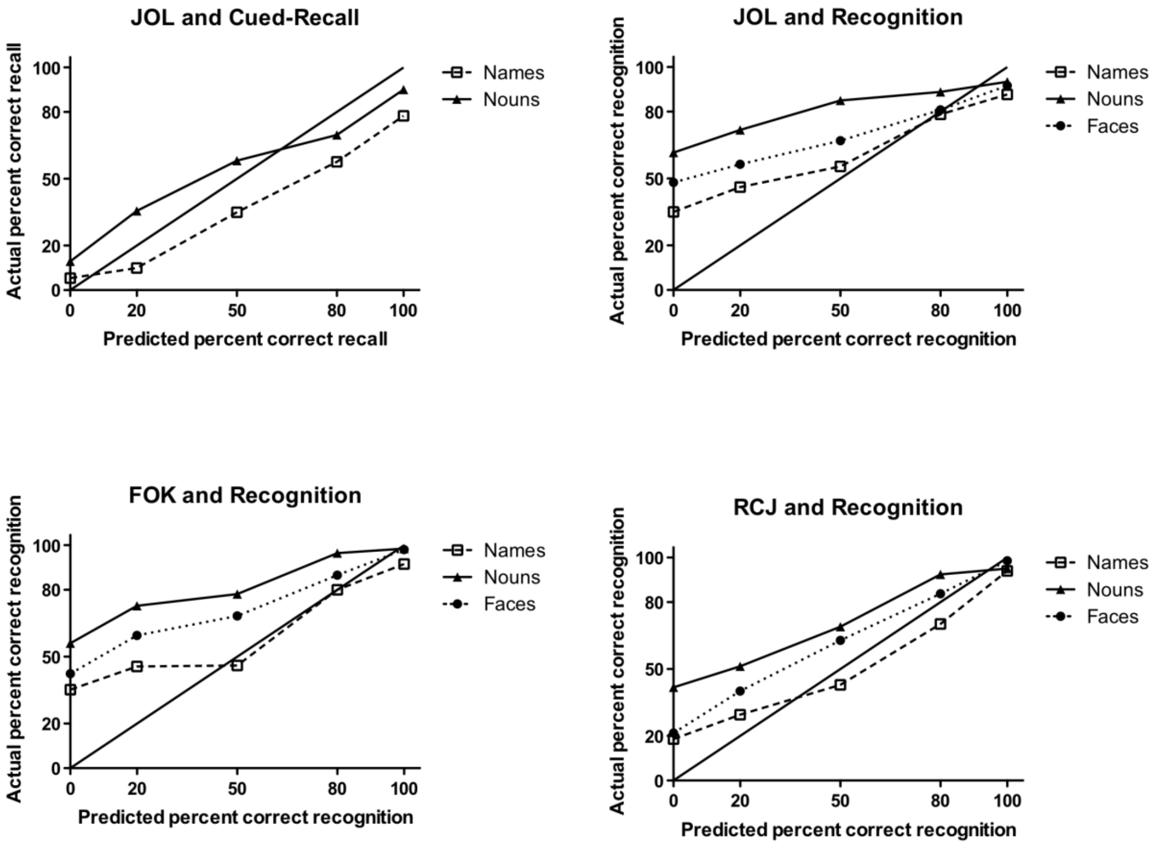

Figure 9. Calibration lines of judgements-of-learning collapsing across immediate and delayed conditions (JOL), feeling-of-knowing judgements for all items (FOK), and retrospective confidence judgements (RCJ) across target conditions. Each panel contains a plot of observed memory performance as a function of predicted memory performance for each target condition. The title of each panel refers to the monitoring judgement and memory test that was used to construct the calibration lines. The diagonal line refers to perfect calibration.

Turning to the correspondence between JOLs and associative recognition accuracy, the calibration lines suggest that participants in the noun target condition underestimated their associative recognition ability. Indeed, the linear parameters of the calibration lines for name (Y = .53X + 34) and face (Y = .43X + 47) target conditions are closer to perfect calibration compared with the noun (Y = .31X + 65) target condition. The statistical significance of this pattern of results was tested with a 3 (target condition: faces vs. names vs. nouns) × 2 (memory: associative recognition accuracy vs. JOL magnitude) mixed ANOVA. The main effect of memory, $F(1,69) = 157.45$, $p < .001$, partial $\eta^2 = 0.69$, and the main effect of target condition, $F(1,69) = 18.14$, $p < .001$, partial $\eta^2 = 0.34$, were significant. Importantly, these main effects were qualified by a significant interaction between target condition and memory, $F(2,69) = 4.29$, $p = .018$, partial $\eta^2 = 0.11$. Simple effects tests revealed that JOL magnitude was significantly lower than associative recognition accuracy in name, $F(1,69) = 26.38$, $p <.001$, face, $F(1,69) = 55.98$, $p <.001$, and noun, $F(1,69) = 82.28$, $p <.001$, target conditions, but the magnitude of the difference was the greatest in the noun target condition, which was sufficient to produce the interaction. Thus, participants in the noun target condition underestimated their future associative recognition performance to a greater degree that participants in the name or face target conditions.

Likewise, the absolute accuracy of FOKs was superior for name (Y = .57X + 31) and face (Y = .54X + 44) targets compared with noun (Y = .42X + 60) targets. A 3 (target condition: faces vs. names vs. nouns) × 2 (memory: associative recognition accuracy vs. FOK magnitude) mixed ANOVA revealed significant main effects of memory, $F(1,69) = 237.194$, $p < .001$, partial $\eta^2 = 0.77$, and target condition, $F(1,69) = 13.52$, $p < .001$, partial $\eta^2 = 0.28$, as well as a significant interaction between memory and target condition, $F(2,69) = 5.26$, $p = .007$, partial $\eta^2 = 0.13$. Simple effects tests revealed that FOK magnitude was significantly lower than associative recognition accuracy in the name, $F(1,69) = 41.03$, $p < .001$, face, $F(1,69) = 98.93$, $p < .001$, and noun, $F(1,69) = 105.16$, $p < .001$, target conditions, but the magnitude of the difference was smaller for the name target condition compared with the face and noun target conditions, which likely produced the interaction. Thus, after a cued-recall test, participants in the name target condition underestimated their associative recognition accuracy to a lesser degree compared with participants in the noun and face target conditions.

The calibration lines of RCJs show a similar pattern, with poorer absolute accuracy for noun targets (Y = .57X + 41) compared with face (Y = .76X + 23) or name (Y = .73X + 14) targets. A 3 (target condition: faces vs. names vs. nouns) × 2 (memory: associative recognition accuracy vs. RCJ magnitude) mixed ANOVA revealed significant main effects of memory, $F(1,70) = 18.67$, $p < .001$, partial $\eta^2 = 0.21$, and target condition, $F(1,70) = 14.61$, $p < .001$, partial $\eta^2 = 0.28$. The interaction between memory and target condition only approached significance, $F(2,70) = 3.05$, $p = .053$. As the bottom right panel of Figure 10 illustrates, the magnitude of RCJs was lower than associative recognition accuracy for face and noun targets, whereas the opposite pattern occurs for name targets. This is likely responsible for driving the interaction towards significance. Altogether, these results reveal that the precision at which JOLs, FOKs and RCJs reflect actual recognition accuracy differs among face, name, and noun targets. Participants underestimate their recognition performance for face targets to a greater degree than name targets. Furthermore, although participants are more confident in their memory for nouns compared with faces and names, they nevertheless underestimate their associative recognition ability for nouns to a greater degree than faces and names.

Discussion.

The results of Experiment 1 reveal that participants can monitor their memory for face-name associations at a level above chance during various stages of learning, and that the relative accuracy of their monitoring judgements does not differ between face and name targets. Thus, after a brief learning period, participants can distinguish between which faces and names will or will not be retrieved from memory at a comparable degree of accuracy. However, there are some differences in absolute monitoring performance between faces and names. Across all of the monitoring judgements, the calibration lines for face targets were consistently above the diagonal, especially for the lower estimates of memory performance (i.e. 0%, 20%, 50%). This effect was more prominent for face targets compared with name targets, which suggests that participants underestimate their associative recognition memory for faces to a greater degree than names.

Interestingly, when cued with a name, the mental image clarity of a face was predictive of face recognition accuracy, and was strongly correlated with the magnitude of FOKs, which suggests that imagery might be used as a cue when making FOK judgements for faces. This interpretation is in line with target accessibility accounts of FOKs (Koriat, 1993; 1995; Koriat & Levy-Sadot, 2001), which posit that the by-products of retrieval act en masse to contribute to the magnitude of FOKs. In the case of face targets, a by-product of retrieval involved mental imagery, which presumably contributed to the magnitude of FOK judgements. However, it should be noted that imagery ratings were always solicited prior to FOKs, therefore the relationship between FOKs and imagery should be interpreted with caution.

Unexpectedly, Experiment 1 yielded mixed results with regards to the hypothesis that relative accuracy would be superior for noun targets compared with face and name targets. On the one hand, the gamma correlations between retrieval performance and the magnitude of EOLs, JOLs, or FOKs were similar across target conditions (see Table 2). This suggests that the relative accuracy of *prospective monitoring* is equivalent among faces, names, and nouns. On the other hand, the gamma correlation between the magnitude of RCJs and recognition accuracy was significantly greater for noun targets compared with name targets. This suggests that the relative accuracy of *retrospective monitoring* is superior for nouns compared with names. A detailed explanation of the pattern of results is presented in Chapter 7.

As expected, the delayed-JOL effect was replicated when names were used as targets on cued-recall and associative recognition tests or when faces were used as targets on an associative recognition test. This indicates that similar metamemory mechanisms operate over different categories of stimuli (e.g. nouns, names, and faces), but taking into consideration the findings mentioned above, retrospective monitoring appears to operate more efficiently for nouns compared with names.

In each target condition, multiple monitoring judgements were solicited on the same set of items. Admittedly, a limitation of this experimental design is that it leaves open the possibility of practice or carry-over effects among monitoring judgements. However, considering that correlations among monitoring tasks are weak (Kelemen, Frost, & Weaver, 2000; Leonesio & Nelson, 1990), neurological impairments differentially affect performance on JOL, FOK, and RCJ tasks (Modirrousta & Fellows, 2008; for a review, see Pannu & Kaszniack, 2005), and each monitoring judgement is thought to rely on different sources of information (Schwartz, 1994), it is unlikely that practice or carry-over effects influenced the relative accuracy results. Nonetheless, to ensure that the significant difference between noun and name targets on the relative accuracy of RCJs is a genuine effect of the metamemory system rather than an artifact of the experimental design, a second experiment was conducted, in which participants were not explicitly required to provide prospective confidence judgements prior to completing the RCJ task.

Experiment 2

This experiment used the same procedure and stimuli as Experiment 1, with the exception that prospective monitoring judgements were not solicited. If practice at monitoring the same items is responsible for the pattern of RCJ results in Experiment 1, then the differences in relative accuracy of RCJs across target conditions in Experiment 2 should be attenuated. The reason being is that participants in Experiment 2 do not receive practice at monitoring their memory prior to completing RCJs. However, if the pattern of RCJ results in Experiment 1 genuinely reflects more accurate retrospective monitoring for noun targets compared with name targets, then the significant difference among target conditions on the relative of RCJs should be replicated even when participants do not complete any prospective monitoring judgements.

Methods.

Participants.

Fifty-eight undergraduate students from the University of Ottawa (38 Females and 20 Males) participated in the experiment. The mean age of the participants was 18.2 years. Each participant received course credit for participating. All participants were tested individually.

Design.

The experiment was a single factor between-subject design with three levels (face targets vs. name targets vs. noun targets). The participants were randomly assigned to each target condition, with 20 participants in the name and noun target conditions and 18 participants in the face target condition.

Materials and apparatus.

The same materials and apparatus used in Experiment 1 were used in Experiment 2.

Procedure.

Participants followed the same procedure as Experiment 1, with the exception that prospective monitoring judgements were not solicited. Thus, participants were presented with each paired-associate for five seconds and were instructed to simply view each item. Afterwards, the order of paired-associates was randomized anew and participants were instructed to study each paired-associate for five seconds. Following the study session, participants completed an unrelated filler task for two minutes, and then completed a cued-recall test. For name and noun target conditions, participants were instructed to type the corresponding name or noun that was previously associated with the cue during the beginning stages of the experiment. For the face target condition, participants were instructed to respond whether they could recall the face by pressing a key labeled with a "Y" for yes or key labeled with a "N" for no. As in Experiment 1, catch trials were included. Following the cued-recall task, participants completed another unrelated filler task for two minutes and then completed an 8-AFC associative recognition test. For every recognition decision, a retrospective confidence judgement was solicited. The procedure for the recognition test and RCJ task was identical to Experiment 1.

Results.

Descriptive statistics for cued-recall accuracy, associative recognition accuracy, the magnitude of RCJs, and relative accuracy of RCJs across target conditions are presented in Table 3. Any outliers that were detected based on the criteria developed by Van Selst and Jolicoeur (1994) were removed from the analysis.[11] All hypothesis tests were conducted with alpha set at .05.

Table 3

Descriptive Statistics for all Dependent Variables across Target Conditions in Experiment 2

	Target Condition								
	Names			Faces			Nouns		
Dependent Variable	*M*	*SEM*	*N*	*M*	*SEM*	*N*	*M*	*SEM*	*N*
Cued-recall accuracy	0.23	0.03	19	-	-	-	0.43	0.06	20
Recognition accuracy	0.45	0.04	20	0.50	0.06	18	0.67	0.06	20
RCJ Magnitude	49.94	2.73	19	49.52	4.29	18	66.69	4.32	20
Relative accuracy of RCJs (γ)	0.77	0.03	18	0.77	0.03	18	0.91	0.02	14

Note. Empty cells occur in conditions where accuracy on the memory test could not be objectively determined (see Chapter 4 for details). Relative accuracy was based on performance on the recognition test. RCJ, retrospective confidence judgements.

Memory performance. An independent t-test revealed than cued-recall accuracy was significantly greater in noun target condition compared with the name target condition, $t(37) = 3.11$, p = .004, $r^2 = .20$. Likewise, a one-way ANOVA revealed that recognition accuracy significantly differed among target conditions, $F(2,55) = 5.09$, p =

[11] Table 3 contains the sample size in each condition for each dependent variable after outliers were removed.

.009, $\omega^2 = .13$. A Scheffe post-hoc test revealed that recognition accuracy was greater for noun targets compared with name ($p = .013$) targets, but not face targets ($p > .05$). Thus, much like Experiment 1, associative memory was superior for noun-noun pairs compared with face-name pairs.

RCJ magnitude. A one-way ANOVA revealed that the mean magnitude of RCJs significantly differed among target conditions, $F(2,54) = 6.88$, $p = .002$, $\omega^2 = .17$. In a like manner to Experiment 1, a Scheffe post-hoc test revealed that the magnitude of RCJs was greater for noun targets compared with name ($p = .009$) and face ($p = .010$) targets.

Relative accuracy of RCJs. Three participants in the noun target condition had indeterminate gammas because of near ceiling performance on the recognition test. On the basis of Levene's test, the assumption of homogeneity of variance was violated, $F(2,47) = 3.69$, $p = .043$. Therefore, as recommended by Field (2009), an *F*-ratio was calculated using the Brown-Forsythe (1974) correction, which resulted in a significant difference among target conditions on the mean gamma correlation between RCJs and recognition accuracy, $F(2,41.76) = 7.74$, $p < .001$, $\omega^2 = .19$. A Scheffe post-hoc test revealed that the mean gamma was greater in the noun target condition compared with the face ($p = .01$) or name ($p = .005$) target conditions. Similarly, using G*, a one-way ANOVA revealed a significant difference among target conditions, $F(2,45) = 3.82$, $p = .029$, $\omega^2 = .10$. A Scheffe post-hoc test revealed that the mean G* was greater in the noun target condition compared with the name ($p = .032$) but not face ($p > .05$) target conditions. These results are consistent with Experiment 1, and indicate that retrospective monitoring is more accurate for common nouns compared with names.

Absolute accuracy of RCJs. The calibration lines for RCJs across target conditions are presented in Figure 10.

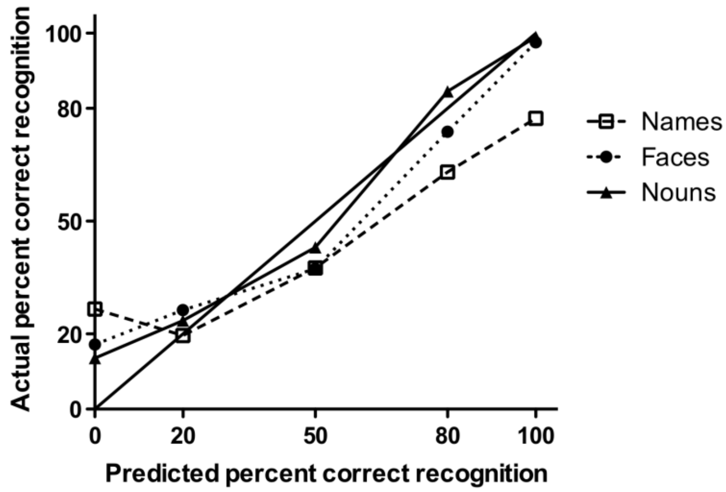

*Figure 10***.** Calibration lines of RCJs across target conditions in Experiment 2. The diagonal line indicates perfect calibration.

As can be seen in Figure 10, the parameters of the calibration lines for face (Y = .80X + 11), and noun (Y = .90X + 7.8) target conditions are closer to perfect calibration compared with the name target condition (Y = .57X + 16). A 3 (target condition: faces vs. names vs. nouns) × 2 (memory: associative recognition accuracy vs. RCJ magnitude) mixed ANOVA revealed a significant main of target condition, $F(2,53) = 7.87$, $p = .001$, partial $\eta^2 = 0.23$, whereas the main effect of memory, $F(1,53) < 1$, or the memory × target condition interaction, $F(2,53) < 1$ were not significant. Thus, across target conditions, mean retrospective confidence corresponded with mean associative recognition accuracy, which is in contrast to Experiment 1. The disparity could be due to the fact that associative recognition accuracy was lower in Experiment 2 compared with Experiment 1, which may have mitigated under confidence across target conditions. This result is elaborated upon in Chapter 7.

Discussion.

Experiment 2 revealed that when prospective monitoring judgements are not solicited, the relative accuracy of RCJs is superior for nouns compared with names. These results suggest that it is unlikely that the difference in relative accuracy between name targets and noun targets in Experiment 1 is due to practice or carry-over effects. Ultimately, the results of both experiments support the hypothesis that participants can monitor their memory for face-name associations at various stages of learning, and retrospective metamemory is more accurate, in a relative sense, with noun-noun pairs compared with face-name pairs. Having established that individuals can monitor their memory for faces and names, one of the aims of Experiments 3 and 4 is to advance this finding by identifying sources of information that individuals use to monitor their memory for faces and names.

Chapter 6 – Experiments 3 and 4: The effects of distinctiveness on metamemory for face-name associations

The intuition that distinct items are easier to remember compared with typical items is supported by many laboratory findings (for reviews, see Hunt & McDaniel, 1993; Schmidt, 1991), and faces and names are no exception. Distinct faces are recognized more accurately and quickly compared with typical faces (e.g. Brown & Lloyd-Jones, 2006; Cohen & Carr, 1975; Courtois & Mueller, 1981; Ellis, Shepherd, Giblin, & Shepherd, 1988; Light et al., 1979; Semmler & Brewer, 2006; Valentine & Bruce, 1986 Winograd, 1981), and distinct names are recognized (Brandt et al., 2006) and recalled (Stanhope & Cohen, 1993) more accurately compared with typical names. Little is known, however, of the influence of distinctiveness on metamemory and associative memory for faces and names. Questions arise as to whether estimates of future and past memory performance for face-name associations are sensitive to manipulations of distinctiveness and whether the strength of the association improves if the face or name is distinct. The purpose of Experiments 3 and 4 is to investigate these questions, and by doing so, to determine if individuals use face or name distinctiveness as a source of information when monitoring their memory for people.

