Metacognitive errors contribute to the difficulty in remembering proper names

Sarah K. Tauber and Matthew G. Rhodes
Colorado State University, Fort Collins, CO, USA

While prior research has shown that proper names are more challenging to learn and remember than other types of information (e.g., occupations), little research has explored the role of metacognitive factors in proper name learning. Thus in four experiments participants learned, made predictions, and were tested on their memory for common nouns (i.e., occupations) and proper nouns (i.e., names). Results showed that memory predictions were consistently overconfident for names, whereas the discrepancy between predictions and performance was smaller for occupations. With experience, participants were able to modify predictions and, critically, Experiment 4 showed that improvements in the accuracy of memory predictions led participants to allocate more study time to names and thus improved memory for names. Such data suggest that theories of proper name learning should make provisions for deficits in metacognitive awareness.

Keywords: Metacognition; Metamemory; Memory; Names.

A growing area of research examines memory for proper names. Such research has consistently demonstrated that proper names (e.g., Mr Smith) are more challenging to learn and remember than other types of information such as an occupation or hobby (e.g., Bredart, Valentine, Calder, & Gassi, 1995; Cohen & Faulkner, 1986; James, 2004; James, Fogler, & Tauber, 2008; McWeeny, Young, Hay, & Ellis, 1987; Peressotti, Cubelli, & Job, 2003). While deficits in proper name learning are well documented, comparatively little work has examined the metacognitive basis of proper name learning (but see Hosey, Peynircioglu, & Rabinovitz, 2009).

Metacognition refers to self-reflection about cognition (Koriat, 2007; Metcalfe, 2000; Nelson & Narens, 1994) and includes monitoring and control processes (Nelson & Narens, 1994). For example, a student preparing for an upcoming exam will be better prepared if she is able to accurately assess (monitor) what information has been learned. Such monitoring can then inform decisions related to the self-regulation of learning (i.e., metacognitive control), such as the decision to stop studying because information is known and understood (e.g., Metcalfe & Kornell, 2005; Nelson, 1996, Nelson & Leonesio; 1988; Thiede & Dunlosky, 1999). To the degree that monitoring is poor, corresponding control processes will not operate optimally.

Metacognition is commonly assessed via judgements of learning (JOLs) in which participants predict how likely they are to later remember some piece of information. These predictions are then compared with subsequent memory performance, allowing for an assessment of the correspondence between memory predictions and memory performance (termed absolute accuracy or calibration). For example, calibration would be excellent if participants predicted they would
remember 50% of the material and indeed they remembered 50%. However, metacognitive errors would be evident when predictions reliably differ from memory performance, exhibiting either overconfidence (i.e., greater predictions than performance) or underconfidence (i.e., lower predictions than performance). Memory predictions are often accurate, yet research has demonstrated a number of discrepancies between predictions and performance (e.g., Koriat & Bjork, 2005; Koriat, Bjork, Sheffer, & Bar, 2004; Rhodes & Castel, 2008, 2009). For example, Rhodes and Castel (2008) reported that participants provided higher JOLs for words presented in a large font compared with a smaller font, indicating that participants expected large words to be better remembered than small words. However, results showed no impact of font size on recall (cf. Rhodes & Castel, 2009). Thus participants made predictions using an accessible cue at encoding (font size) that was unrelated to memory performance (see also Kornell & Bjork, 2009). In the current study we were likewise concerned with whether participants would base JOLs on available cues such that memory predictions would distinguish between recall for names versus other information (i.e., occupations) that is generally more memorable.

Theories attempting to account for the difficulty of learning proper names emphasise the accessibility (activation) of proper name information compared with other information (see Cohen & Burke, 1993, for review). For example, Burton and Bruce (1992) posited an interaction and activation competitive network (IAC) in which information can be stored in three interactive units: face recognition units (FRU), person identity nodes (PIN), and semantic information units (SIU). From this perspective, proper names are difficult to retrieve because they are unique to each individual whereas other information (e.g., occupations) is more likely to be shared across individuals (see also Hanley, 1995). Thus proper name information is more difficult to retrieve because retrieval relies on few PIN units, whereas common noun information has the benefit of input from several PINs.