Research into the mechanisms underlying metamemory has revealed that participants rely on a variety of cues and heuristics when making monitoring judgements. In particular, researchers have identified the following as potential sources of information that participants can use, implicitly or explicitly, when evaluating their memory: encoding fluency (e.g. Benjamin & Bjork, 1996; Begg et al., 1989; Koriat & Ma'ayan, 2005), retrieval fluency (e.g. Benjamin & Bjork, 1996; Kelley & Lindsay, 1993), cue familiarity (e.g. Koriat & Levy-Sadot, 2001; Metcalfe & Finn, 2008), accessibility of information relating to the target (Koriat, 1993; 1995; Koriat & Levy-Sadot, 2001), item difficulty (Dunlosky & Matvey, 2001), and the interval between the learning episode and the solicitation of the monitoring judgement (Connor et al., 1997; Dunlosky & Nelson, 1992; 1994; Nelson et al., 2004; Thiede & Dunlosky, 1994). In regards to learning the identity of an individual, the distinctiveness of a face or name is likely to be another source of information that participants can use to monitor their memory. For instance, if a recently learned name or face is out of the ordinary, then participants may feel very confident in their ability to remember that individual at some point in the future.

Two lines of research contributed to the development of this hypothesis. The first is that the phenomenological experience of face and name retrieval differs depending on the degree of distinctiveness. By using the remember-know procedure (Tulving, 1985), researchers have found that distinct faces (Brandt et al., 2003; Brown & Lloyd-Jones, 2006; Mantyla, 1997) and names (Brandt et al., 2006) are more likely to entail recognition decisions based on explicit recollection compared with typical faces and names. In contrast, typical faces (Brandt et al., 2003; Vokey & Read, 1992) are more likely to evoke a feeling of familiarity compared with distinct faces. In short, typical faces seem to invoke a sense of familiarity during recognition whereas both distinct faces and names are more likely to invoke recollection of the context in which the stimulus was learned. Note that the remember-know procedure can be regarded as a form of retrospective memory monitoring since participants are asked to reflect on their experience of retrieving an item from memory. If remember-know judgements vary as a function of distinctiveness, it is reasonable to hypothesize that other types of monitoring judgements (e.g. EOLs, JOLs, FOKs, or RCJs) might also be sensitive to variations of distinctiveness.

The second line of research examined the sensitivity of monitoring judgements to target set size (e.g. Izaute et al., 2000). For instance, word targets that contain a small set of semantic associations in memory tend to produce greater FOK judgements compared with word targets that contain a large set of semantic associations (Schreiber, 1998), which suggests that the uniqueness of a target in memory is a factor that affects monitoring. As Koriat and Lieblich (1977) contend, a cue that is associated with a unique target will serve as the best indicator for an accurate monitoring judgement because of less interference from neighboring targets. Considering that some computational models of face recognition posit that distinct faces occupy points in low-density regions of a multi-dimensional "face space" – which enables distinct faces to be recognized more accurately relative to typical faces because less noise is generated from neighboring points (e.g. Lewis, 2004; Lewis & Johnston, 1999; Valentine, 1991; 2001) – it is reasonable to hypothesize that the distinctiveness of a face can influence the monitoring of a face-name association (i.e., because of the face's relative uniqueness in memory).

This possibility has some empirical support. Sommer et al. (1995) reported that the accuracy of JOLs in predicting face recognition performance is dependent on facial distinctiveness. However, the authors did not use an objective measure of recognition performance to validate the accuracy of JOLs, which undermines the legitimacy of their finding. Nonetheless, the same study reported a significant positive correlation between distinctiveness ratings and JOLs, which indicates that participants are more confident in their ability to recognize a face as the distinctiveness of the face increases. In a similar vein, Semmler and Brewer (2006) found that participants are more confident in the accuracy of their recognition decisions when faces are distinct compared to when they are typical. Together, these results suggest that prospective and retrospective monitoring is sensitive to variations of facial distinctiveness. It has yet to be determined, however, whether these findings generalize to other monitoring judgements (i.e. EOLs and FOKs), to manipulations of name distinctiveness, or to tasks that test associative memory for face-name pairs as opposed to tasks that strictly test memory for faces.

In addition to examining metamemory, Experiments 3 and 4 will also determine if distinctiveness strengthens the face-name association. There are scattered findings in the literature that point to the possibility that the distinctiveness of a face can modulate the learning and retrieval of its associated name (e.g. Pantelis et al., 2008; Valentine & Moore, 1995). To the best of my knowledge, only one study has examined the extent to which the distinctiveness of an unfamiliar face can improve memory for a name or other identity-related semantic information. It showed that cued-recall accuracy for typical names is superior when the name is learned with a distinct face as opposed to a typical face (Bruyer & Courvoisier, 1990). This finding highlights the possibility that the memory-enhancing effects of facial distinctiveness can transfer over to names. This begs the question as to whether the beneficial effects of distinctiveness on face-name associations are symmetrical, such that name distinctiveness can improve memory for faces, and whether the symmetry (or asymmetry for that matter) is apparent in metamemory.

It is reasonable to suggest that if one item in a face-name pair is distinct, then it will improve memory for the other item, regardless of whether the other item is distinct or whether the distinct item is a cue or target on an associative memory test. This suggestion derives from the contentions that distinct items engender elaborative encoding of the learning episode (e.g. Brandt et al., 2003; Hirshman, Whelley, & Palij, 1989; Schmidt, 1991) and recollection of the context in which the item was learned (Rajaram, 1996). As a result, if the face or name in a face-name pair is distinct, then the face-name pair as a whole should benefit not only from elaborative encoding but also from context recollection at the time of retrieval, which ultimately should improve memory for the target face or name.

The purpose of the Experiments 3 and 4 is to determine if the distinctiveness of a face or name affects memory for its associated name or face, respectively, and if monitoring judgements are sensitive to manipulations of distinctiveness. To this end, the experiments followed Nelson and Narens (1990) framework in order to evaluate associative memory and metamemory performance between face-name pairs that contained distinct versus typical components. In Experiment 3, the face in a face-name pair was either distinct or typical, whereas the name was neither typical nor distinct. In Experiment 4, the name in a face-name pair was either distinct or typical, whereas the face was neither typical nor distinct.

In both experiments, the type of target used in the cued-recall and associative recognition tests was manipulated. This was done to determine if the effects of distinctiveness on monitoring vary depending on whether the distinct or typical item is a cue or a target. Several of the inferential theories of monitoring judgements contend that information from the cue or target is sampled during monitoring. Thus, it is possible that monitoring performance might vary depending on the role of the distinct item in the cue-target pair. For example, the sensitivity of FOK judgements to distinctiveness might only be present when the distinct item acts as a target rather than a cue. This aspect of the experiments is exploratory, and although I do not make any explicit predictions as to how this factor might interact with distinctiveness, I nonetheless wanted to ensure that I could examine its potential affects on metamemory performance.

Distinctiveness Ratings

In Experiments 3 and 4, distinctiveness was operationalized according to ratings obtained during pilot testing. For the typical and distinct faces used in Experiment 3, an independent sample of observers rated a series of faces based on how memorable the face is or whether it easily would blend into a crowd, which is a common procedure in studies of facial distinctiveness (e.g. Brandt et al., 2003; Semmler & Brewer, 2006; Valentine, 1991). Similarly, for the typical and distinct names used in Experiment 4, the same sample rated a series of names based on how ordinary or remarkable it is. For both faces and names, a median split was conducted. The lowest and highest 18 items from the medians were grouped as typical and distinct, respectively. Subsets of the remaining names and faces were randomly selected as associates to the faces and names in Experiment 3 and Experiment 4, respectively. Thus, in Experiment 3 the faces were either typical or distinct and the names were neither typical nor distinct, whereas in Experiment 4, the names were either typical or distinct, and the faces were neither typical nor distinct. The method for the pilot test is described below.

Method.

Participants.

Seven students (5 Females and 2 Males), ages 20-27 (median value: 22), from the University of Ottawa were asked to rate a series of faces and names in terms of distinctiveness. All participants were tested individually and received monetary compensation for participating.

Materials and apparatus.

One hundred and five male faces and names were used in the distinctiveness rating tasks. Eighty-seven face images were selected from an online database at the University of Stirling's Psychology Department PIC set (http://pics.psych.stir.ac.uk). Of the 87 images, 50 belonged to the "nothingham_scans" set, 20 belonged to the "Stirling_faces" set, and 17 belonged to the "aberdeen" set. Thirteen images were selected from the Georgia Institute of Technology face database (http://www.anefian.com/research/face_reco.htm). The remaining five images were selected from a database at the University of Ottawa and were used for practice trials. All of the faces were Caucasian males who appeared to be between the ages of 18-40 and depicted a neutral expression. None of the faces contained particularly identifiable features (e.g. scars, facial hair, glasses, etc.). The images were edited to remove any clothing or background cues, and cropped to depict the neck, face and hairstyle. All faces were presented in the frontal (0°) view. Each image was reduced to 8-bit grey-scale and scaled so the image dimensions were approximately 5 cm in width and 7 cm in height (visual angle of 5° x 7° at the viewing distance of 57 cm). All the images were equated for mean luminance and RMS contrast. After normalization, the images were imbedded in a 256 x 256 pixel uniform black background.

One hundred and five male first names were selected from the US census bureau and US social security card applications. The names reflect Western European and North American naming conventions (e.g. Bryce, Luke, Mark, Leland, Noah, Maddox). All of the names were presented in white Arial font, size 21. Of the 105 names, five were randomly selected for practice trials.

The experiment was run with Matlab 7.6 (Mathworks Inc., Natick, MA; http://www.mathworks.com) on a Macintosh Dual 2 GHz PowerPc G5. Stimuli were presented on an NEC AccuSync 95F 17” monitor.

Procedure.

Participants were seated at a computer and informed that the experiment is composed of a face distinctiveness rating task and a name distinctiveness rating task. Prior to beginning the experiment, participants were presented with instructions and a set of practice trials for each task. Participants completed five practice trials for each task, and progressed through each task in a sequential manner. For both tasks, all ratings were made on a 7-point scale from 1 (very typical) to 7 (very distinct). After the practice trials were complete, participants were given the opportunity to ask the researcher any questions. Otherwise, they could begin the experiment proper by pressing any key. During the experiment proper, the procedure was identical to the practice, with the exception that participants completed 100 trials per task. The order of tasks in the practice and experiment proper was counterbalanced across participants.

The instructions for the face distinctiveness rating task were adapted from Semmler and Brewer (2006), and were as follows: “In this section, you will be looking at a series of faces. Your task is to rate how distinctive or typical each face is on a scale from 1 to 7, with 1 being very typical to 7 being very distinct. A typical looking face is one that would easily blend into a crowd, and that you would not take notice of it. If you think that the face you are seeing is typical, rate it low on the scale (i.e. give it a 1 or 2). A distinctive face is one that would stand out from a crowd and make you take notice of it. A face might be distinctive because it is particularly attractive, or it might be distinctive because it is particular ugly. Or it may have a particular prominent feature or set of features that do not match well. If you think that a face is distinctive, rate it high on the scale (i.e. give it a 6 or a 7). The faces you will be shown may vary in terms of quality of the image, please do not let this influence your ratings. At the bottom of the screen, there will be a response prompt; for each face, use the keyboard to type the number on the scale that you think best represents how typical or distinctive the face looks. Don’t spend too much time looking at each face, just go with your first reaction. You will only have 8 seconds to rate each face before the next one is presented. After you have rated a face, a new face will immediately appear on the monitor. You cannot go back and change any of you ratings. Try not to miss any of the faces. You may press any key to begin.”

The instructions and procedure for the name distinctiveness rating task were similar to the face distinctiveness rating task, with the exception that names were rated rather than faces. The instructions were modified to inform participants that a typical name is one that is very common, highly usual and ordinary, whereas a distinct name is one that is very uncommon, highly unusual and remarkable.

Thus, for both the face and name distinctiveness rating tasks, a face or name appeared at the centre of the monitor for eight seconds and participants were prompted to rate the distinctiveness of the item on a 7-point scale. Immediately after participants made their rating, a new item appeared on the monitor. The procedure continued until ratings were made for all items. Participants received a 20 second break after every 25^{th} trial, and in between tasks. For each participant, the order of presentation of the items in each task was randomized.

Results.

For every face and name presented in the experiment proper, the mean rating across participants was calculated. A median split was then conducted. The median ratings for faces and names were 3.78 and 4.21, respectively. For both faces and names, the 18 highest and lowest items from the median were selected as distinct and typical items, respectively. The mean ratings for the distinct and typical face sets were 5.27 (SD = 0.49) and 2.60 (SD = 0.28), respectively. The mean ratings for the distinct and typical name sets were 6.21 (SD = .43) and 1.60 (SD = .29). For both faces and names, paired-samples t-tests were conducted on the distinctiveness ratings to ensure that the distinct and typical sets were significantly different from each other. For both faces, $t(17) = 18.13$, $p <.001$, and names, $t(17) = 35.70$, $p <.001$, the ratings for the distinct set were significantly greater compared with the typical set. Of the remaining 64 faces and 64 names, 36 names and 36 faces were randomly selected as associates for the faces and names in Experiment 3 and Experiment 4, respectively. The mean distinctiveness ratings for these names and faces were 3.94 (SD = .80), and 3.53 (SD = .38), respectively. In Experiment 3, for each participant, the 18 distinct and 18 typical faces were randomly paired with the 36 names that were neither typical nor distinct. In Experiment 4, for each participant, the 18 distinct and 18 typical names were randomly paired with the 36 faces that were neither typical nor distinct.

Experiment 3

The purpose of Experiment 3 is to examine the influence of facial distinctiveness on associative memory and metamemory for face-name associations. Following the reasoning outlined in the beginning of this chapter, cued-recall and associative recognition accuracy are expected to be superior for face-name pairs that contain a distinct face compared with a typical face. This will occur regardless of whether the distinct face is a cue or target. Thus cued-recall and associative recognition accuracy for names are expected to be greater when the name is associated with a distinct face compared with a typical face. Furthermore, when names are used as cues, it is expected that distinct faces will be recognized more accurately compared with typical faces.

On the basis of previous findings (Semmler & Brewer, 2006; Sommer et al., 1995) and theoretical accounts of metamemory (e.g. Koriat, 1995; 1997; 2007), it is expected that prospective and retrospective estimates of associative memory performance will be greater when the face in a face-name pair is distinct compared to when it is typical. Moreover, considering that relative accuracy is greater for cues associated with unique targets in memory (Koriat & Lieblich, 1977) and distinct faces can be considered unique targets (Valentine, 1991), it is expected that participants' estimates of memory performance will be more accurate at discriminating between targets that will be (or were) and will not be (or were not) retrieved on tests of associative memory when a face-name pair contains a distinct face compared with a typical face.

Method.

Participants.

Sixty undergraduate students (10 Males and 50 Females) from the University of Ottawa volunteered to participate in this experiment in exchange for course credit. The age of the participants ranged from 17 to 30 (median = 18). All participants were tested individually.

Design.

A 2 × 2 mixed factorial design was employed, with target condition (name vs. face) as a between-subject factor and face type (distinct vs. typical) as a within-subject factor. The participants were randomly assigned to the between-subject conditions, with thirty participants in each condition.

Materials and apparatus.

Forty-four faces and names were used in the experiment proper. Of the 44 faces, 18 were distinct and 18 were typical. The remaining eight faces were neither typical nor distinct and were used for practice trials only. All of the names were neither typical nor distinct. All of the stimuli were selected on the basis of distinctiveness ratings obtained during pilot testing, as described above. The same apparatus from the pilot test was used in Experiment 3.

Procedure.

The procedure was identical to the general procedure outlined in Chapter 4, with following exceptions. First, each judgement was made on a 5-point 0-100 scale, using keys on a keyboard that were labeled 0, 25, 50, 75, and 100. The change in key labels from Experiments 1 and 2 to Experiments 3 and 4 was due to a programming technicality. Second, participants were informed that the image quality of the faces might vary, such that some of the images might appear “grainier” than others, but to not let that influence their judgements. Third, half of the face-name pairs contained a distinct face, whereas the other half contained a typical face. Finally, during the study-JOL stage, on each trial a paired-associate was presented on the monitor for up to ten seconds. If a participant did not require the full ten seconds to study the item, they were instructed to press the space bar to end the study period and move on to the confidence judgement. Otherwise, JOLs were solicited immediately after ten seconds elapsed. Thus, the amount of study time per item was partly determined by each participant. This was done in an attempt to gauge how distinctiveness influences control of studying. Only the cue (i.e., a name or a face, depending on the condition) was presented on the monitor when the JOL was solicited.

Results.

Descriptive statistics for associative memory accuracy across target condition and face type are presented in Table 4. Figure 11 depicts the magnitude of the monitoring judgements for face-name pairs that contain a typical or a distinct face, collapsing across target conditions. Any outliers that were detected based on the criteria developed by Van Selst and Jolicoeur (1994) were removed from the analysis. All hypothesis tests were conducted with alpha set at .05.

Table 4

Descriptive Statistics for Cued-Recall and Associative Recognition Accuracy across Target Condition and Face Type in Experiment 3

	Target Condition and Face Type							
	Face Targets				Name Targets			
	Typical Faces		Distinct Faces		Typical Faces		Distinct Faces	
Memory Test	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
Cued-Recall	-	-	-	-	0.26	0.03	0.40	0.03
Associative Recognition	0.62	0.04	0.75	0.03	0.53	0.04	0.67	0.04

Note. Empty cells occur in conditions where accuracy on the memory test could not be objectively determined (see Chapter 4 for details).

Memory performance.

A paired-samples t-test revealed that names were more accurately recalled when paired with a distinct face compared with a typical face, $t(28) = 5.10$, $p < .001$, $r^2 = .47$. The type of name recall errors that participants committed on the cued-recall test also varied according to facial distinctiveness. Two paired samples t-test revealed that typical face types yielded significantly more commission errors than distinct face types, $t(29) = 4.28$, $p < .001$, $r^2 = .39$, whereas the number of omission errors did not significantly differ between face types, $t(29) < 1$. This last finding is consistent with the predictions of "face-space" models since an incorrect name was less likely to be recalled when prompted with a distinct face compared with a typical face. The difference between distinct and typical face types on response latencies during the cued-recall test was not significant, $t(29) < 1$.

A 2 (face type) × 2 (target condition) mixed ANOVA was conducted on associative recognition accuracy with face type as a within-subject factor. The ANOVA revealed a significant main effect of face type, $F(1,57) = 45.47$, $p < .001$, partial $\eta^2 = .44$, with distinct faces yielding superior associative recognition accuracy compared with typical faces. The main effect of target condition revealed a trend towards significance, $F(1,57) = 3.51$, $p = .066$, with recognition accuracy greater for face targets (M = .68, SE = .03) than for name targets (M = .60, SE = .03). A 2 (face type) × 2 (target condition) mixed ANOVA was conducted on recognition response latencies and revealed significant main effects of face type, $F(1,58) = 23.51$, $p < .001$, partial $\eta^2 = .28$, and target condition, $F(1,58) = 5.72$, $p = .02$, partial $\eta^2 = .09$. Response latencies were lower for distinct (M = 7.29, SD = 2.07) compared with typical (M = 8.30, SD = 2.20) faces, and for name targets (M = 7.51, SD = 1.89) compared with face targets (M = 8.70, SD = 1.99). Thus, associative recognition performance was superior for face-name pairs that contained a distinct face compared with a typical face, and there is some evidence to suggest that face targets were more accurately recognized than name targets, much like Experiment 1.

Visual inspection of the histograms of study times across face type and target conditions revealed a disproportionate amount of participants used the full ten seconds during the study period in every trial to learn the face-name association. As a result, the distributions are non-normal. The violation of normality was confirmed using four Kolmogorov-Smirnov tests, which revealed that the distributions of study times in the name target condition for typical, $D(30) = .22$, $p < .05$, and distinct, $D(30) = .19$, $p < .05$, face types, and in the face target condition for typical, $D(30) = .27$, $p < .05$, and distinct, $D(30) = .19$, $p < .05$, face types, were significantly non-normal. Thus, for both target conditions, non-parametric statistics were conducted to determine if the amount of study time differed between distinct and typical face types. Two Wilcoxon signed rank tests revealed that for face targets, the amount of time spent studying the face-name pair was greater for typical faces (*Mdn* = 9.34) compared with distinct faces (*Mdn* = 8.40), z =

2.80, $p = .005$, $r^2 = 0.26$. In contrast, for name targets, the difference between face types was not significant, z = .211, $p > .05$.