As an alternative to IAC, node structure theory (NST) accounts for the difficulty in learning proper names as well as word retrieval errors and age-related differences in memory for names (Burke, MacKay, & James, 2000). Common nouns (semantic information) and proper nouns are presumed to be stored as separate lexical nodes, with this division maintained even if a name and occupation are homonyms and have identical phonology (e.g., Mr Baker and a baker). The surname “Baker” can be connected with an individual’s first name (if known) by a semantic node. However, the occupation “baker” is connected to several semantic nodes relevant to that occupation (e.g., wears a white hat) as well as information about known bakers (e.g., I know a Mr Smith who is a baker.). Therefore semantic information is more easily activated than proper name information because semantic information has supplementary links to other related information.

Both the IAC model (Burton & Bruce, 1992) and NST model (Burke et al., 2000) can account for the ubiquitous finding that proper names are more difficult to remember than other types of information (e.g., Bredart et al., 1995; Cohen & Faulkner, 1986; James, 2004; James et al., 2008; McWeeny et al., 1986; Peressotti et al., 2003). Critically, these models of proper name learning do not account for a potential deficit in metacognition as a contributing factor to the difficulty of learning proper names. For example, if individuals do not understand that names are more difficult to remember than other types of information, they may not engage in control strategies that would improve name learning, such as allocating additional time to studying a name. Thus deficits in proper name learning may be exacerbated by a lack of metacognitive awareness of the difficulty of learning names, a factor not addressed by current theoretical accounts of proper name learning.

**OVERVIEW OF CURRENT RESEARCH**

In the experiments reported, participants studied and made JOLs for faces paired either with an occupation or name, and memory was then examined with a cued recall test. Given prior work indicating that names are less likely to be remembered than occupations (e.g., Bredart et al., 1995; Cohen & Faulkner, 1986; James, 2004; James et al., 2008; McWeeny et al., 1987; Peressotti et al., 2003), the first two experiments examined whether participants are aware of this effect. To anticipate, participants were not aware of the difficulty of learning names and provided equivalent memory predictions for names and occupations. Experiment 3 attempted to improve memory predictions by allowing participants to learn through task experience. Finally,
in Experiment 4 we allowed participants to exert control over learning via self-paced study.

EXPERIMENT 1

Part of what drives research in learning names in association with faces is the fact that some words, such as baker, exist in the English language as both proper names (surnames) and common nouns (occupations). McWeeny et al. (1987) were the first to formally examine this, reporting that proper names were harder to remember than occupations even when identical words were used (e.g., Mr Baker vs employment as a baker), with the only distinction being the status of a word as a name or occupation (a finding referred to as the Baker–baker paradox). Subsequent proper name learning studies have replicated these findings (e.g., James, 2004; James et al., 2008). In Experiment 1 we asked participants to study and make JOLs for words that exist as both surnames and occupations (e.g., baker, carpenter, butler, dean); these items will henceforth be referred to as Baker–baker items. Each Baker–baker item was paired with a face. Half of the faces were paired with an occupation (baker) and half with a surname (e.g., Mr Baker). If participants are sensitive to the greater difficulty of learning names compared with occupations then memory predictions should be lower for names that may either result in equivalent calibration for names and occupations or potentially better calibration for names.

Method

Participants. A total of 44 students from Colorado State University (M age = 18.73, SD = 0.97) participated for credit in psychology courses.

Materials. A total of 24 colour photographs of unfamiliar male (12 photos) and female (12 photos) faces taken from Minear and Park (2004) were used in the face–noun associated learning task. The ages of the individuals pictured ranged from approximately 18 to 25 years. Then 24 surnames that are also occupations (e.g., baker, dean, butler) were selected based on US Census Bureau (1990) frequencies for proper names that indicate the percentage of the population who possess each name (M name frequency = .05%, SD = .04%). Log transformed English lexicon frequencies were used to indicate how common each occupation is in the English language (Balota et al., 2002; M occupation frequency = 8.77, SD = 1.09).