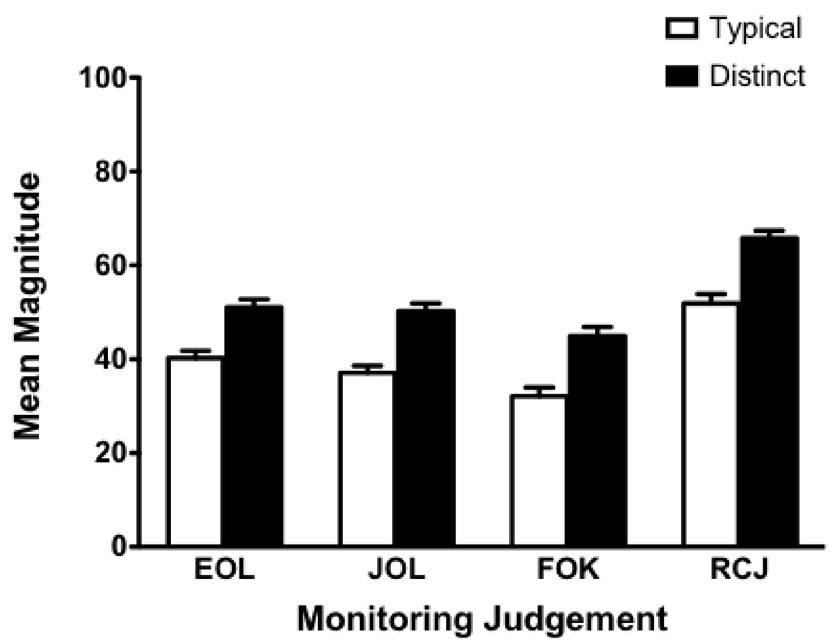

Figure 11. The mean magnitude of monitoring judgements in Experiment 3 for distinct and typical face types, collapsing across target condition. EOL, ease-of-learning judgements; JOL, judgements-of-learning; FOK, feeling-of-knowing judgements for all items; RCJ, retrospective confidence judgements. Error bars represent 1 SEM.

Magnitude of monitoring judgements.

As Figure 11 illustrates, the magnitude of each monitoring judgement was greater for face-name pairs that contained a distinct face compared with a typical face. The statistical significance of these differences was assessed with four 2 (face type) × 2 (target condition) mixed ANOVAs. As expected, there were main effects of face type on the magnitude of EOLs, $F(1,57) = 31.44$, $p < .001$, partial $\eta^2 = .35$; on the magnitude of JOLs, $F(1,56) = 44.54$, $p < .001$, partial $\eta^2 = .44$; on the magnitude of FOKs, $F(1,57) =$

68.33, $p < .001$, partial $\eta^2 = .55$; and on the magnitude of RCJs, $F(1,58) = 83.56$, $p < .001$, partial $\eta^2 = .59$. For all monitoring judgements, neither the main effect of target condition nor the interaction between face type and target condition was significant. Hence, monitoring judgements were sensitive to manipulations of facial distinctiveness.

In keeping with previous studies that only solicited FOKs for incorrect responses, I also compared the magnitude of FOKs between distinct and typical face types for names that were correctly recalled and incorrectly recalled. Two paired sample t-tests revealed that for correctly recalled names, the mean magnitude of FOKs was significantly greater for face-name pairs with a distinct face (M = 80.15, SD = 14.68) compared with a typical face (M = 68.69, SD = 19.00), $t(29) = 4.58$, $p < .001$, $r^2 = .42$. However, for names that were incorrectly recalled, the difference between distinct (M = 18.36, SD = 12.91) and typical (M = 14.90, SD = 10.83) faces was not significant, $t(29) = 1.85$, $p > .05$. Thus, even after accurately recalling a name, participants were more confident in their ability to recognize that name in the future if it was paired with a distinct face compared with a typical face. However, if participants could not recall the name, distinctiveness did not affect their prospective confidence.

In regards to the imagery data in the face target condition, the mean imagery rating for distinct faces (M = 42.90, SD = 14.57) was significantly greater compared with typical faces (M = 31.54, SD = 13.67), $t(29) = 3.82$, p = .001, $r^2 = .35$. This suggests that on average, participants could image the target face and image clarity was superior for distinct faces compared with typical faces.

Relative accuracy.

Three participants in the name target condition scored below 5% on the cued-recall test. In addition, two participants in the face target condition scored above 95% on the recognition test. These participants were not included in the analyses of relative accuracy because of the limited variability in their data (e.g. Connor et al., 1997). Table 5 presents descriptive statistics for relative accuracy using gamma across face type and target conditions. Samples sizes in Table 5 do not include outliers.

Table 5

Descriptive Statistics for Relative Accuracy across Target Condition and Face Type in Experiment 3

	Target Condition and Face Type											
	Face Targets						Name Targets					
	Typical Faces			Distinct Faces			Typical Faces			Distinct Faces		
Relative Accuracy (γ)	*M*	*SEM*	*N*	*M*	*SEM*	*N*	*M*	*SEM*	*N*	*M*	*SEM*	*N*
EOL-Recall	-	-	-	-	-	-	0.39	0.09	26	0.49	0.06	26
EOL-Recognition	0.36	0.06	27	0.26	0.07	27	0.29	0.08	30	0.36	0.06	30
JOL-Recall	-	-	-	-	-	-	0.70	0.05	24	0.63	0.06	24
JOL-Recognition	0.49	0.08	26	0.46	0.05	26	0.55	0.07	29	0.64	0.05	29
FOK-Recognition	0.67	0.05	26	0.68	0.04	26	0.54	0.07	29	0.67	0.06	29
Imagery-Recognition	0.54	0.09	25	0.63	0.07	25	-	-	-	-	-	-
RCJ-Recognition	0.83	0.03	25	0.91	0.02	25	0.73	0.05	27	0.88	0.02	27

Note. Empty cells occur in conditions where a monitoring judgement was not solicited or where accuracy on the memory test could not be objectively determined (see Chapter 4 details). EOL, ease-of-learning judgements; JOL, judgements-of-learning; FOK, feeling-of-knowing judgements for all items; RCJ, retrospective confidence judgements. The suffixes -recall and -recognition refer to the objective memory test that was used in the calculation of Goodman-Kruskal's Gamma (γ).

Ease-of-learning judgements. Two one sample t-tests revealed that the mean gamma correlations between cued-recall accuracy for names and the magnitude of EOLs for distinct and typical face types were significantly greater than zero, $t(25) = 7.75$, $p < .001$, $r^2 = .71$, $t(25) = 4.26$, $p < .001$, $r^2 = 0.42$, respectively. Likewise, four one sample t-tests revealed that the mean gamma correlation between associative recognition accuracy and the magnitude of EOLs was significantly greater than zero for both distinct and typical face types in the name target condition, $t(29) = 6.01$, $p < .001$, $r^2 = 0.55$, $t(29) = 3.61$, $p < .001$, $r^2 = 0.32$, respectively, and in the face target condition, $t(26) = 3.89$, $p = .001$, $r^2 = 0.36$, $t(26) = 6.00$, $p < .001$, $r^2 = 0.58$, respectively. Thus, estimates of learning difficulty are predictive of future associative memory performance for face and name targets, regardless of the degree of distinctiveness of the face in the face-name pair.

Two paired-samples t-tests were conducted on gamma and G* to determine if the predictive validity of EOLs on cued-recall accuracy for name targets significantly differed between distinct and typical face types. For gamma, and for G*, the difference between typical and distinct face types was not significant, $t(25) < 1$ and $t(23) = 1.66$, $p > .05$, respectively. Two 2 (face type) × 2 (target condition) mixed ANOVAs with face type as a within-subject factor were conducted to evaluate the predictive validity of EOLs on associative recognition accuracy. For gamma, the main effect of face type, $F(1,55) < 1$, the main effect of target condition, $F(1,55) < 1$, and the face type × target condition interaction, $F(1,55) = 1.85$, $p > .05$, were not significant. Likewise, for G*, the main effect of face type, $F(1,47) < 1$, the main effect of target condition, $F(1,47) < 1$, and the face type × target condition interaction, $F(1,47) = 1.77$, $p > .05$, were not significant. Hence, the predictive validity of EOLs did not differ among face types or target conditions.

Judgements-of-learning. Two one sample t-tests revealed that the mean gamma correlations between cued-recall accuracy and the magnitude of JOLs for distinct and typical face types were significantly greater than zero, $t(23) = 10.46$, $p < .001$, $r^2 = 0.82$,

$t(23) = 12.79$, $p < .001$, $r^2 = 0.88$, respectively. Similarly, four one sample t-tests revealed that the mean gamma correlation between associative recognition accuracy and the magnitude of JOLs was significantly greater than zero for both distinct and typical face types in the name target condition, $t(28) = 13.31$, $p < .001$, $r^2 = 0.86$, $t(28) = 8.46$, $p < .001$, $r^2 = 0.72$, respectively, and in the face target condition, $t(25) = 8.90$, $p < .001$, $r^2 = 0.76$, $t(25) = 6.17$, $p < .001$, $r^2 = 0.60$, respectively. In like manner to EOLs, JOLs were predictive of future cued-recall and associative recognition performance, regardless of facial distinctiveness.

Two paired-samples t-tests were conducted on gamma and G* to determine if the predictive validity of JOLs on cued-recall accuracy for name targets significantly differed between distinct and typical face types. For gamma, the difference between typical and distinct face types was not significant, $t(23) < 1$. Similarly, for G* the difference was not significant, $t(23) = 1.66$, $p > .05$. Two 2 (face type) × 2 (target condition) mixed ANOVAs with face type as a within-subject factor were conducted on gamma and G* to determine if the predictive validity of JOLs on recognition accuracy differed among face types and target conditions. For gamma, the main effect of target condition approached significance, $F(1,53) = 3.40$, $p = .071$,whereas the main effect of face type, $F(1,53) < 1$, and the face type × target condition interaction, $F(1,53) = 1.16$, $p > .05$, were not significant. For G*, there was a significant main effect of target condition, $F(1,43) = 4.92$, $p = .032$, partial $\eta^2 = 0.10$, whereas the main effect of face type, $F(1,43) = 2.22$, $p > .05$, and the interaction between target condition and face type, $F(1,43) = 1.10$, $p > .05$, were not significant. The predictive validity of JOLs on associative recognition accuracy was superior for face-cues/name-targets compared with name-cues/face-targets.

Feeling-of-knowing judgements. Four one sample t-tests revealed that the mean gamma correlation between associative recognition accuracy and the magnitude of FOKs was significantly greater than zero for typical and distinct face types in the name target condition, $t(28) = 8.18$, $p < .001$, $r^2 = 0.71$, $t(28) = 12.22$, $p < .001$, $r^2 = 0.84$, and in the face target condition, $t(25) = 14.66$, $p < .001$, $r^2 = 0.89$, $t(25) = 15.43$, $p < .001$, $r^2 = 0.91$, respectively. The relative accuracy of FOKs for the subset of items that were not recalled during the cued-recall test could not be calculated for distinct or typical face types because of the limited number of cases.

A 2 (face type) × 2 (target condition) mixed ANOVA with face type as a within-subject factor was conducted on gamma to evaluate the predictive validity of FOKs on associative recognition accuracy. The main effect of face type, $F(1,53) = 2.02$, $p > .05$, the main effect of target condition, $F(1,53) = 1.16$, $p > .05$, and the face type × target condition interaction, $F(1,53) = 1.31$, $p > .05$, were not significant. For G*, the assumption of homogeneity of covariances was violated. As a result, two non-parametric Wilcoxon signed rank tests were conducted to compare G* between distinct and typical face types. For both face and name targets, the difference between distinct and typical face types was not significant, $z = .82$, $p > .05$, and $z = .47$, $p > .05$, respectively. Thus, the predictive validity of FOKs did not differ among face types or target conditions.

Imagery ratings. Two one sample t-test revealed that the mean gamma correlation between imagery ratings and associative recognition accuracy for both distinct and typical faces was significantly greater than zero, $t(26) = 10.03$, p < .001, $r^2 = 0.79$, $t(27) = 5.98$, p < .001, $r^2 = 0.57$, respectively. The difference between distinct and typical face types on gamma, $t(24) < 1$, or on G*, $t(14) < 1$, was not significant.

Retrospective confidence judgements. Four one sample t-tests revealed that the mean gamma correlation between associative recognition accuracy and the magnitude of RCJs for distinct and typical face types in the name target condition, $t(26) = 44.28$, $p < .001$, $r^2 = 0.98$, $t(26) = 16.16$, $p < .001$, $r^2 = 0.91$, respectively, and in the face target condition, $t(24) = 41.92$, $p < .001$, $r^2 = 0.98$, $t(24) = 25.17$, $p < .001$, $r^2 = 0.96$, respectively, were significantly greater than zero. Hence, like prospective monitoring, retrospective monitoring was above chance, regardless of facial distinctiveness.

Two 2 (face type) × 2 (target condition) mixed ANOVAs with face type as a within-subject factor were conducted on gamma and G* to determine if the predictive validity of RCJs on associative recognition accuracy differed among face type and target conditions. For gamma, the main effect of face type was significant, $F(1,49) = 15.98$, $p < .001$, partial $\eta^2 = 0.25$, whereas the main effect of target condition, $F(1,49) = 2.41$, $p > .05$, and the interaction between face type and target condition, $F(1,49) = 1.15$, $p > .05$, were not significant. A similar pattern emerged in G*, where the main effect of face type was significant, $F(1,35) = 7.72$, $p = .009$, partial $\eta^2 = 0.18$, whereas the main effect of target condition, $F(1,35) < 1$, and the interaction between face type and target condition, $F(1,35) = 1.34$, $p > .05$, were not significant. Thus, retrospective monitoring is more accurate for distinct faces compared with typical faces.

Absolute accuracy.

Figure 12 contains plots of calibration lines for JOLs, FOKs, and RCJs across face types and target conditions in Experiment 3. Beginning with JOLs and cued-recall, distinct face types (Y = .57X + 13) and typical face types (Y = .54X + .95) yielded similar levels of absolute accuracy. In both conditions, however participants overestimated their future memory performance. This was statistically evaluated with a 2 (face type: typical vs. distinct) × 2 (memory: cued-recall accuracy vs. JOL magnitude) repeated-measures ANOVA. The main effects of face type, $F(1,26) = 10.70$, $p = .003$, partial $\eta^2 = 0.29$, and memory, $F(1,26) = 37.73$, $p < .001$, partial $\eta^2 = 0.59$ were significant, whereas the interaction between memory and face type was not significant, $F(1,26) < 1$. The significant main effect of face type indicates that facial distinctiveness affected JOLs and cued-recall accuracy in a similar manner, such that JOLs and cued-recall accuracy were lower for typical face types compared with distinct face types, which is consistent with the hypothesis that participants use distinctiveness as a source of information to monitoring their memory. The main effect of memory indicates that for both distinct and typical faces, JOLs for names overestimated cued-recall accuracy, much like Experiment 1.

The calibration lines for JOLs and associative recognition performance indicate that absolute accuracy was superior for distinct (Y = .51X + 41) compared with typical (Y = .36X + 39) face types in the name target condition, but not in the face target condition, where distinct face types (Y = .29X +59) were poorly calibrated compared with typical face types (Y = .46X +42). This pattern of results, however, was not statistically significant. A 2 (face type: typical vs. distinct) × 2 (memory: associative recognition accuracy vs. JOL magnitude) × 2 (target condition: faces vs. names) mixed ANOVA with target condition as a between subject factor revealed significant main effects of memory, $F(1,55) = 66.51$, $p < .001$, partial $\eta^2 = 0.55$, and face type, $F(1,55) = 90.74$, $p < .001$, partial $\eta^2 = 0.62$, but none of the interactions were significant, nor was the main effect of condition. As the top right panel of Figure 12 illustrates, participants in all conditions underestimated future recognition performance. This accounts for the significant main effect of memory. The significant main effect of face type is due to the fact that distinct face types yielded greater levels of associative recognition accuracy and JOL magnitude compared with typical face types.

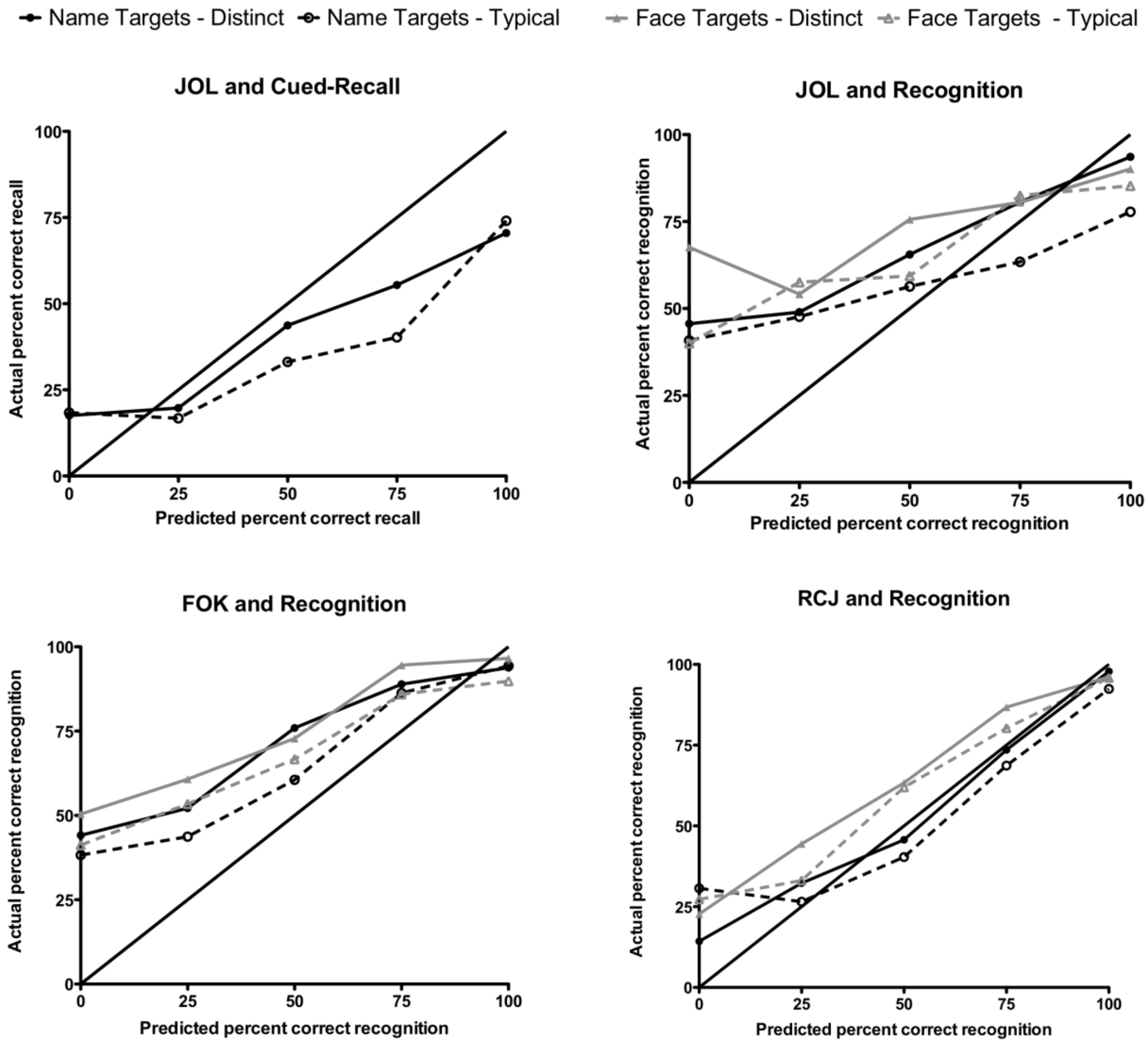

Figure 12. Calibration lines of judgements-of-learning (JOL), feeling-of-knowing judgements for all items (FOK), and retrospective confidence judgements (RCJ) across face type and target conditions in Experiment 3. Each panel contains a plot of observed memory performance as a function of predicted memory performance for each target condition. The title of each panel refers to the monitoring judgement and memory test that was used to construct the calibration lines. The solid black diagonal line refers to perfect calibration. The dotted lines refer to typical faces whereas the solid lines refer to distinct faces.

As for FOKs, participants appear to underestimate their future recognition performance. For name targets, distinct (Y = .54X + 44) and typical (Y = .62X + 34) face types yielded similar levels of calibration accuracy. Likewise, for face targets, the slopes of the calibration lines were approximately identical between distinct (Y = .50X + 55) and typical (Y = .52X + 41) face types. However, in all conditions the calibration lines were consistently above the diagonal. This was confirmed with a 2 (face type: typical vs. distinct) × 2 (memory: associative recognition accuracy vs. FOK magnitude) × 2 (target condition: faces vs. names) mixed ANOVA. The main effects of memory, $F(1,56) = 197.54$, $p < .001$, partial $\eta^2 = 0.77$, and face type, $F(1,56) = 89.69$, $p < .001$, partial $\eta^2 = 0.61$, were significant, whereas the main effect of condition was not significant nor were any of the interactions. As with JOLs, the main effect of memory indicates that the magnitude of FOKs was significantly lower than recognition accuracy across all conditions, and the main effect of face type indicates that the magnitude of FOKs and recognition accuracy were greater for distinct faces compared with typical faces.

Finally, visual inspection of the calibration lines for RCJs reveal that participants are well calibrated in name target condition, with distinct (Y = .83X + 11) and typical (Y = .66X +19) face types close to perfect calibration, but participants in the face target condition reveal under confidence, with distinct (Y = .76X + 25) and typical face types (Y = .74X + 23) slightly above the diagonal. The statistical significance of this pattern of results was evaluated with a 2 (face type: typical vs. distinct face types) × 2 (memory: associative recognition accuracy vs. RCJ magnitude) × 2 (target condition: faces vs. names) mixed ANOVA. The main effects of memory, $F(1,57) = 9.14$, $p = .004$, partial $\eta^2 = 0.14$, and face type, $F(1,57) = 79.83$, $p < .001$, partial $\eta^2 = 0.58$, were significant. The main effect of memory was qualified by a memory × target condition interaction, $F(1,57) = 7.42$, $p = .009$, partial $\eta^2 = 0.15$. Follow-up simple effects tests revealed that associative recognition accuracy and RCJ magnitude did not significantly differ for name targets,

$F(1,57) < 1$, but recognition accuracy was significantly greater than RCJ magnitude for face targets, $F(1,57) = 16.25$, $p < .001$, which indicates that participants in the face target condition underestimated their performance on the recognition test.

Discussion.

Experiment 3 yielded several interesting results in regards to the effects of facial distinctiveness on participants' memory and metamemory for face-name associations. As expected, not only were distinct faces recognized more accurately than typical faces, but names paired with distinct faces were also recalled and recognized more accurately compared with names paired with typical faces. This finding reveals that the memory-enhancing effects of facial distinctiveness (e.g. Brown & Lloyd-Jones, 2006; Cohen & Carr, 1975; Courtois & Mueller, 1981; Light et al., 1979; Semmler & Brewer, 2006; Winograd, 1981) can transfer over to memory for names, which is consistent with previous findings (e.g. Bruyer & Courvoisier, 1990; Valentine & Moore, 1985) and the contention that distinct items prompt elaborative encoding at the time of learning (e.g. Brandt et al., 2006; Hirshman et al.,1989; Schmidt, 1991). Chapter 7 elaborates on the possible mechanisms underlying this effect.

The results also revealed that monitoring face-name associations is sensitive to facial distinctiveness. The magnitude of EOL judgements was greater for distinct face types compared with typical face types, which indicates that participants consider face-name pairs easier to learn when the face is distinct. The study time results support this interpretation since participants spent less time studying face-name pairs that contained a distinct face compared with a typical face. The magnitude of JOLs and FOKs were also greater for distinct face types, which indicates that participants are more confident in their ability to retrieve a face if it is distinct or retrieve a name if it is paired with a distinct face. Finally, the magnitude of RCJs indicates that participants are more confident in their associative recognition responses when faces are distinct compared to when they are typical. This occurs regardless of whether participants are attempting to recognize a face or recognize a name when cued with a face.

The calibration lines revealed that facial distinctiveness yielded proportionate effects on the magnitude of monitoring judgements and performance on objective memory tests. Prospective and retrospective estimates of cued-recall and associative recognition accuracy were greater for face-name pairs that contained a distinct face compared with a typical face, and cued-recall and associative recognition accuracy was greater for face-name pairs that contained a distinct face compared with a typical face. The significant main effect of face type across all monitoring judgements statistically supports this interpretation of the calibration lines. Thus, it appears that participants are aware, either implicitly or explicitly, of the memory-enhancing effects of facial distinctiveness. However, like Experiment 1, participants tended to underestimate their associative recognition ability, especially in the face target condition.

Interestingly, the predictive validity of retrospective monitoring was superior for face-name pairs that contain a distinct face compared with a typical face. This suggests that participants are better at distinguishing between faces that were and were not accurately recognized when the face is distinct compared to when it is typical, which is consistent with the predictions of “face-space” models (e.g. Valentine, 1991; 2001). Moreover, when the distinct face is a cue, it improves participants’ ability to distinguish between correct and incorrect name recognition decisions. Although there were no significant differences between distinct and typical face types on the relative accuracy of three prospective monitoring judgements, it is worth noting that, by and large, the effect sizes for gamma were greater for distinct face types compared with typical face types.

JOLs were more accurate predictors of name recognition compared with face recognition. In Experiment 1, I was unable to detect any differences between conditions that used face-cues/name-targets and name-cues/face-targets on the relative accuracy of three prospective monitoring judgements. Furthermore, I was unable to replicate this effect in Experiment 4 (see below). Thus, it is unlikely that this effect is due to faces versus names, *qua* faces and names. Considering that faces were either typical or distinct, whereas names were neither typical nor distinct, it is possible that JOLs were more accurate predictors of associative recognition performance for face-cues/name-targets compared with name-cues/face-targets because the relative distinctiveness of the face cues provided an additional heuristic, one that was diagnostic of future memory performance, that participants could use to make their JOLs.

Experiment 4

Experiment 3 revealed that facial distinctiveness influences memory and metamemory for face-name associations. As outlined in the beginning of this chapter, it is expected that if one item in a face-name pair is distinct, then it can improve memory for the other item. If this hypothesis is correct, then the beneficial effects of distinctiveness should be symmetrical between faces and names, that is – since distinct faces improve memory for names, distinct names should also improve memory for faces. Moreover, since monitoring judgements are sensitive to variations of facial distinctiveness, they should also be sensitive to variations of name distinctiveness. Experiment 4 tests these predictions. Its design is identical with Experiment 3, with the exception that the distinctiveness of the name was manipulated instead of that of the face.

Method.

Participants.

Seventy-two undergraduate students (12 Males and 60 Females) from the University of Ottawa volunteered to participate in this experiment in exchange for course credit. The age of the participants ranged from 16 to 35 (median value: 18). All participants were tested individually.

Design.

A 2 × 2 mixed factorial design was employed, with target condition (name vs. face) as a between-subject factor and name type (distinct vs. typical) as a within-subject factor. The participants were randomly assigned to the between-subject conditions, with 37 participants in the name target condition and 35 participants in the face target condition.

Materials and apparatus.

Forty-four faces and names were used in this experiment. Of the 44 names, 18 were distinct and 18 were typical. The remaining eight names were neither typical nor distinct and were used for practice trials only. All of the faces were neither typical nor distinct. All of the stimuli were selected on the basis the distinctiveness ratings obtained during pilot testing, as described above. The apparatus was the same as was used in Experiment 3.

Procedure.

The procedure was identical to that of Experiment 3, with the exception that name distinctiveness was manipulated rather than face distinctiveness.

Results.

As in Experiment 3, memory performance is summarized before metamemory performance. Any outliers that were detected based on the criteria developed by Van Selst and Jolicoeur (1994) were removed from the analysis. Alpha was set at .05 for all analyses. Descriptive statistics for memory performance across target condition and name type are presented in Table 6. Figure 13 depicts the magnitude of the monitoring judgements for each name type, collapsing across target condition.

Table 6

Descriptive Statistics for Cued-Recall and Associative Recognition Accuracy across Target Condition and Name Type in Experiment 4

	Target Condition and Name Type							
	Face Targets				Name Targets			
	Typical Names		Distinct Names		Typical Names		Distinct Names	
Memory Test	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
Cued-Recall	-	-	-	-	0.29	0.03	0.33	0.03
Associative Recognition	0.62	0.03	0.72	0.03	0.52	0.03	0.63	0.03

Note. Empty cells occur in conditions where accuracy on the memory test could not be objectively determined (see Chapter 4 for details).

Memory performance.

Two paired samples t-tests revealed that cued-recall accuracy and response latency did not significantly differ between typical and distinct name types, $t(35) = 1.76$, $p > .05$, and $t(36) = 1.31$, $p > .05$, respectively. Similarly, two paired samples t-tests revealed that the number of omission, $t(31) < 1$, and commission errors, $t(35) < 1$, did not significantly differ between name types. In contrast to cued-recall, a 2 (name type) × 2 (target condition) mixed ANOVA with name type as a within-subject factor revealed significant main effects of name type, $F(1,67) = 15.55$, $p < .001$, partial $\eta^2 = 0.18$, and target condition, $F(1,67) = 7.71$, $p = .007$, partial $\eta^2 = 0.10$, on associative recognition accuracy. Analogous to Experiment 3, faces associated with distinct names were more accurately recognized than faces associated with typical names, and distinct names were more accurately recognized than typical names. Also, like Experiments 1, face targets were more accurately recognized than name targets. A 2 (name type) × 2 (target condition) mixed ANOVA revealed a main effect of name type on recognition response latency, $F(1,67) = 39.79$, $p < .001$, partial $\eta^2 = 0.37$. Response latencies were lower for distinct names (M = 6.63, SD = 1.69) compared with typical names (M = 7.77, SD = 2.14). Altogether, these results are consistent with the hypothesis that distinct names can improve face recognition performance relative to typical names.

Once again, the assumption of normality was violated in the study time data because a disproportionate amount of participants used the full ten seconds to study the paired-associates. For both target conditions, non-parametric statistics were conducted to determine if the amount of study time differed between distinct and typical name types. Two Wilcoxon signed rank tests revealed that for face targets, $z = 1.13$, $p > .05$, and name targets, $z = .36$, $p > .05$, the difference between distinct and typical name types on the duration of study time was not significant.

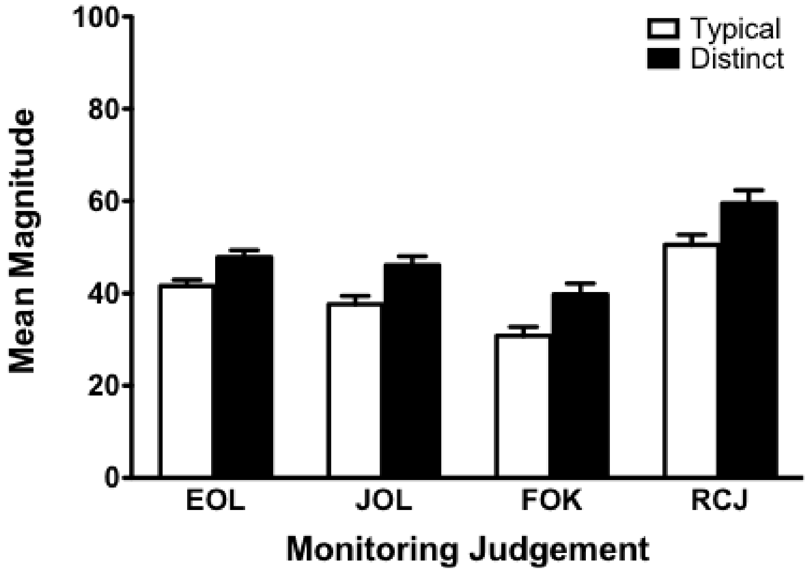

Figure 13. The mean magnitude of monitoring judgements in Experiment 4 for distinct and typical name types, collapsing across target conditions. EOL, ease-of-learning judgements; JOL, judgements-of-learning; FOK, feeling-of-knowing judgements for all items; RCJ, retrospective confidence judgements. Error bars represent 1 SEM.

Magnitude of monitoring judgements.

As Figure 13 illustrates, the magnitude of each monitoring judgement was greater for face-name pairs that contained a distinct name compared with a typical name. Four 2 (name type) × 2 (target condition) mixed factorial ANOVAs with name type as the within-subject factor were conducted on the magnitude of the monitoring judgements. As expected, there were main effects of name distinctiveness on the magnitude of EOLs, $F(1,65) = 14.25$, $p < .001$, partial $\eta^2 = .18$; on the magnitude of JOLs, $F(1,68) = 16.54$, $p < .001$, partial $\eta^2 = .20$; on the magnitude of FOKs, $F(1,70) = 12.27$, $p < .001$, partial $\eta^2 = .15$; and on the magnitude of RCJs, $F(1,70) = 9.72$, $p = .003$, partial $\eta^2 = .12$. For each monitoring judgement, neither the main effect of target condition nor the interaction between name type and target condition was significant. The mean imagery rating was significantly greater for faces paired with distinct names (M = 47.72, SD = 17.80) compared with faces paired with typical names (M = 31.88, SD = 15.59), $t(34) = 6.17$, $p < .001$, $r^2 = .53$. This suggests that the distinctiveness of a name can influence the recollection of its associated face. Altogether, these results reveal that monitoring face-name associations is sensitive to manipulations of name distinctiveness, which is consistent with my hypothesis and the results of Experiment 3.

Relative accuracy.

Six participants in the name target condition yielded indeterminate gammas because of near floor performance on the cued-recall test. Similarly, two participants in the name target condition and four participants in the face target condition yielded indeterminate gammas because of ceiling performance on the associative recognition test. These participants were removed from the inferential analyses. Table 7 presents descriptive statistics for gamma across name type and target conditions.

Table 7

Descriptive Statistics for Relative Accuracy across Target Condition and Name Type in Experiment 4

	Target Condition and Name Type											
	Face Targets						Name Targets					
	Typical Names			Distinct Names			Typical Names			Distinct Names		
Relative Accuracy (γ)	*M*	*SEM*	*N*	*M*	*SEM*	*N*	*M*	*SEM*	*N*	*M*	*SEM*	*N*
EOL-Recall	-	-	-	-	-	-	0.32	0.06	30	0.43	0.05	30
EOL-Recognition	0.33	0.07	29	0.36	0.06	29	0.12	0.08	34	0.29	0.06	34
JOL-Recall	-	-	-	-	-	-	0.43	0.07	31	0.56	0.04	31
JOL-Recognition	0.32	0.07	30	0.50	0.07	30	0.34	0.06	33	0.49	0.05	33
FOK-Recognition	0.58	0.05	30	0.69	0.05	30	0.47	0.06	33	0.54	0.06	33
Imagery-Recognition	0.63	0.09	31	0.81	0.04	33	-	-	-	-	-	-
RCJ-Recognition	0.69	0.04	31	0.77	0.06	31	0.67	0.03	34	0.63	0.05	34

Note. Empty cells occur in conditions where a monitoring judgement was not solicited or where accuracy on the memory test could not be objectively determined (see Chapter 4 for details). EOL, ease-of-learning judgements; JOL, judgements-of-learning; FOK, feeling-of-knowing judgements for all items; RCJ, retrospective confidence judgements. The suffixes -recall and -recognition refer to the objective memory test that was used in the calculation of Goodman-Kruskal's Gamma (γ).

Ease-of-learning judgements. The predictive validity of EOLs on cued-recall accuracy was assessed with two one sample t-tests, which revealed that the mean gamma correlation was significantly greater than zero for both typical, $t(29) = 5.70$, $p < .001$, $r^2 = .52$, and distinct, $t(29) = 7.99$, $p < .001$, $r^2 = .69$, name types. Similarly, two one sample t-tests revealed that the mean gamma correlation between associative recognition accuracy and the magnitude of EOLs was significantly greater than zero for both distinct and typical name types in the face target condition, $t(28) = 5.63$, $p < .001$, $r^2 = 0.53$, $t(28) = 4.96$, $p < .001$, $r^2 = 0.46$, respectively. However, for name targets, the relative accuracy of EOLs was above chance for distinct name types, $t(33) = 4.82$, $p < .001$, $r^2 = 0.47$, but not for typical name types, $t(33) = 1.53$, $p > .05$. Thus, estimates of learning difficulty are predictive of future associative memory performance for face targets, but name distinctiveness modulates the predictive validity of EOLs on recognition for name targets.

Two paired-samples t-tests were conducted on gamma and G* to determine if the predictive validity of EOLs on cued-recall accuracy for name targets significantly differed between distinct and typical name types. For gamma, and for G*, the difference between typical and distinct name types was not significant, $t(29) = 1.49$, $p > .05$, and $t(28) = 1.04$, $p > .05$, respectively. Two 2 (name type) × 2 (target condition) mixed ANOVAs with name type as a within-subject factor were conducted to evaluate the predictive validity of EOLs on associative recognition accuracy. For gamma, the main effect of name type, $F(1,61) = 2.41$, $p > .05$, the main effect of target condition, $F(1,61) = 3.39$, $p > .05$, and the name type × target condition interaction, $F(1,61) = 1.36$, $p > .05$, were not significant. Likewise, for G*, the main effect of name type, $F(1,59) = 3.61$, $p > .05$, the main effect of target condition, $F(1,59) = 3.12$, $p > .05$, and the name type × target condition interaction, $F(1,59) = 1.27$, $p > .05$, were not significant.

Judgements-of-learning. The predictive validity of JOLs on cued-recall accuracy was assessed with two one sample t-tests. The outcome of the tests revealed that the mean gamma correlation was significantly greater than zero for both distinct, $t(30) = 12.76$, $p < .001$, $r^2 = .84$, and typical, $t(30) = 6.42$, $p < .001$, $r^2 = .58$, name types. The same can be said for the predictive validity of JOLs on associative recognition accuracy. Four one-sample t-tests revealed that the mean gamma correlation was significantly greater than zero for both distinct and typical name types in the name target condition, $t(32) = 8.92$, $p < .001$, $r^2 = 0.71$, and $t(32) = 5.28$, $p < .001$, $r^2 = 0.46$, respectively, and in the face target condition, $t(29) = 7.01$, $p < .001$, $r^2 = 0.63$, and $t(29) = 4.45$, $p < .001$, $r^2 = 0.41$, respectively.

Two paired-samples t-tests were conducted on gamma and G* to determine if the predictive validity of JOLs on cued-recall accuracy for name targets significantly differed between distinct and typical name types. For gamma, the difference between typical and distinct name types approached significance, $t(30) = 1.91$, $p = .066$. For G*, the difference was significant, $t(25) = 2.06$, $p = .05$, $r^2 = 0.14$. Two 2 (name type) × 2 (target condition) mixed ANOVAs with name type as a within-subject factor were conducted on gamma and G* to determine if the predictive validity of JOLs on recognition accuracy differed among name types and target conditions. For gamma, the main effect of name type was significant, $F(1,61) = 7.38$, $p = .009$, partial $\eta^2 = 0.11$, whereas the main effect of target condition, $F(1,61) < 1$, and the interaction between name type and target condition, $F(1,61) < 1$, were not significant. For G*, there was a significant main effect of name type, $F(1,50) = 5.90$, $p = .019$, partial $\eta^2 = 0.11$, whereas the main effect of target condition, $F(1,50) < 1$, and the interaction, $F(1,50) < 1$, were not significant. Altogether, these results indicate that JOLs were more accurate predictors of cued-recall and associative recognition accuracy when the face-name pair contained a distinct name compared with a typical name.

Feeling-of-knowing judgements. Four one sample t-tests revealed that the mean gamma correlation between associative recognition accuracy and the magnitude of FOKs was significantly greater than zero for distinct and typical name types in the name target condition, $t(32) = 8.61$, $p < .001$, $r^2 = 0.70$, and $t(32) = 8.35$, $p < .001$, $r^2 = 0.69$, respectively, and in the face target condition, $t(29) = 14.15$, $p < .001$, $r^2 = 0.87$, and $t(29)$

= 11.00, $p < .001$, $r^2 = 0.81$, respectively. Two 2 (face type) × 2 (target condition) mixed ANOVAs with name type as a within-subject factor were conducted on gamma and G* to evaluate the predictive validity of FOKs on recognition accuracy. For gamma, the main effect of name type, $F(1,61) = 3.24$, $p > .05$, and the name type × target condition interaction, $F(1,61) < 1$, were not significant, whereas the main effect of target condition was significant, $F(1,61) = 4.55$, $p = .03$, partial $\eta^2 = 0.07$, with gamma significantly greater in the face target condition compared with the name target condition. This finding is analogous to the significant difference between target conditions on the relative accuracy of JOLs in Experiment 3. In this case, the relative distinctiveness of the name cues in the name-cue/face-target condition might have aided monitoring relative to the face-cue/name-target condition. Chapter 7 elaborates on this possibility. For G*, the main effect of name type, $F(1,48) = 1.39$, $p > .05$, the main effect of target condition, $F(1,48) = 1.42$, $p > .05$, and the name type × target condition interaction, $F(1,48) < 1$, were not significant.