Participants saw 24 faces in total, 12 female faces, 6 paired with surnames and 6 paired with occupations, and 12 male faces, 6 paired with surnames and 6 paired with occupations. Of the 24 pairs, the first and last 2 pairs served as primacy and recency buffers (1 from each gender and type of item) and were excluded from all analyses reported. Nouns were counterbalanced such that each word appeared equally often with a male or female face and as an occupation or surname.

Procedure. Participants were instructed that they would study a face paired with a name or occupation that they were to remember in preparation for a later memory test. The study phase then began with face–noun pairs presented one at a time at a 4-s rate in a fixed-randomised sequence. Immediately following each pair participants made a JOL predicting the likelihood of later recalling a surname or occupation if given the face as a cue on a scale from 0 (not likely at all) to 100 (very likely). Participants were encouraged to use the entire range of the scale and were given 4 s to write down their JOL. A 500-ms interstimulus interval was included before the presentation of the next study item after each JOL. Immediately following the study phase, participants engaged in a 3-minute filler task (solving maths problems) prior to proceeding to the test phase. During the test phase, each face was shown again for 4 s in a new fixed-randomised sequence and participants were instructed to write down each name or occupation that was remembered. Participants were given 10 s to answer and were directed to write down “Mr” or “Ms” if a name was remembered. Finally, participants were debriefed and thanked for participation.

Results

Only names that included “Mr” or “Ms” were scored as correct. Thus, if a participant reported the correct noun but as an occupation rather than as a name (e.g., correct answer was Mr Baker, but the participant only wrote down baker), this was scored as incorrect. Calibration
was of primary interest in the current research. As previously noted, calibration refers to the overall correspondence between JOLs and recall performance. To the extent that one is well calibrated, memory predictions will be consistent with memory performance.

Predicted and actual recall performance are depicted in Figure 1. These data were examined in a 2 (Item Type: surname, occupation) × 2 (Measure: JOL, recall) repeated-measures analysis of variance (ANOVA). The alpha level was set to .05 for this and all subsequent analyses reported. Main effects of Item Type, \( F(1, 43) = 8.79, \eta^2_p = .17 \), and Measure, \( F(1, 43) = 88.17, \eta^2_p = .67 \), were qualified by a reliable Item Type × Measure interaction, \( F(1, 43) = 4.72, \eta^2_p = .10 \). Follow-up tests revealed that while JOLs did not differ between names and occupations, \( t(43) = 1.60, p = .12, \) Cohen’s \( d = .13 \), participants recalled reliably fewer surnames than occupations, \( t(43) = 2.83, \) Cohen’s \( d = .86 \). Thus, while names were more difficult to recall than occupations (cf. McWeeny et al., 1987), participants’ JOLs were not sensitive to this discrepancy.

**EXPERIMENT 2**

Results from Experiment 1 indicated that participants’ JOLs were insensitive to the type of item (name or occupation) to be remembered resulting in poor calibration for names compared with occupations. In contrast, memory was significantly poorer for names than occupations, consistent with prior research (Baressi, Obler, & Goodglass, 1998; Bredart et al., 1995; Cohen & Faulkner, 1986; James, 2004; James et al., 2008; McWeeny et al., 1987; Peressotti et al., 2003). One potential limitation of this finding is that the overall level of memory performance, while not at floor, was quite low in Experiment 1. We thus attempted to replicate the results of Experiment 1 using a methodology that would enhance memory performance. Specifically, participants in Experiment 2 studied 20 rather than 24 face–noun pairs, and these pairs were shown twice during the study phase, yielding two JOLs for each face–noun pair. We expected to replicate the primary results of Experiment 1 such that JOLs should be similar for names and occupations while recall would be superior for occupations. In addition, while recall of names should remain lower than recall of occupations, exposing participants to each pair twice should elevate the overall level of recall compared with Experiment 1.