Imagery ratings. Two one sample t-tests revealed that the predictive validity of imagery ratings was above chance for both distinct, $t(30) = 19.61$, $p < .001$, $r^2 = 0.92$, and typical $t(32) = 6.88$, $p < .001$, $r^2 = 0.59$, name types, but on the basis of gamma, $t(29) = 1.15$, $p > .05$, or G*, $t(10) < 1$, the difference between name types on the predictive validity of imagery ratings was not significant.

Retrospective confidence judgements. The predictive validity of RCJs on associative recognition accuracy was assessed with four one sample t-tests. For both distinct and typical name types, the mean gamma correlation was significantly greater than zero in the face target condition, $t(30) = 17.41$, $p < .001$, $r^2 = 0.90$, and $t(30) = 13.95$,

$p < .001$, $r^2 = 0.86$, respectively, and in the name target condition, $t(33) = 12.82$, $p < .001$, $r^2 = 0.83$, and $t(33) = 20.96$, $p < .001$, $r^2 = 0.93$, respectively. Two 2 (name type) × 2 (target condition) mixed ANOVAs with name type as a within-subject factor were conducted on gamma and G* to determine if the predictive validity of RCJs differed among name type and target conditions. For gamma, the main effect of name type, $F(1,63) < 1$, the main effect of target condition, $F(1,63) < 1$, and the interaction between face type and target condition, $F(1,63) = 2.31$, $p > .05$, were not significant. Likewise, for G*, the main effect of name type, $F(1,63) < 1$, the main effect of target condition, $F(1,63) = 3.41$, $p > .05$, and the interaction between face type and target condition, $F(1,63) = 2.99$, $p > .05$, were not significant. Thus, in contrast to Experiment 3, name distinctiveness did not improve retrospective monitoring accuracy.

Absolute accuracy.

Figure 14 contains plots of calibration lines for JOLs, FOKs and RCJs across name types and target conditions in Experiment 4. As the top left panel of Figure 14 illustrates, the absolute accuracy of JOLs on cued-recall is similar for distinct (Y = .48X + 12) and typical (Y = .43X + 15) name types. This was evaluated statistically with a 2 (name type: typical vs. distinct) × 2 (memory: cued-recall accuracy vs. JOL magnitude) repeated-measures ANOVA, which revealed that the main effects of name type, $F(1,36) = 11.46$, $p = .002$, partial $\eta^2 = 0.24$, and memory, $F(1,36) = 7.85$, $p = .008$, partial $\eta^2 = 0.18$, were significant, whereas the interaction between memory and name type only approached significance, $F(1,36) = 4.09$, $p = .053$. The main effect of name type does not necessarily contradict the non-significant difference between distinct and typical names on cued-recall accuracy. Keep in mind that JOL magnitude and cued-recall accuracy are aggregated when testing the main effect of name type, thus the significance of the main effect is likely due to JOLs rather than cued-recall.

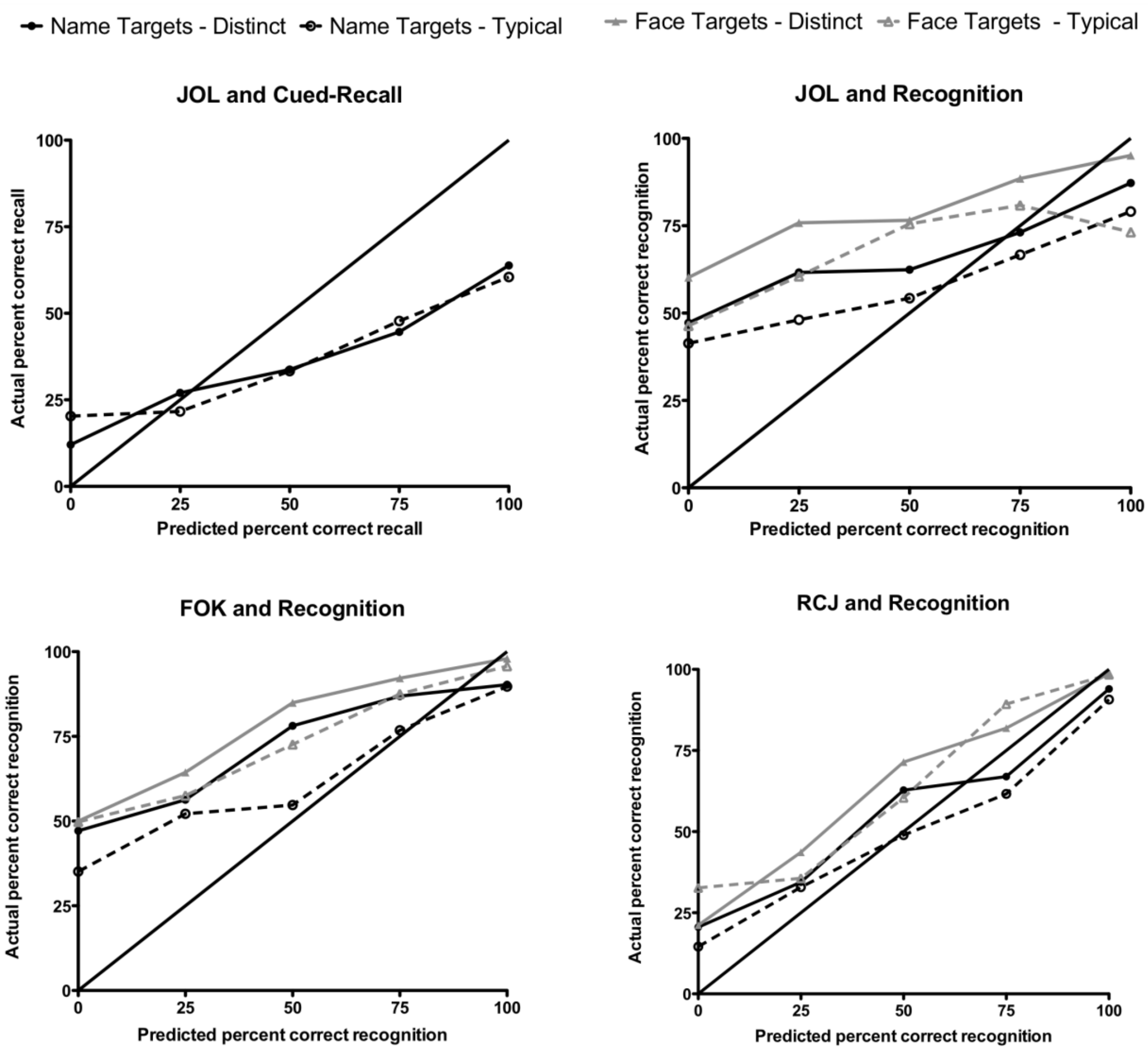

Figure 14. Calibration lines of judgements-of-learning (JOL), feeling-of-knowing judgements for all items (FOK), and retrospective confidence judgements (RCJ) across name type and target conditions in Experiment 4. Each panel contains a plot of observed memory performance as a function of predicted memory performance for each target condition. The title of each panel refers to the monitoring judgement and memory test that was used to construct the calibration lines. The solid black diagonal line refers to perfect calibration. The dotted lines refer to typical name types whereas the solid lines refer to distinct name types.

The linear parameters of the calibration lines for JOLs on associative recognition are similar across name types. For name targets, the linear parameters for distinct (Y = .37X + 48) and typical (Y = .38X + 39) name types are comparable. The same can be said for the linear parameters of face targets, as distinct (Y = .33X + 63) and typical (Y = .30X + 53) name types are alike. But across target conditions, participants appear to have underestimated associative recognition performance for face targets to a greater degree compared with name targets. The statistical significance of this pattern of results was evaluated with a 2 (name type: typical vs. distinct) × 2 (memory: associative recognition accuracy vs. JOL magnitude) × 2 (target condition: faces vs. names) mixed ANOVA with target condition as a between-subject factor. The main effects of memory, $F(1,68) = 75.49$, $p < .001$, partial $\eta^2 = 0.52$, and name type, $F(1,68) = 27.13$, $p < .001$, partial $\eta^2 = 0.29$, were significant, but none of the interactions were significant, nor was the main effect of condition. Once again, the magnitude of JOLs was significantly lower than associative recognition accuracy. The calibration lines reveal that this effect is more prominent for face targets compared with name targets

The calibration lines for FOKs also reveal underconfidence, especially in the face target condition. For name targets, distinct name types (Y = .47X + 48) yielded slightly poorer levels of absolute accuracy compared with typical name types (Y = .54X + 35). For face targets, the slopes and intercepts of the calibration lines were approximately identical between distinct (Y = .49X + 53) and typical (Y = .49X + 48) name types. In all conditions the calibration lines were consistently above the diagonal. A 2 (name type: typical vs. distinct) × 2 (memory: associative recognition accuracy vs. FOK magnitude) × 2 (target condition: faces vs. names) mixed ANOVA with target condition as a between-subject factor revealed significant main effects of memory, $F(1,69) = 156.25$, $p < .001$, partial $\eta^2 = 0.69$, name type, $F(1,69) = 18.86$, $p < .001$, partial $\eta^2 = 0.21$, and target condition, $F(1,69) = 4.35$, $p = .043$, partial $\eta^2 = 0.05$. None of the interactions were significant. The magnitude of FOKs was significantly lower than associative recognition accuracy across all conditions, and distinct names had greater FOKs and greater associative recognition accuracy compared with typical names. The main effect of target condition is likely due to the fact that recognition accuracy is significantly greater for face targets compared with name targets.

The calibration lines for RCJs are similar to those of Experiment 3. Participants are well calibrated in name target condition, with distinct (Y = .72X + 20) and typical (Y = .72X +14) name types approaching perfect calibration, but participants in the face target condition reveal under confidence, with distinct (Y = .77X + 25) and typical name types (Y = .74X + 26) above the diagonal. The statistical significance of this pattern of results was evaluated with a 2 (name type: typical vs. distinct) × 2 (memory: associative recognition accuracy vs. RCJ magnitude) × 2 (target condition: faces vs. names) mixed ANOVA with target condition as a between-subject factor. The main effects of memory, $F(1,69) = 10.57$, $p = .002$, partial $\eta^2 = 0.13$, and name type, $F(1,69) = 18.66$, $p < .001$, partial $\eta^2 = 0.21$, were significant, whereas the main effect of condition was not significant nor were any of the interactions. Across all conditions, participants underestimated the accuracy of their associative recognition decisions.

Discussion.

The results of Experiment 4 largely parallel those of Experiment 3. As expected, distinct names were better recognized that typical names, and faces associated with distinct names were better recognized than their typical counterparts. Thus, the beneficial effects of name distinctiveness (e.g. Brandt et al., 2006; Stanhope & Cohen, 1993) can transfer over to memory for faces. Indeed, analysis of the imagery ratings revealed that mental images of faces paired with distinct names were more vivid compared with mental images of faces paired with typical names. Once again, these findings are consistent with the idea that distinct items prompt elaborative encoding of the learning episode. However, in general the effect sizes were smaller for the distinctiveness effect in names (i.e. Experiment 4) compared with the distinctiveness effect in faces (i.e. Experiment 3).

Unexpectedly, the distinctiveness effect in names on cued-recall accuracy was not replicated. The failure to do so could be due to the context in which the names were learned. If there are an equal proportion of distinct and typical names, then the beneficial effects of name distinctiveness on cued-recall accuracy are attenuated (Stanhope & Cohen, 1993). In Experiment 4, the equal proportion of distinct and typical names could have diminished the effect, which might also account for the smaller effect sizes in Experiment 4 compared with Experiment 3. Alternatively, the failure to replicate could also be due to the fact that distinctiveness effects on recall memory are attenuated when distinctiveness is defined in terms of frequency (Schmidt, 1991). Participants in the pilot test may have based their name distinctiveness ratings on frequency to a greater degree than on remarkableness, which could account for the findings on the cued-recall test. Nonetheless, the distinctiveness effect in names appeared on associative recognition accuracy, and was sufficient to influence participants' ability to recognize the name or the associated face.

Importantly, the effects of name distinctiveness were not confined to memory performance. Not only do participants consider face-name pairs with a distinct name easier to learn, but also they are more confident in their prospective and retrospective memory for the pair when it contains a distinct name compared with a typical name. Thus, in the same way that distinct faces increase the magnitude of monitoring judgements relative to typical faces, distinct names also increase the magnitude of monitoring judgements relative to typical names. Ultimately, this pattern of results indicates that the participants use distinctiveness as a source of information when monitoring their memory for faces and names.

The calibration lines reveal that subjective estimates of associative recognition performance are below objective performance, especially for face targets. In like manner to Experiments 1 and 3, this effect was more pronounced in the lower estimates of memory performance (i.e. 0%, 25%, 50%). It is only when participants are 100% confident in their future or past memory performance is calibration nearly perfect. Reasons as to why participants underestimate face recognition to a greater degree compared with name recognition are discussed in Chapter 7.

The predictive validity of prospective monitoring, but not retrospective monitoring, varied according to distinctiveness. Specifically, JOLs were more accurate at distinguishing between faces or names that will be and will not be recognized when the face-name pair contained a distinct name compared with a typical name. This is in contrast with Experiment 3, where facial distinctiveness had a similar effect on the predictive validity of RCJs but not JOLs. This pattern of results highlights another difference between prospective and retrospective monitoring, and is discussed in more detail in the following chapter.

Chapter 7 – General Discussion

This chapter begins with a brief re-iteration of the rationale for conducting the present research and then discusses the global findings that are prevalent among all of the experiments. How these findings add to our understanding of metamemory and face identification is emphasized throughout. Afterwards, the idiosyncratic local findings of each experiment are addressed and related to theoretical and practical issues in metamemory and face identification. Suggestions for future research are included throughout the general discussion, and a detailed rationale for a specific research program is presented towards the end of the chapter. The chapter closes with the conclusions that can be drawn from the dissertation.

The major objective of the present research was to elucidate metamemory performance for novel face-name associations during various stages of learning. This objective was motivated by practical and theoretical reasons. First, relatively little attention has been paid to individuals' knowledge of their face identification capabilities. Monitoring and control processes are not explicitly included in theoretical models of person identification. As a result, a theoretical framework of face identification that incorporates metamemory mechanisms is lacking. Second, while some have found that participants are capable of predicting their future face recognition performance (Busey et al., 2000), others have found that individuals are quite poor at evaluating their face recognition ability (Olsson & Juslin, 1999; Perfect, 2004; Sommer et al., 1995). Thus there is conflicting evidence regarding individuals' ability to monitor their memory for faces. Finally, although there is a rich body of literature on metamemory, the majority of metamemory studies have relied on verbal paired-associates or general knowledge questions to probe monitoring and control. As a result, their findings might not generalize to tasks that involve face identification. The present research aimed to shed light on all of these issues by examining face-name learning through the prism of metamemory.

The primary aim of Experiments 1 and 2 was to determine if participants could accurately monitor their memory for faces and names, and if metamemory performance differs between face and name targets. A secondary aim of these experiments was to assess the validity of inferring metamemory performance during face-name learning from findings with verbal paired associates. To this end, participants learned a series of novel face-name or noun-noun paired associates and were solicited for a variety of monitoring judgements during different stages of learning. Metamemory performance was compared among conditions that used faces, names or nouns as items to be retrieved from memory. As a supplement to the secondary aim, Experiment 1 attempted to reproduce a reliable metamemory phenomenon, the delayed-JOL effect, using non-verbal stimuli.

Experiments 3 and 4 complemented Experiments 1 and 2 by attempting to identify sources of information that individuals might use to monitor their memory for faces and names. On the basis of empirical and theoretical reasons, the relative distinctiveness of a face or name was hypothesized to be one such source of information. As such, the primary aim of Experiments 3 and 4 was to determine if distinctiveness influences metamemory for face-name associations. Additionally, a second aim of Experiments 3 and 4 was to examine how distinctiveness affects associative memory for faces and names, such that memory for one item in a face-name pair might be enhanced if the other is distinct.

There were several points of departure between the present research and previous studies that examined metamemory performance for face-name pairs. First, multiple monitoring judgements for novel paired-associates were solicited at various stages of learning rather than only soliciting one monitoring judgement at a single stage of learning (Hosey et al., 2009; Tauber & Rhodes, 2010). Second, the accuracy of monitoring judgements was compared between face and name targets as opposed to exclusively focusing on name targets (Hosey et al., 2009; Modirrousta & Fellows, 2008; Pannu et al., 2005; Tauber & Rhodes, 2010). Finally, monitoring accuracy was assessed using both relative and absolute measures, rather than reporting only one or the other (Hosey et al., 2009; Tauber & Rhodes, 2010). By and large, the present research revealed that participants can monitor their memory during face-name learning, and monitoring is sensitive to variations of distinctiveness. These findings are elaborated upon below.

Monitoring Face-Name Associations

Several findings were common among all of the experiments. The most important of which was that participants were capable of monitoring their memory for face-name associations at a level above chance. Across all experiments, the gamma correlations between monitoring judgements and objective memory performance were significantly greater than zero. The only exception to this was EOL judgements for typical name targets in Experiment 4. Aside from this anomalous finding, all of the monitoring judgements were predictive of memory performance. Thus, participants can form direct links between faces and names, and the links are accessible to monitoring, which is in contrast to the notion that face-name associations cannot be formed without the presence of identity-related semantic information (e.g. Bruce & Young, 1986; Craigie & Hanley, 1997).

In Experiment 1, the magnitude and relative accuracy of the four monitoring judgements did not significantly differ between face and name targets. This was an unexpected result, especially considering that names were more difficult to recognize than faces, but it does highlight an interesting property of our memory for people. Namely, our memory for names is poorer compared with faces, but we can nonetheless distinguish between names that will and will not be retrieved from memory at a degree of accuracy that is comparable to faces.

In Experiments 3 and 4, the relative accuracy of JOLs or FOKs varied according to the type of target on the memory test. Specifically, JOLs were more accurate at predicting recognition performance for name targets in Experiment 3, whereas FOKs were more accurate at predicting recognition performance for face targets in Experiment 4. This is most likely due to the relative distinctiveness of the cue that was present during the solicitation of the monitoring judgement rather than a fundamental difference between faces and names in their capacity as faces and names. In Experiment 3, faces varied according to distinctiveness, and the mean gamma correlation between JOLs and associative recognition accuracy was greater for face-cues/name-targets compared with name-cues/face-targets. Conversely, in Experiment 4, names varied according to distinctiveness, and the mean gamma correlation between FOKs and associative recognition accuracy was greater for name-cues/face-targets compared with face-cues/name-targets. In both cases, the relative distinctiveness of the cues that were present during the solicitation of the monitoring judgement varied among the paired-associates.

This may have provided an additional heuristic that participants used to predict future memory performance. If so, it would account for the differences between target conditions in Experiments 3 and 4, and the discrepancy between these experiments and Experiment 1.

In contrast to relative accuracy, absolute accuracy varied according to whether a face or name must be retrieved from memory. A consistent finding amongst Experiments 1, 3, and 4 was that participants underestimated their associative recognition accuracy for face targets to a greater degree than name targets. In Experiment 1, the calibration lines of JOLs, FOKs and RCJs for face targets were predominantly above the diagonal. This was not the case for name targets, which showed fairly accurate levels of calibration, especially for judgements at the higher end of the rating scale (i.e. greater than 50%). Likewise, in Experiments 3 and 4, the calibration lines of JOLs, FOKs, and RCJs for face targets were above the diagonal to a greater degree than name targets. In some cases, the relative difference in absolute accuracy between face and name targets was confirmed statistically. Simple effects testing of the interaction between levels of target condition (i.e. faces versus names) and levels of memory performance (i.e. estimated memory performance versus actual memory performance) for FOKs and JOLs in Experiment 1 or for RCJs in Experiment 3 revealed that the difference between the mean magnitude of the monitoring judgement and mean associative recognition accuracy was greater for face targets compared with name targets. The only exception to these findings was Experiment 2, in which the absolute accuracy of RCJs did not differ among target conditions.

One possible explanation for this pattern of results is the hard – easy effect (e.g. Lichtenstein & Fischhoff, 1977; Juslin, 1993; for a review, see Juslin, Winman, &

Olsson, 2000). The hard – easy effect refers to the phenomenon whereby participants tend to underestimate their memory performance for items that are easy to retrieve from memory (e.g. when objective memory performance is above 75%) and overestimate their memory performance for items that are difficult to retrieve from memory (e.g. objective memory performance is below 60%). In Experiments 1, 3, and 4, recognition accuracy was greater for face targets compared with name targets. Thus, if associative recognition performance is used as a criterion, then faces were easier to retrieve than names, and as a result, participants may have underestimated their memory for faces compared with names. This may explain the relative differences in absolute accuracy among faces and names in Experiments 1, 3, and 4. In contrast, recognition accuracy in Experiment 2 did not differ between face and name target conditions, and in-line with the predictions of a hard – easy effect, the absolute accuracy of RCJs did not differ between face and name target conditions. In short, the greater underconfidence for face targets compared with name targets might be due to the fact that faces were easier to retrieve than names, and participants in the face target condition failed to adjust their estimates of memory performance accordingly.

Going beyond differences among target conditions, participants were generally underconfident in their associative recognition ability, and underconfidence was more prominent for lower estimates of memory performance (i.e. ratings below 75-80%). Thus, unless participants were very confident in their prospective or retrospective memory, they tended to underestimate their recognition accuracy. This pattern of results has been reported elsewhere in the literature. Thiede and Dunlosky (1994) compared JOL accuracy between conditions that used cued-recall or associative recognition as tests of objective memory performance. Regardless of the type of test that participants anticipated, the magnitudes of JOLs were consistently lower than recognition accuracy, especially for ratings below 80%. The relative accuracy of JOLs was also poorer when an associative recognition test was presented compared with a cued-recall test, which is consistent with the gamma correlations reported in the present research (see Tables 2, 5, and 7). As Thiede and Dunlosky point out, poorer monitoring accuracy on tests that involve recognition could be due to correct guessing. For instance, a participant may judge that a target item has only a 20% likelihood of being remembered, but because the target item is present at time of recognition and the participant can eliminate possible alternatives on the recognition test, his objective memory performance for that item is likely to be greater than expected. On cued-recall however, participants do not necessarily have access to the target or possible alternatives, which mitigates the effects of correct guessing. Although the recognition test in the present research was an 8-AFC, there might not have been enough distracters to prevent correct guessing.

Correct guessing may also explain the inconsistent pattern of results in the calibration of JOLs for names. Across all experiments, the calibration lines of JOLs for names on cued-recall revealed overconfidence, whereas underconfidence was evident when JOLs were plotted against associative recognition. The overconfidence on cued-recall accuracy for names is consistent with the findings of Tauber and Rhodes (2010), and suggests that participants erroneously believe that their memory for a name is better than what it actually is. But this contradicts the calibration of JOLs on associative recognition. Bearing in mind that participants were not explicitly informed of the nature of the memory test when making their JOLs, they might not have taken into consideration the possibility of correct guessing on the recognition test. As a result, their JOLs will have underestimated recognition accuracy even though they overestimated cued-recall accuracy.

The discrepancy between the absolute and relative accuracy results highlights the importance of distinguishing between both measures of monitoring accuracy (see Mengelkamp & Bannert, 2010; Schraw, 2009). Measures of relative accuracy reflect the degree to which predictions of memory performance can be used to discriminate between remembered and non-remembered items, whereas measures of absolute accuracy reflect the degree to which predictions of memory performance are in line with objective memory performance. Bearing this distinction in mind, the relative accuracy results of the present research indicate that the accuracy with which individuals discriminate between items that will be (or were) or will not be (or were not) remembered is comparable between faces and names. In regards to absolute accuracy, one can conclude that predictions of face recognition performance will tend to underestimate actual recognition performance. This will also occur for predictions of name recognition, but to a lesser degree, especially if the predictions are retrospective and highly confident.

These results complement previous studies that examined metamemory during face-name learning or face recognition. While some have found that the correlation between JOLs and old-new recognition ratings for faces is not significant when controlling for facial distinctiveness (Sommer et al., 1995), others have reported that JOLs are accurate predictors of face recognition accuracy regardless of the perceptual properties of the face (Busey et al., 2000). The results of the present research are consistent with the latter as opposed to the former. In three experiments, JOLs were significant predictors of face recognition accuracy, even for typical faces in Experiment 3 or faces associated with typical names in Experiment 4.

The analysis of EOLs and RCJs for face targets reveals an interesting relationship between the subjective difficulty of learning and recognizing faces with actual face recognition performance. Previous research within the field of eyewitness memory has reported that the confidence-accuracy relationship is weak (for reviews, see Bothwell, Deffenbacher, & Brigham, 1987; Krug, 2007; Wells & Murray, 1984) and that self-reported memory ability for faces is a poor predictor of face recognition performance (Olsson & Juslin, 1999; Perfect, 2004). However, across all of the experiments in the present research, the correlations between EOLs or RCJs on associative recognition accuracy for face targets were significantly above chance and displayed moderate effect sizes. Thus, beliefs concerning the difficulty of learning and recognizing a face when cued with a name are predictive of actual recognition performance. The disparity between the present research and previous studies could be due to the fact that I solicited judgements on an item-by-item basis, used multiple study-test trials, and tested associative recognition, whereas eyewitness studies typically solicit judgements on an aggregate of items, use three or fewer study-test trials, and test familiarity-based recognition. Such differences may prove to be important in assessing the confidence-accuracy relationship in eyewitness memory.

These findings can be used to clarify and constrain aspects of the higher order cognitive processes in models of person identification. For example, the executive functions of Bruce and Young’s (1986) model are delegated to a general purpose cognitive system. These functions include, but are not limited to, directing resources to components of the model or searching semantic memory. Monitoring and control mechanisms are presumably subsumed under the cognitive system. Considering that PINs signal the identity of a particular individual and mediate access between faces and names, PINs are likely to be the component of the model that is monitored. Indeed, PINs have bi-directional links with the cognitive system, whereas FRUs and name-codes do not.

Monitoring the activity of PINs fits nicely with the various theoretical accounts of monitoring judgements. For example, if a face cue is presented and an individual is asked to estimate the likelihood of recalling the appropriate name (i.e. a JOL), then the time it takes for the PIN to activate a name-code can be used to determine retrieval fluency. Furthermore, if the appropriate name-code cannot be accessed (i.e. a recall failure), but the PIN remains active, then an FOK is produced. The magnitude of the FOK will be proportional to strength of the signal from the FRU to the PIN (i.e. cue-familiarity) and the amount of identity-related semantic codes that the PIN activates (i.e. accessibility of partial information related to the target). In a similar manner, a name cue could prompt retrieval of a face through a backwards flow of activation throughout the architecture of the model (Groninger & Murray, 2004). First, a name-code would be activated, which would send a signal to the appropriate PIN. The PIN would then attempt to activate the appropriate FRU. The level of activation in the FRU could give rise to a mental image of the target face. As the FRU accumulates more and more activation, the clarity of the mental image would increase, and if it passes a certain threshold, then the face would be recollected. The fact that the magnitudes and relative accuracies of monitoring judgements are equivalent between face-name pairs and name-face pairs accords well with the idea that PINs are monitored as opposed to FRUs or the name-codes since PINs are the mediators between faces and names.

Of course, additional research is needed to fully elaborate the metamemory components of Bruce and Young's (1986) model. For example, little is known of the control mechanisms that govern learning the identity of an individual and how these control mechanisms relate to the activity or formation of PINs or FRUs. One novel way to advance our understanding in this area would be to investigate metamemory of people in a naturalistic setting. This would not only diminish criticisms of ecological validity (see Bruce, Burton, & Walker, 1994) that typically haunt face identification studies that use paired-associate learning paradigms, but would also allow researchers to investigate monitoring and control during learning of faces, names, and other forms of identity-related semantic information.

Imagery and face retrieval.

The present research provided several new lines of evidence to suggest that mental imagery plays a prominent role during retrieval of faces (Groninger, 2000; 2006; Lobmaier & Mast, 2008). Of considerable significance was the reliable finding that imagery ratings in the presence of name cues were significant predictors of face recognition accuracy. Intuitively, this finding makes sense. The degree to which an individual can recollect details about a face in the presence of a name will be related to his ability to recognize that face. But not only were imagery ratings correlated with memory performance, they were also correlated with metamemory performance. Imagery ratings and FOKs varied concomitantly, such that as the clarity of the mental image of the face increases, the estimated likelihood of future recognition also increases. This finding is consistent with accessibility accounts of FOKs (e.g. Koriat & Levy-Sadot, 1999). When prompted with a name cue, a mental image of a face is brought to mind. The clarity of this mental image, regardless of its objective accuracy, contributes to the subjective sense that the face will or will not be recognized. Hence the magnitude of FOKs will increase as mental image clarity increases. The findings of the present research, coupled with the fact that imagined and perceived faces activate similar neural mechanisms (Farah, Peronnet, Gonon, & Giard, 1988; Ishai, Haxby, & Ungerleider, 2002), indicate that imagery has a prominent role in face retrieval.

Retrospective versus Prospective Monitoring

Another pattern that emerged from all of the experiments was the dissociation between prospective and retrospective memory monitoring. Experiment 1 revealed that in terms of relative accuracy, retrospective memory monitoring is more accurate for nouns compared with names, whereas prospective memory monitoring is equivalent between nouns, faces, and names. Dissociations between prospective and retrospective monitoring are commonly reported in the metamemory literature (e.g. Dougherty et al., 2005; Shimamura & Squire, 1986; 1988; Thiede & Dunlosky, 1994) including studies that used facial stimuli (Busey et al., 2000; Modirrousta & Fellows, 2008; Pannu et al., 2005). Of particular relevance is the finding that extrinsic cues, such as study time, can affect the relative accuracy of prospective monitoring judgements without affecting the relative accuracy of retrospective monitoring judgements (Dougherty et al., 2005). The results of Experiments 1 expand on this finding by revealing that intrinsic cues, such as the category of stimulus, can also produce differential effects on the relative accuracy of prospective and retrospective monitoring judgements.

Along similar lines, the predictive validity of prospective and retrospective memory monitoring varied as a function of distinctiveness. Experiments 3 and 4 revealed that facial distinctiveness affected the relative accuracy of RCJs, whereas name distinctiveness affected the relative accuracy of JOLs. It seems then, that name distinctiveness is more diagnostic of future memory performance than facial distinctiveness, but that facial distinctiveness is more diagnostic of past memory performance than name distinctiveness. These findings indicate that after a brief meeting with a group of unfamiliar individuals, our predictions as to whether we will be able to recognize a particular individual are more accurate if he has an unusual or remarkable name. However, after we recognize someone who we recently met, our confidence in our recognition decision is more valid when that individual has an unusual face.

It is also worth noting that in general, the gamma correlations between the magnitude of the monitoring judgement and associative recognition performance were greater for RCJs compared with EOLs, JOLs, and FOKs (see Tables 2, 5, and 7). This occurred regardless of the category of stimuli and regardless of the distinctiveness of the target on the memory test. Others have also reported that relative accuracy is greater for retrospective monitoring compared with prospective monitoring (e.g. Dougherty et al., 2005; Pannu et al., 2005; Thiede & Dunlosky, 1994). One reason for the difference could be due to the fact that prospective monitoring judgements are made under a greater degree of uncertainty compared with retrospective monitoring judgements (Van Overschelde, 2008). That is to say, the target has yet to be accessed and incorrect alternatives have yet to be eliminated in prospective monitoring tasks, which is not the case in retrospective monitoring tasks. This uncertainty could manifest as noise during monitoring, which would attenuate accuracy.

Generalizing Metamemory Performance to Faces and Names

A secondary objective of Experiments 1 and 2 was to investigate any potential differences in metamemory performance across different categories of stimuli. As a preliminary investigation into this matter, the accuracy of four monitoring judgements was compared between face-name pairs and noun-noun pairs. Several outcomes support the hypothesis that metamemory findings from one category of stimuli can generalize to another.

First, for both face-name pairs and noun-noun pairs, the relative accuracies of four monitoring judgements (i.e. EOLs, JOLs, FOKs, and RCJs) were significantly above chance. Thus, regardless of the category of stimuli, participants can accurately estimate the likelihood of retrieving one item relative to another. Second, for each of the three prospective monitoring judgements, the magnitudes of the gamma correlations were approximately equal among name, face, and noun targets. This finding reveals that although names seem to be learned in a distinctively different manner than other types of semantic information (e.g. McWeeny et al., 1987), and are more difficult to retrieve than nouns (e.g. Burke et al., 1991), the ability to predict future retrieval performance for names is similar to that for nouns. Finally, manipulations that affected the relative accuracy of JOLs for nouns also produced similar effects for face-name pairs. Although smaller in magnitude compared with nouns, the delayed-JOL effect was replicated using faces and names as targets in an associative recognition test or using names as targets in a cued-recall test.

In contrast to prospective metamemory, retrospective metamemory was more accurate with nouns compared with names. In Experiment 1, the relative accuracy of RCJs was greater for noun targets compared with name targets. Likewise, in Experiment 2, when participants were not required to provide prospective monitoring judgements, the relative accuracy of RCJs was superior for nouns compared with names. These findings are consistent with previous research that has reported dissociations between general knowledge questions and eyewitness performance on the relative accuracy of retrospective confidence judgements (Perfect & Hollins, 1996).

The difference between face-name pairs and noun-noun pairs on the relative accuracy of retrospective monitoring could be due to the fact that participants have more experience distinguishing between correct and incorrect noun targets in memory compared with face or name targets (see Perfect, 2004 for a similar explanation). The participants in the present research were sampled from a population of university students, who are often required to study for and complete multiple choice (i.e. recognition) and fill in the blank (i.e. cued-recall) tests. While studying for recognition or cued-recall tests, students obtain practice at retrieving a target word from memory and estimating if the retrieved word is correct. In most cases, they can obtain feedback on the accuracy of their response. Therefore, participants have experience at estimating whether a retrieved target word is correct relative to other possible targets and receive feedback on the accuracy of their retrospective monitoring judgement. Now, consider the case where a participant must retrieve a face or name from memory. The opportunity to estimate whether a retrieved face or name is correct relative to other possible face or name targets occurs less frequently compared with nouns. Moreover, it is much less common for a participant to receive feedback on the accuracy of his retrospective confidence for face or name targets. Thus, the poorer relative accuracy of RCJs for face-name pairs compared with noun-noun pairs could be due to a lack of experience.

The absolute accuracy results of Experiment 1 are also consistent with the distinction between face-name monitoring and noun-noun monitoring. The precision at which JOLs, FOKs and RCJs reflected actual recognition performance was greater in the face and name target conditions compared with the noun target condition. Participants in the noun target condition underestimated their associative recognition ability to a greater degree than name or face targets. Considering that associative recognition accuracy for nouns in Experiment 1 was above 80%, but in Experiment 2, where underconfidence was not observed, associative recognition accuracy for nouns was 67%, these results could be due to a hard – easy effect. Nevertheless, the outcome of Experiments 1 and 2 suggests that generalizations of retrospective monitoring from studies that used verbal stimuli to face-name associations should be made with caution.

On a final note, the results of Experiment 1 also speak to the mechanisms underlying the delayed-JOL effect. Previous studies have found that the predictive validity of delayed JOLs is attenuated if the memory test involves recognition compared with recall (Thiede & Dunlosky, 1994). Likewise, in the present research the advantageous effects of delay on the predictive validity of JOLs were attenuated when the memory test involved associative recognition. This occurred regardless of the target condition. Thus, the delayed JOL effect appears to be more sensitive to the nature of the memory test than to the category of stimuli.

In summary, the answer to the question, "Can previous metamemory findings with verbal stimuli generalize to face identification tasks?" is: "Yes, with some qualifications." In a like manner to verbal paired-associates, individuals are capable of monitoring their memory for face-name associations above chance. Indeed, the relative accuracy of three prospective monitoring judgements does not differ among face-name pairs and noun-noun pairs. Furthermore, delayed-JOL effect, a robust metamemory phenomenon that has previously only been observed with verbal stimuli, can be replicated with face-name associations. However, the relative accuracy of retrospective monitoring is more accurate for nouns compared with faces and names, and underestimations of future and past recognition performance occur to a greater degree for nouns compared with faces and names. Thus, generalizations of retrospective relative accuracy or overall absolute accuracy should be made with caution.

Distinctiveness Affects Memory and Metamemory for Face-Name Associations

It is often reported that distinct faces and names benefit from superior memory compared to their typical counterparts (e.g. Brandt et al., 2003; 2006, for a review see Valentine, 1991). The aim of the Experiments 3 and 4 was to extend these findings into the domains of associative memory and metamemory. In Experiment 3, names were more likely to be recalled or recognized when paired with a distinct face compared with a typical face. Similarly, in Experiment 4, faces were more likely to be recognized when paired with a distinct name compared with a typical name. Moreover, the symmetrical effects of distinctiveness between faces and names were also apparent in metamemory. Participants were more confident in their prospective and retrospective memory for face-name associations that contained a distinct item compared with a typical item.

Associative memory.

Previous studies have found that facial distinctiveness can improve performance on associative memory tasks (Bruyer & Courvoisier, 1990; Pantelis et al., 2008; Valentine & Moore, 1995). The results of Experiments 3 and 4 advance these findings by providing direct evidence that face *and* name distinctiveness can enhance cued-recall and associative recognition accuracy. Of particular relevance to these results are the findings of Stanhope and Cohen's (1993) third experiment, in which they compared recall memory between distinct and typical forenames. The names were paired with unfamiliar faces, and after a learning phase, recall memory for the names was tested using free recall and cued-recall with faces acting as cues. Only under cued-recall were the effects of name distinctiveness apparent. The authors conclude that "the effect of distinctiveness is not simply to facilitate access to a name but to strengthen the name-face link." (Pg. 64). The results of Experiments 3 and 4 are directly in line with their conclusion.

I consider three possible mechanisms that might allow for the distinctiveness of one item in a face-name pair to affect memory for the other item. The first involves extending "face-space" models (e.g. Valentine, 1991; 2001) to incorporate identity-related information. Instead of only a face being represented in a multi-dimensional space, other aspects of an individual's identity, including his name or occupation, could be represented in an "identity-space". Like the "face-space", each dimension in the "identity-space" would be normally distributed around some measure of central tendency. However, the dimensions in this space would not only include physical characteristics of a face, but also identity-related information such as familiarity with a name or degree of personal relevance. A point in this space would correspond to an individual's identity.

Points in low-density regions of this space would be individuals that are outliers on some of the dimensions, that is – individuals with a particularly unusual name, face, or occupation, or that have a high degree of personal relevance (e.g. a father or a best friend). The location of an individual's point in the "identity-space" will vary as we learn more about the individual. In situations where we only learn an individual's face and name, the values on the other dimensions of identity are assumed to be equal to the value of the dimension's measure of central tendency (e.g. when we meet someone at a party, we assume that he or she is like everyone else until we learn more information about that person). In this "identity-space", an individual with a distinct name will occupy a point in a low-density region of the space. Ultimately, this will have the effect of improving memory for the face associated with the distinct name because the face is less susceptible to be confused with another from a nearby point.

The second possibility makes use of Burton and Bruce's (1993) extension of Burton, Bruce and Johnston's (1990) interactive activation and competition (IAC) model of person recognition. Briefly, the IAC model is based on connectionist architecture, where information about an individual is stored in pools of processing units. Three pools are particularly relevant for the purposes of explaining the current results: face recognition units (FRUs), name recognition units (NRUs), and person identity nodes (PINs). FRUs and NRUs store a structural code of a face and name, respectively, and become activated when there is some degree of similarity between an input and the stored structural code. PINs signal the identity of a particular individual. FRUs and NRUs are connected to PINs, but not each other. Within each pool, units are linked to each other in an inhibitory manner, whereas excitatory links form between pools. All links are bi-directional and equal strength, and activation of one unit in a pool cascades to connected units in other pools.

An explanation of the distinctiveness results in the context of this model is as follows. When a NRU of a common name is activated, PINs that are connected to that NRU also become activated, albeit to a lesser extent because of inhibition within the pool of PINs. The activation of these PINs cascades to their corresponding FRUs. Thus, when a common name is activated, several face representations are also activated. This has the effect of increasing the noise with the pool of FRUs, making it difficult for the target FRU associated with the common name to reach asymptotic activation. Now consider the case where a NRU of an uncommon name is activated. A smaller number of PINs will be connected to the NRU of an uncommon name. As a result, it will be easier for the target FRU's activation to reach asymptote. The same reasoning applies to effects of facial distinctiveness. The number of FRUs that are similar to a typical face will be greater than the number of FRUs that are similar to a distinct face. The activation of several FRUs will eventually cascade to the pool of NRUs, which will increase the amount noise within the pool. Once again, this will have the effect of making it more difficult for a target NRU to reach asymptotic activation when presented with a typical face input compared with a distinct face input. In short, the extension of the IAC model can explain how the beneficial effects of distinctiveness transfer between faces and names by the fact that a distinct item activates fewer representations in memory relative to a typical item, which subsequently reduces the amount of noise at the time of retrieval for the target associated with the distinct item.

A final possibility as to how distinctiveness might strengthen face-name associations draws on general theoretical accounts of distinctiveness. According to Schmidt's (1991) incongruity hypothesis, distinctiveness effects arise from encoding items that are incongruent with an active conceptual framework. The incongruity prompts increased attention to the item, as well as elaboration and rehearsal of information related to the event that gave rise to the incongruity. The additional cognitive resources allocated to the incongruent item results in superior memory performance for that item compared with the congruent items. In the case where a distinct name or face is presented alongside typical names or faces, the distinct face or name will be incongruent with typical members of its class. The incongruity prompts elaborative encoding of the learning episode, which in the present case involves a face-name association. Thus, the face-name association as a whole benefits when the face or name is distinct. By allocating additional cognitive resources to the context in which the incongruity arose, memory for an item associated with a distinct face or name is improved compared with an item associated with a typical face or name.

It would be interesting to determine whether the locus of distinctiveness effects in face identification tasks is during encoding (e.g. Brandt et al., 2003; 2006; Hirshman et al., 1989; Schmidt, 1991), which the incongruity hypothesis seems to favor, or during retrieval (e.g. Burton & Bruce, 1993; Hunt & McDaniel, 1993; Valentine, 1991), which the "identity-space" and IAC explanations seem to favor. Of course, as McDaniel and Geraci (2006) point out, it is likely that that both encoding and retrieval factors play a role in distinctiveness effects and the present results do not rule out this possibility.

Ultimately, additional research is required to unify empirical findings in face identification with theoretical accounts of distinctiveness.

The memory-enhancing effect of name distinctiveness on faces is at odds with the findings of Diana and Reder (2006). In their first experiment, low or high frequency words were superimposed over pictures of objects and participants were instructed to read aloud the word in the picture-word pair. Recognition memory for the pictures alone and the words alone was tested using an old/new paradigm. Low frequency words had more hits and fewer false alarms than high frequency words, thus reproducing a mirror effect (e.g. Glanzer & Adams, 1990; Guttentag & Carroll, 1997). However, pictures presented with low frequency words had fewer hits compared with pictures presented with high frequency words. Diana and Reder (2006) argue that the detrimental effect of low frequency words on memory for associated pictures is due to resource limitations at the time of encoding. Low frequency words demand more attentional resources at encoding compared with high frequency words (de Zuicaray et al., 2005). The additional resources that are allocated to the low frequency word occur at the expense of encoding the associated picture, which results in poorer memory for the picture.

Turning to the present research, Experiment 4 revealed that faces are more likely to be recognized if they are associated with a distinct name compared with a typical name. Distinct names, like low frequency words, are thought to demand additional cognitive resources during encoding compared with typical names (Brandt et al., 2006). Also, distinct names, like low frequency words (Gardiner & Java, 1990), are more likely to evoke recollection during recognition as opposed to familiarity (Brandt et al., 2006). On the basis of Diana and Reder's (2006) results, one would expect that recognition memory for faces should have been poorer if they were associated with a distinct name compared with a typical name. The disparity between the present research and Diana and Reder could be due to the nature of the recognition test. In Experiment 4, a cue (i.e. a name) was presented at the time of recognition, which was not the case in Diana and Reder's study. If the cue was distinct, it might have prompted recollection of the context in which the face-name pair was encoded, which subsequently enhanced recognition for the target face and mitigated the effects of resource limitations during encoding.

Metamemory.

Distinctiveness appears to be a source of information that participants use to estimate their associative memory for faces and names. The magnitude of four monitoring judgements was greater when a face or name in a face-name pair was distinct compared to when it was typical. This occurred regardless of whether the distinct item was a cue or target on associative memory tasks. Moreover, facial distinctiveness affected how participants control their learning: study time duration was lower for face-name pairs that contained a distinct face compared with a typical face. These results are consistent with previous studies that have shown that JOLs and RCJs are sensitive to facial distinctiveness (Sommer et al., 1995; Semmler & Brewer, 2006). However, these results go beyond these studies in several ways. First, they reveal that other metamemory judgements, namely EOLs, FOKs, and imagery ratings, are also sensitive to facial distinctiveness. Second, they reveal that name distinctiveness exerts similar affects as face distinctiveness on the magnitude of monitoring judgements. Finally, they reveal that facial distinctiveness affects control in addition to monitoring.

These findings bear on theoretical issues in distinctiveness research. Others have found that JOLs are not sensitive to isolation effects (e.g. seeing a red letter in the context black letters) when the isolate occurs at the beginning of a list (Dunlosky, Hunt, & Clark, 2000). This finding can be interpreted as evidence that distinct items are not salient at the time of encoding (Hunt, 2006), which undermines encoding explanations of distinctiveness. However, as McDaniel and Geraci, (2006) point out, different definitions of distinctiveness may differentially rely on encoding and retrieval operations. Using Schmidt's (1991) classification of distinctiveness, they argue that encoding and retrieval processes underlie secondary distinctiveness effects (i.e. memory is superior for atypical items of an active conceptual class compared with typical counterparts), whereas only retrieval processes underlie primary distinctiveness effects (i.e. memory is superior for items that are atypical with respect to the immediate context). The isolation effect is an instance of primary distinctiveness; therefore, isolated items are not expected to be salient at the time of encoding, which would account for Dunlosky et al.'s (2000) findings. Conversely, distinct names and faces are instances of secondary distinctiveness, and are expected to be salient at the time of encoding. Indeed, McDaniel and Geraci (2006) explicitly predict that JOLs should be greater for items that reflect secondary distinctiveness. Clearly, the results of Experiments 3 and 4 support their prediction.

The findings of Experiments 3 and 4 can also be incorporated into current theoretical accounts of metamemory. First, the results are consistent with the predictions of Koriat's (1997) cue-utilization theory. Recall that this theory posits that JOLs are based upon a variety of cues and heuristics that are available during the solicitation of the judgement. Intrinsic cues, which reflect the relative recallability of items within a list, are thought to be particularly influential to JOLs. The distinctiveness of an item is an intrinsic cue; thus, it is not surprising then that JOLs were sensitive to variations of face and name distinctiveness. Second, the accessibility account of FOK (Koriat, 1993; 1995) contends that FOKs reflect the amount of partial information that comes to mind when attempting to retrieve a target. One could argue that the amount of partial information related to distinct face-name pairs is greater compared with typical face-name pairs because distinct items prompts elaborative encoding (e.g. Brandt et al., 2003; 2006) and recollection (Rajaram, 1996) of the learning episode. If this were correct, then it would account for the fact that the magnitude of FOKs was greater for face-name pairs that contain a distinct face or name compared to a typical face or name. Alternatively, the greater magnitude of FOKs for distinct items could be due to the smaller target set size compared with typical items (Schreiber, 1998). Finally, the ease with which an answer comes to mind is thought to be one basis for RCJs (Kelley & Lindsay, 1993; Nelson et al., 1990). In Experiments 3 and 4, response latencies on the associative recognition test were lower for face-name pairs that contained a distinct item compared with a typical item. On the basis of these findings, one could argue that the reason why RCJs were greater for distinct face-name pairs compared to their typical counterparts was because they were easier to retrieve.

In summary, Experiments 3 and 4 shed light on the sources of information that individuals use to monitor their memory for faces and names. Individuals believe that their associative memory for faces and names improves if the face or name is distinct compared to when it is typical. This belief is concordant with associative recognition memory performance. Additionally, the results of Experiments 3 and 4 support

Experiments 1 and 2 by showing that participants can monitor their memory for face-name associations at a level above chance, but face and name distinctiveness differentially affect the predictive validity of prospective and retrospective monitoring judgements.

Future Research

Much more research is required to delineate the subjective concomitants of perceptual memory in general and face identification in particular. One research program that would clarify and complement several aspects of the present work involves examining the perceptual strategies that participants use to learn new faces. Face perception depends on two types of facial information: configural and featural. Featural information typically refers to the anatomically distinct facial features, such as the shape of an eye or the size of the mouth. Although a universally accepted operational definition of configural information has yet to be proposed (for a review, see Maurer, Le Grand, & Mondloch, 2002), configural information typically refers to the metric distances between facial features, such has the distance between the eyes or the distance between the eyes and the mouth. This type of information has also been referred to as “relational” (e.g. Leder & Bruce, 2000) or “second-order” (e.g. Diamond & Carey, 1986) or “spatial-relational” (Searcy & Barlett, 1996).

The face inversion effect was instrumental in distinguishing between featural and configural information. Numerous studies have found that face recognition is disproportionately affected by stimulus orientation compared with object recognition (e.g. Boutet, Collin, & Faubert, 2003; Diamond & Carey, 1986; Yin, 1969). The detrimental effects of inversion in face recognition were thought to occur because configural processing is more sensitive to orientation compared to featural processing and recognition of faces relies more on the relative positions of face features than on the appearance of the individual face parts. Subsequent studies into the face inversion effect support the relative importance of configural information compared with featural information in face recognition (Colishaw & Hole, 2002; Friere, Lee, & Symons, 2000; Leder & Bruce, 2000; Rhodes, Brake, & Atkison, 1993; Searcy & Barlett, 1996; Tanaka & Sengco, 1997; Watier, Collin, & Boutet, 2010).

Although the importance of configural information has been shown experimentally, individuals do not seem to be aware that they rely on the configuration of facial features when they recognize a face. Winograd (1981) reported that the majority of first-year psychology students believe that focusing on a physical feature is the best strategy for recognizing faces. Laugherty, Duval, & Wogalter (1986) also found that the majority of participants rely on a feature/analytic strategy when attempting to recall a face from memory. Consistent with these findings, participants believe that distinctive feature judgements result in superior recognition performance (Sporer, 1991). However, not all studies have found a bias for a featural encoding strategy. In Olsson and Juslin (1999), the majority of participants reported using an encoding strategy that emphasizes the holistic properties of a face following an eyewitness task.

On the basis of the aforementioned results, it appears that the subjective experience of face recognition does not always correspond with what is found experimentally. Configural information is critical for face recognition, yet attending to featural information is a commonly reported encoding strategy. Comparing metamemory performance for faces that can only be distinguished based on featural information or configural information is one avenue for determining the most effective encoding strategies used by participants. If participants adapt their encoding strategy depending on the properties of the face, then their prospective and retrospective monitoring judgements for featurally or configurally distinct faces should be in line with their performance. However, if participants erroneously rely on only one strategy, then the accuracy of their monitoring might vary depending on the properties of the face. Follow-up studies could examine the extent to which monitoring and control is effected by training or by explicitly informing participants of the optimal face processing strategy.

Another area of face identification research that could benefit from exploring monitoring and control processes is the verbal-overshadowing effect (VOE). Verbally describing a face can subsequently impair recognition memory for that face (Schooler & Engstler-Schooler, 1990). Within the literature, this effect is referred to as "verbal-overshadowing" because it is argued that the visual memory for a difficult to describe stimulus is ignored in favor of its verbal memory. In a typical verbal-overshadowing task, participants are presented with a face and are either subsequently asked to describe the face for five minutes following an unrelated filler task or continue to work on the unrelated filler task for five more minutes. A forced-choice recognition test is then administered where participants have to identify the face that was previously presented to them. Participants who describe a face subsequently perform poorer on the recognition test compared with participants who do not describe a face (Brandimonte, Schooler, & Gabbino, 1997; Dodson, Johnson, & Schooler, 1997; Fallshore & Schooler, 1995; Finger & Pezdeck, 1999; Schooler & Engstler-Schooler, 1990).

Inaccurate monitoring after study may contribute to the VOE. In the majority of VOE studies, participants are forced to spend five minutes describing a face, which may inflate their confidence in their prospective memory for that face. At the time of the recognition test, participants would erroneously expect their memory for the target face to be enhanced. This may lead participants to give greater weight to facial features that come to mind, irrespective of whether they belong to the target, or give a greater weight to their first choice as a possible target without fully considering the other alternatives.

Strong support for the influence of monitoring effectiveness comes from two experiments conducted by Meissner, Brigham and Kelly (2001). In their first experiment, the instructions for the description task were manipulated to either encourage accuracy or quantity. Specifically, participants were required to fill 25 lines of descriptors, and therefore raise the likelihood that erroneous descriptors would be provided, or participants were encourage to only give accurate descriptors that they were confident in. Aside from manipulating the instructions, the standard verbal-overshadowing paradigm was followed. They found that recognition accuracy was significantly poorer in the quantity condition compared to the accuracy condition. Instructing participants to fill 25 lines of descriptors may have disproportionately increased their confidence in recognizing a face or encouraged participants to rely on erroneous facial information, which subsequently decreased their recognition accuracy. Indeed, in a second experiment the authors found that if participants are explicitly told to monitor the source of their memory prior to the recognition task, the VOE can be attenuated. Based on these results, it appears that when participants are encouraged to be conservative in their description accuracy or reflect on the content of their memories, the VOE does not occur. Suffice it to say, metamemory may play a role in the interfering effects of verbalization and warrants future research.

Conclusions

The results of the experiments reported in this dissertation indicate that subjective estimates of memory performance for faces and names are valid predictors of objective memory performance. After learning a set of faces and names, individuals can accurately predict which faces or names will or will not be remembered at some point in the future. However, subjective estimates of face recognition accuracy will tend to underestimate objective face recognition accuracy. A second conclusion is that prospective and retrospective metamemory are differentially affected by the type of item that must be retrieved from memory. Generalizations of retrospective metamemory from nouns to faces and names should be made with caution. Finally, a tertiary conclusion that can be drawn from this work is that face and name distinctiveness not only strengthens the association between a face and a name, but are also sources of information that individuals use to monitor their memory for faces and names.

References

Bahrick, H.P. (1984). Memory for people. In J.E. Harris & P.E. Morris (Eds.), *Everyday memory, actions and absentmindedness* (pp. 19-34). London: Academic Press.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*, 610-632.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory and Cognition*, *31*(2), 297-305.

Benjamin, A.S., & Bjork, R.A. (1996). Retreival fluency as a metacognitive index. In L. Reder (Ed.), *Metacognition and implicit memory* (pp. 309-338). Hillsdale, NJ: Erlbaum.

Benjamin, A. S. & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Memory and Metamemory* (pp. 73-94). New York, New York: Psychology Press.

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied, 5,* 76-88.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72,* 691-695.

Boutet, I., Collin, C.A., & Faubert, J. (2003). Configural face encoding and spatial frequency information. *Perception & Psychophysics, 65,* 1078-1093.

Brandimonte, M.A., Schooler, J.W., Gabbino, P. (1997). Attenuating verbal-overshadowing through visual retrieval cues. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*, 915-231.

Brandt, K.R., Gardiner, J.M., & Macrae, C.N. (2006). The distinctiveness effect in forenames: The role of subjective experiences and recognition memory. *British Journal of Psychology*, 97, 269-280.

Brandt, K.R., Macrae, C.N., Schloerscheidt A.M., & Milne, A.B. (2003). Remembering or knowing others? Person recognition and recollective experience. *Memory,11*(1), 89-100.

Bredart, S., & Valentine, T. (1998). Descriptiveness and Proper Name Retrieval. *Memory, 6*(2), 199-206.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11-30.

Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behaviour*, *5,* 325-337.

Brown, C., & Lloyd-Jones, T.J. (2005). Verbal facilitation of face recognition. *Memory & Cognition, 33*(8*)*, 1442-1456.

Brown, C., & Lloyd-Jones, T.J. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Memory & Cognition,* 34(*2),* 277-286.

Brown, M.B., & Forsythe, A.B. (1974). The small sample behaviour of some statistics which test the equality of several means. *Technometrics*, *16*, 129-132.

Bruce, V., Burton, A.M. & Dench, N. (1994). What's distinctive about a distinctive face? *Quarterly Journal of Experimental Psychology***,** *47A,* 119-141.

Bruce, V., Burton, A.M., & Walker, S. (1994). Testing the models? New data and commentary on Stanhope & Cohen (1993). *British Journal of Psychology*, *85*(3), 335-349.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Bruyer, R., & Courvoisier, P. (1990). Distinctiveness of Faces or Distinctiveness of Persons? *Perceptual and Motor Skills*, *71,* 967-984.

Burke, D., MacKay, D. G., Worthley, J., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30*, 542-579.

Burton, A.M. (1994). Learning new faces in an interactive activation and competition model. *Visual Cognition*, 1(*2/3)*, 313-348.

Burton, A.M., & Bruce, V. (1993). Naming faces and naming names: exploring an interactive activation model of person recognition. *Memory, 1*(4), 457-480.

Burton, A.M. & Bruce, V. (1992). I recognize your face but I can't remember your name: A simple explanation? *British Journal of Psychology, 83,* 45-60.

Burton, A.M., Bruce, V., & Hancock, P.J.B. (1999). From Pixels to People: A Model of Familiar Face Recognition. *Cognitive Science, 23*(1), 1-31.

Burton, A.M, Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81*, 361-380.

Burton, A.M., Young, A.W., Bruce, V., Johnston, R.A. & Ellis, A.W. (1991). Understanding covert recognition. *Cognition, 39*, 129-166.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*, 26-48.

de Carvahlo, M.K., & Yuzawa, M. (2001). The Effect of Social Influences and General Metacognitive Knowledge on Metamemory Judgements. *Contemporary Educational Psychology*,*26*, 571-587.

Cavanaugh, J. C. (1989). The importance of awareness in memory aging. In L. W. Poon, D. Rubin, & B. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 416-436). New York: Cambridge University Press.

Chua, E.F., Schacter, D.L., & Sperling, R.A. (2009). Neural correlates of metamemory: a comparison of feeling-of-knowing and retrospective confidence judgements. *Journal of Cognitive Neuroscience*, *21*(9), 1751-1765.

Clark, S. E. (1997). A familiarity-based account of confidence accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 232-238.

Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology, 81*, 287-297.

Cohen, M. E., & Carr, W. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, *6*, 383–384.

Cohen, G. & Faulkner, D. (1986). Memory for proper names: Age difference in retrieval. *British Journal of Developmental Psychology, 4,* 187-197.

Coin, C., & Tiberghien, G. (1997). Encoding activity and face recognition. *Memory*, 5, 545-568.

Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240-247.

Collishaw, S.M., & Hole, G.J. (2002) Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception, 31*(3), 287-296.

Connor, L.T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*,*12*(1), 50-71.

Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 142-150.

Courtois, M. R., & Mueller, J. H. (1981). Target and distractor typicality in facial recognition. *Journal of Applied Psychology, 66*, 639–645.

Craigie, M., & Hanley, R.J. (1997). Putting faces to names. *The British Journal of Psychology, 88*, 157-171.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about the relationship? *Law and Human Behavior, 4,* 243-260.

Deffenbacher, K.A., Carr, T, H., & Leu, J.R. (1981). Memory for Words, Pictures, and Faces: Retroactive Interference, Forgetting, and Reminiscence. *Journal of Experimental Psychology: Human Learning and Memory,7*(4), 299-305.

Dewhurst, S.A., & Conway, M.A. (1994). Pictures, Images, and Recollective Experience. *Journal of Experimental Psychology: Human Learning and Memory, 20*(5), 1088-1098.

Dewhurst, S.A., Hay, D.C., & Wickham, L.H. (2005). Distinctiveness, typicality, and recollective experience in face recognition: A principal components analysis. *Psychonomic Bulletin & Review*, *12*(6), 1032-1037.

de Zubicaray, G.I., McMahon, K.L., Eastburn, M.M., Finnigan, S., & Humphreys, M. (2005). fMRI evidence of word frequency and strength effects during episodic memory encoding. *Cognitive Brain Research*, *22*, 439-450.

Diana, R. & Reder, L.M. (2004). Visual vs. verbal metacognition: Are they really different? In Levin, D.T. (Ed), *Thinking and Seeing: Visual Metacognition in Adults and Children* (pp. 187-201). Cambridge: MIT Press.

Diana, R.A., & Reder, L.M. (2006). The Low-Frequency Encoding Disadvantage: Word Frequency Affects Processing Demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 805-815.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General, 115,* 107-117.

Dien, J. (2009). A tale of two recognition systems: Implications of the fusiform face area and the visual form area for lateralized object recognition models. *Neuropsychologia, 47,* 1-16.

Dodson, C.S., Johnson M.K., & Schooler J.W. (1997). The verbal overshadowing effect:

why descriptions impair face recognition. *Memory & Cognition, 25*, 129-139.

Dougherty, M.R., Scheck, P., Nelson, T.O., & Narens, L.(2005). Using the past to predict the future. *Memory & Cognition,33*(6), 1096-1115.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249-276). Hillsdale, NJ: Erlbaum.

Dunlosky, J., Hunt, R. R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 649-657.

Dunlosky, J., & Nelson, T.O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 374-380.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33*, 545-565.

Dunlosky, J., & Matvey, G. (2001). Empirical Analysis of the Intrinsic-Extrinsic Distinction of Judgments of Learning (JOLs): Effects of Relatedness and Serial Position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition,27*(5), 1180-1191.

Ellis, H. D., Shepherd, J. W., Gibling, F, & Shepherd, J. (1988). Stimulus factors in face learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 136-144). Chichester, UK: Wiley.

Fallshore, M., & Schooler J.W. (1995). The verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning Memory and Cognition, 21*, 1608-1623.

Farah, M.J. (2004). *Visual Agnosia 2nd ed*. Cambridge, MA: The MIT Press.

Farah, M.J., Peronnet, F., Gonon, M.A., & Giard, M.H. (1988). Electrophysiological evidence for a shared representational Medium for visual images and visual percepts. *Journal of Experimental Psychology: General, 117*, 248–257

Field, A. (2009). *Discovering Statistics Using SPSS, 3rd edition*. Sage Publications: California.

Finn, B., & Metcalfe, J. (2007). The Role of Memory for Past Test in Underconfidence with Practice Effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*(1), 238-244.

Finger, K. (2002). Mazes and music: Using perceptual processing to release verbal overshadowing. *Applied Cognitive Psychology. Special Issue: Investigation of the effects of verbalization on memory, 16*(8), 887-896.

Finger, K., & Pezdek, K. (1999). The effect of verbal description on face identification accuracy: release from verbal overshadowing. *Journal of Applied Psychology, 84*, 340-348.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906-911.

Friere, A., Lee, K., & Symons, L. (2000). The face inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception, 29*, 159-170.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*, 23-30.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506-528.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 5-16.

Gottman, J.M., Katz, L.F., & Hooven, C. (1996). Parental meta-emotion philosophy and the emotional life of families: Theoretical models and preliminary data. *Journal of Family Psychology*, *10*(3), 243-268.

Groninger, L. D. (2000). Face-name mediated learning and long-term retention: the role of images and imagery processes. *The American Journal of Psychology, 113*, 199-219.

Groninger, L. D. (2006). Variables influencing memory over long periods of time for recently learned face-name pairs. *The American Journal of Psychology*, *119*, 175-192.

Groninger, L.D., & Murray, K.N. (2004). Reminiscence, forgetting, and hypermnesia using face-name learning: Isolating the effects using recall and recognition memory measures. *Memory*, *12*(3), 351-365.

Guttentag, R., & Carroll, D. (1998). Memorability judgments for high- and low-frequency words. *Memory & Cognition, 26*, 951-958.

Hacker, D. (1998). Definitions and empirical foundations. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1-23). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hanley, J.R. & Cowell, E.S. (1988). The effects of different types of retrieval cues on the recall of names of famous faces. *Memory & Cognition, 16*, 545-555.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56, 208-216.

Hart, J.T. (1967). Memory and memory-monitoring process. *The Journal of Verbal Learning and Verbal Behaviour, 6*, 685-691.

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences,1*(2), 78-82.

Hirshman, E., Whelley, M.M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory & Language, 28*, 594-609.

Hosey, L.A., Peynircioğlu, Z.F., & Rabinovitz, B.E. (2009). Feeling of knowing for names in response to faces. *Acta Psychologica*, *130,* 214 - 224.

Hunt, R.R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3-26). New York: Oxford University Press.

Hunt, R.R., & Ellis, H.C. (2004). *Fundamentals of Cognitive Psychology*, *7th Edition*. New York, New York: McGraw-Hill Higher Education.

Hunt, R.R., & McDaniel, M.A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421-445.

Ishai, A., Haxby, J.V., & Ungerleider, L.G. (2002).Visual imagery of famous faces: Effects of memory and attention revealed by fMRI. *Neuroimage, 17*, 1729–1741.

Izaute, M., Chambers, P., & Larochelle, S. (2002). Feeling-of-Knowing for Proper Names. *Canadian Journal of Experimental Psychology, 56*(4), 263-272.

James, W. (1890). *Principles of psychology*. New York: Henry Holt & Co.

Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1990). The influence of subthreshold priming on metamemory and recall. *Acta Psychologica, 73,* 55-68.

Jonsson, F.U., & Olsson, M.J. (2003). Olfactory Metacognition. *Chemical Senses, 28* (*7*),651-658.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological Review, 107*, 384–396.

Kelemen, W.L., Frost, P.J., & Weaver, C.A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, *28*(1),92-107.

Kelemen, W.L., & Weaver, C.A. (1997). Enhanced Metamemory at Delays: Why Do Judgments of Learning Improve Over Time? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*(6).

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1-24.

Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica, 77*, 217-273.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93,* 329-343.

Kim, J.J., Andreasen, N.C., O'Leary, D.S., Wiser, A.K., Boles Ponto, L.L., Watkins, G.L, & Hichwa, R.D. (1999). Direct comparison of the neural substrates of recognition memory for words and faces. *Brain, 122*, 1069-1083.

Kluwe, R. H. (1982). Cognitive knowledge and executive control: Metacognition. In D. R. Griffin (Ed.), *Animal mind–human mind* (pp. 201–224). New York, NY: Springer–Verlag

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609-639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124,* 311-333.

Koriat, A. (1997). Monitoring one's knowledge during study: A cue utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349-370.

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

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297-315.

Koriat, A., & Goldsmith, M. (1996a). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences, 19*, 167-228.

Koriat, A., & Goldsmith, M. (1996b). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(*1*), 34-53.

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 583-402). New York, NY: Guilford Press.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107-118.

Koriat, A., & Lieblich, I. (1977). A study of memory pointers. *Acta Psychologica*, *41*, 151-164.

Koriat, A., Sheffer, L., & Ma’ayan, H. (2002). Comparing Objective and Subjective Learning Curves: Judgments of Learning Exhibit Increased Underconfidence With Practice. *Journal of Experimental Psychology: General.131*, 147-162.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52,* 478-492.

Koriat, A., Ma’ayan, H., & Nussinson, R. (2006). The Intricate Relationships Between Monitoring and Control in Metacognition: Lessons for the Cause-and-Effect Relation between Subjective Experience and Behavior. *Journal of Experimental Psychology: General*, *135,* 36-69.

Krinsky, R., & Nelson, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica, 58*, 141-158.

Krug, K. (2007). The Relationship Between Confidence and Accuracy: Current Thoughts of the Literature and a New Area of Research. *Applied Psychology in Criminal Justice, 3*(1),7-41.

Krug, K. S., & Weaver, C. A. (2005). Eyewitness memory and product identification: Evidence for familiarity biases. *Journal of General Psychology, 132*, 429-445.

Laughery, K.R., Duval, C., & Wogalter, M.S. (1986). Dynamics of facial recall. In H.D. Ellis, M.A. Jeeves, F. Newcombe, & A.W. Young (Eds.), *Aspects of face processing* (pp.373-387). Boston, MA: Kluwer.

Leonesio, J. R., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 464-470.

Leder, H.,& Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology,53A(*2): 513-536.

Levin, D.T. (2004). *Thinking and seeing: Visual metacognition in adults and children*. Cambridge, MA: MIT Press.

Lewis, M.B. (2004). Face-space-R: Towards a unified account of face recognition. *Visual Cognition, 11*(1), 29-69.

Lewis, M. B., & Johnston, R. A. (1999). A unified account of the effect of caricaturing faces. *Visual Cognition, 6*(1), 1-41.

Lichtenstein, S., & Fischoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, *20*, 159-183.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-384). New York: Cambridge University Press.

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning & Memory, 5*, 212–228.

Lindsay, D.S., Read, D.J., & Sharma, K. (1998). Accuracy and Confidence in Person Identification: The Relationship is Strong when Witnessing Conditions Vary Widely. *Psychological Science*, *9*, 215-218.

Lobmaier, J.S., & Mast, F.W. (2008). Face Imagery is Based on Featural Representations. *Experimental Psychology*, *55*(1), 47-53.

Lovelace, E.A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 756-766.

Lupker, S.J., Harbluk, J.L., & Patrick, A.S. (1991). Memory for Things Forgotten. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 897-907.

Maki, R.H., & Berry, S.L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition,10,* 663-679.

Maki, R. H., Shields, M., Wheeler, A.E., & Zacchilli, T, L. (2005). Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology*, *97*(4), 723-731.

Mantyla, T. (1997). Recollections of Faces: Remembering Differences and Knowing Similarities. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*(5), 756-766.

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytical basis for JOLs? *Memory & Cognition, 29*, 222-233.

Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255-260.

Mazzoni, G., & Nelson, T. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory & Cognition, 21*, 1263–1274.

McDaniel, M. A., & Geraci, L. (2006). Encoding and retrieval processes in distinctiveness effects: Toward an integrative framework. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory*. (pp. 65-88). New York: Oxford University Press.

McWeeny, K.H., Young, A.W., Hay, D.C. & Ellis, A.W. (1987). Putting names to faces. *British Journal of Psychology, 78*, 143-149.

Meissner C., & Brigham J. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15,*60-616.

Meissner, C., Brigham J., Kelly, C. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29(*1*), 176-186.*

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94,* 339-354.

Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: Stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, *38*(4), 441-451.

Metcalfe, J. (2000). Metamemory: Theory and data. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 197-211). London: Oxford University Press.

Metcalfe, J., & Finn, B. (2008). Familiarity and Retreival Processes in Delayed Judgements of Learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 34(*5*),* 1084-1097.

Miner, A.C., & Reder, L.M. (1994). A new look at feeling of knowing: Its metacogntive role in regulating question answering. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47-70). Cambridge, MA: MIT Press.

Modirrousta, M., & Fellows, L.K. (2008). Media prefrontal cortex plays a critical and selective role in 'feeling of knowing' meta-memory judgements. *Neuropsychologia, 46,* 2958 - 2965.

Narens, L., Jameson, K. A., & Lee, V. A. (1994). Subthreshold priming and memory monitoring. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 71-92). Cambridge, MA: MIT Press.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of knowing predictions. *Psychological Review*, *95,* 109-133.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *5*, 102-116.

Nelson, T.O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, *2*, 267-270.

Nelson, T.O., Dunlosky, J., White, D., Steinberg, J., Townes, B.D., & Anderson, D. (1990). Cognition and metacognition at extreme altitudes on Mt. Everest. *Journal of Experimental Psychology*: *General*, *119,* 367-374.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 676-686.

Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. F., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 279-288.

Nelson, T.O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In Bower G. H. (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-141). New York: Academic Press.

Nelson, T. O., & Narens, L. (1994). Why Investigate Metacognition? In J. Metcalfe and A. P. Shimamura (Eds.) *Metacognition* (pp. 1-26). Cambridge, MA: The MIT Press.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment Recall And Monitoring (PRAM). *Psychological Methods*, *9*, 53-69.

Nelson, T.O., Gerler, D., & Narens, L. (1984). Accuracy of Feeling-of-Knowing Judgements for Predicting Perceptual Identification and Relearning. *Journal of Experimental Psychology: General*,*113*, 282-300.

Olsson, N., & Juslin, P. (1999). Can Self-Reported Encoding Strategy and Recognition Skill be Diagnostic of Performance in Eyewitness Identifications? *Journal of Applied Psychology*, 84(*1*), 42-49.

Paivio, A. (1991). Dual Coding Theory: Retrospect and Current Status. *Canadian Journal of Psychology*, 45(*3)*, 255-287.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1, Pt. 2), 1-25.

Pannu, J.K., & Kaszniak, A.W. (2005). Metamemory Experiments in Neurological Populations: A Review. *Neuropsychology Review*,*15*(3), 105-130.

Pannu, J.K., Kaszniak, A.W., & Rapcsak, S.Z. (2005). Metamemory for faces following frontal lobe damage. *Journal of the International Neuropsychological Society*,*11*, 668-676.

Pantelis, P. C., van Vugt, M. K., Sekuler, R., Wilson, H. R., & Kahana, M. J. (2008). Why are some people's names easier to learn than others? The effects of face similarity on memory for face-name associations. *Memory & Cognition, 36*(6), 1182-1195.

Pappas, B. A., Sunderland, T., Weingartner, H. M., Vitiello, B., Martinson, H., and

Putnam, K. (1992). Alzheimer's disease and the feeling-of-knowing for knowledge and episodic memory. *Journal of Gerontology: Psychological Sciences*, *47*, 159–164.

Pelli, D.G. (1999). Close encounters: An artist shows that size affects shape. *Science*, *285*, 844- 846.

Perfect, T.J. (2004). The Role of Self-rated Ability in the Accuracy of Confidence Judgements in Eyewitness Memory and General Knowledge. *Applied Cognitive Psychology*,*18*, 157-168.

Perfect, T.J., & Hollins, T.S. (1999). Feeling-of-Knowing Judgements Do Not Predict Subsequent Recognition Performance for Eyewitness Memory. *Journal of Experimental Psychology: Applied*, *5*(3), 250-264.

Perfect, T., & Hollins, T.S. (1996). Predictive Feeling of Knowing Judgements and Postdictive Confidence Judgements in Eyewitness Memory and General Knowledge. *Applied Cognitive Psychology, 10*, 371-382.

Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology, 78,* 144-147.

Quillian, M.R. (1967). Word Concepts: A Theory and Simulation of Some Basic Semantic Capabilities. *Computers in Behavioural Science, 12*, 410-430.

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 365-377.

Rapcsak, S.Z., Reminger, S.L., Glisky, E.L., Kaszniak, A.W., & Comer, J.F. (1999). Neuropsychological Mechanisms of False Facial Recognition Following Frontal Lobe Damage. *Cognitive Neuropsychology, 16*, 267-292.

Reason, J.T. & Lucas, D. (1984). Using cognitive diaries to investigate naturally occurring memory blocks. In J.E. Harris & P.E. Morris (Eds.), *Everyday memory, actions and absentmindedness* (pp. 53–69). San Diego, CA: Academic Press.

Reder, L.M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19, 90-138.

Reder, L.M. (1988). Strategic control of retrieval strategies. In Bower G. H. (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 227-259). San Diego, CA: Academic Press.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 435-451.

Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? *Cognition, 47*(1), 25-57.

Ruiz-Soler, M., & Beltran, F.S. (2006). Face perception: An integrative review of the role of spatial frequencies. *Psychological Research, 70*, 273-292.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19,* 523-542.

Schmitter-Edgecombe, M., & Anderson, J.W. (2007). Feeling of Knowing in Episodic Memory Following Moderate to Severe Close-Head Injury. *Neuropsychology, 21*(2), 224-234.

Schnyer, D.M., Verfaellie, M., Alexander, M.P., LaFleche,G., Nicholls, L., & Kaszniak, A.W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: evidence from patients with lesions to frontal cortex. *Neuropsychologia, 42,* 957-966.

Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology, 17*, 36-71.

Schooler, J.W. (2002). Verbalization Produces a Transfer Inappropriate Processing Shift. *Applied Cognitive Psychology*, 16 ,989-997.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition Learning, 4,* 33-45.

Schreiber, T.A. (1998). Effects of target set size on feelings of knowing and cued recall: Implications for the cue effectiveness and partial-retrieval hypotheses. *Memory & Cognition, 26*(3), 553-571.

Schwartz, B.L. (2006). The Feeling of Going: Judgments of Learning (JOLs) for Maps and Directions. *The Korean Journal of Thinking & Problem Solving, 16*(2), 5-16.

Schwartz, B.L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1*(3), 357-375.

Schwartz, B.L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments. *Memory & Cognition, 36*(1), 9-19.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1074-1083.

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93-113). Cambridge, MA: MIT Press.

Schwaninger, A., Wallraven., & Bulthoff, H.H. (2004). Computational Modeling of Face Recognition Based on Psychophysical Experiments. *Swiss Journal of Psychology, 63*(3), 207-215.

Searcy, J.H., & Bartlett, J.C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 904-915.

Semenza, C., & Zettin, M. (1989). Evidence from aphasia for the role of proper names as pure referring expressions. *Nature, 342*(6250), 678-679.

Semmler, C., & Brewer, N. (2006). Postidentification Feedback Effects on Face Recognition Confidence: Evidence for Metacognitive Influences. *Applied Cognitive Psychology*, *20*, 895-916.

Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning Memory and Cognition*, *12*(3), 452–460.

Shimamura, A. P., & Squire, L. R. (1988). Long-term memory in amnesia: Cued recall, recognition memory, and confidence ratings. *Journal of Experimental Psychology: Learning Memory and Cognition*, *14*(4), 763–770.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 204-221.

Sommer, W., Heinz, A., Leuthold, H., Matt, J., & Schweinberger, S. R. (1995). Metamemory, distinctiveness, and event-related potentials in recognition memory for faces. *Memory & Cognition*, *23*, 1-11.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, *3*, 315-316.

Sporer, S.L. (1991). Deep-Deeper-Deepest? Encoding Strategies and the recognition of human faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 323-333.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315-327.

Stanhope, N. & Cohen, G. (1993). Retrieval of proper names: Testing the models. *British Journal of Psychology, 84,* 51-65.

Tanaka, J.W., & Sengco, J.W. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*(5), 583-592.

Tauber, S.K., & Rhodes, M.G. (2010). Metacognitive errors contribute to the difficulty in remembering proper names. *Memory*, 18(*5*), 522-532.

Thiede, K.W. & Dunlosky, J. (1994). Delaying students' metacognitive monitoring improves their accuracy in predicting the recognition performance. *Journal of Educational Psychology, 86,* 290-302.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of selfregulated study: An analysis of selection of items for study and selfpaced learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 25*, 1024–1037.

Tulving, E. (1985). Memory and consciousness. *Canadian Journal of Psychology*, 26, 1-12.

Tulving, E. (1989). Memory: Performance, Knowledge, and Experience. *European Journal of Cognitive Psychology,1*(1), 3-26.

Tulving, E., & Schacter, D.L. (1990). Priming and Human Memory Systems. *Science, 247(*4940), 301-306.

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71-86.

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free-recall learning. *Journal of Experimental Psychology, 71*, 673-679.

Valentine, T. & Bruce, V. (1986). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40,* 300-305

Valentine, T. & Moore, V. (1995). Naming faces: The effects of facial distinctiveness and surname frequency. *Quarterly Journal of Experimental Psychology, 48*, 849-878.

Valentine, T. (1988). Upside down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*, 471-491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161-204.

Valentine, T. (2001). Face-space models of face recognition. In M. J. Wenger & J. T. Townsend (Eds.), *Computational geometric and process perspectives on facial cognition: Contexts and challenges* (pp. 83-113). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier Elimination. *The Quarterly Journal of Experimental Psychology*, *47*, 631-650.

Van Overschelde, J.P. (2008). Metacognition: Knowing about knowing. Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 47-71). New York, NY: Taylor and Francis.

Van Overschelde, J.P,, & Nelson, T.O. (2006). Delayed judgments of learning cause both a decrease in absolute accuracy (calibration) and an increase in relative accuracy (resolution). *Memory & Cognition*, *34*(7), 1527-1538.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20*(3), 291-302.

Watier, N.N., Collin, C.A., & Boutet, I. (2010). Spatial-frequency thresholds for featural and configural discriminations in upright and inverted faces. *Perception*, 39, 502-513.

Weber, N., & Brewer, N. (2004). Confidence–Accuracy Calibration in Absolute and Relative Face Recognition Judgments. *Journal of Experimental Psychology: Applied*, *10*(3), 156-172.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155-170). New York, NY: Cambridge University Press.

Widner, R.L., Smith, S.M., & Granziano, W.G. (1996). The effects of demand characteristics on the reporting of tip-of-the-tongue and feeling-of-knowing states. *The American Journal of Psychology*, *109*, 525-538.

Winograd, T. (1981). Elaboration and Distinctiveness in Memory for Faces. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *7*(3), 181-190.

Woodhead, M.M, & Baddeley, A.D. (1981). Individual differences and memory for faces, pictures, and words. *Memory & Cognition*, *9*(4), 368-370.

Yin, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1): 141-145.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517.

Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The faces that launched a thousand slips. *British Journal of Psychology, 76*,495-523.

Young, A.W., McWeeny, K.H., Ellis, A.W. & Hay, D.C. (1986). Naming and categorizing faces and names. *Quarterly Journal of Experimental Psychology 38A*, 297-318.