**Method**

**Participants.** A total of 40 students (\( M_{\text{age}} = 19.13, SD = 1.14 \)) participated for credit in Colorado State University psychology courses.

**Materials and procedure.** Materials were identical to those used in Experiment 1 with the exception that 20 rather than 24 pairs were presented in Experiment 2. (US Census Bureau frequencies for the proper names, 1990; \( M = .06, SD = .05 \); log transformed English lexicon frequencies for common nouns; Balota et al., 2002; \( M = 8.74, SD = 1.18 \)).

The procedure was identical to that of Experiment 1, with the exception that the entire 20-item face–noun list was presented twice, consecutively, in an identical order and without any breaks between presentations. This created two blocks in which participants studied and made a JOL for each face–noun pair. Participants were informed that face–noun pairs would be presented more than once.

**Results**

**JOLs.** Mean JOLs (Figure 2) were analysed in a 2 (Block: 1, 2) × 2 (Item Type: Name, Occupation) repeated-measures ANOVA. Results showed that the magnitude of JOLs increased from Block 1 (\( M = 40.45, SD = 12.99 \)) to Block 2, \( M = 52.63, SD = 16.95 \); \( F(1, 39) = 28.60, \eta^2_p = .43 \). However, JOLs did not differ for names compared with occupations, \( F < 1 \), nor did Block interact with Item Type, \( F < 1 \).
24.39, h = main effect of Measure was evident, ANOVA was conducted with Block 2 JOLs. A 2 (Measure: JOL, Recall) repeated-measures ANOVA was conducted. A Cohen’s d = 2.75, all recall performance was higher in Experiment 2 across both experiments. Most importantly, over-
to examine the percentage of targets recalled 1, 2) mixed-factor ANOVA was conducted
A second 2 (Item Type: Name, Occupation) \times 2 (Measure: JOL, Recall) repeated-measures ANOVA was conducted with Block 2 JOLs. The main effects of Measure, F(1, 39) = 3.96, p = .05, ηp² = .09, and Item Type, F(1, 39) = 4.23, p = .05, ηp² = .10, were marginally reliable. Consistent with Experiment 1, an Item Type \times Measure interaction was evident, F(1, 39) = 9.05, ηp² = .19. Follow-up tests demonstrated that while JOLs for names and occupations did not differ (t < 1), recall was reliably higher for occupations than for names, t(39) = 2.75, Cohen’s d = .43.

A second ANOVA was conducted with Block 2 JOLs. A main effect of Measure was evident, F(1, 39) = 24.39, ηp² = .39, while the main effect of Item Type was marginally reliable, F(1, 39) = 3.96, p = .05, ηp² = .09. The Item Type \times Measure interaction was reliable, F(1, 39) = 8.12, ηp² = .17, with follow-up tests demonstrating that JOLs for names and occupations did not differ for Block 2 (t < 1). The same cued recall performance was used for both ANOVAs because memory was measured once.

Recall. Experiment 2 was designed to improve memory performance in comparison with the low levels of recall observed in Experiment 1. A 2 (Item Type: Name, Occupation) \times 2 (Experiment: 1, 2) mixed-factor ANOVA was conducted to examine the percentage of targets recalled across both experiments. Most importantly, overall recall performance was higher in Experiment 2

\( M = 32.44, SD = 22.09 \) compared with Experiment 1 \( M = 18.26, SD = 10.76; F(1, 84) = 14.90, ηp² = .15 \). In addition, recall of occupations \( M = 29.33, SD = 22.61 \) reliably exceeded recall of names, \( M = 20.38, SD = 19.79; F(1, 84) = 15.00, ηp² = .15 \). Finally Item Type did not interact with Experiment, \( F < 1 \).

**Discussion**

The primary results of interest from Experiment 1 were replicated in Experiment 2. Specifically, while JOLs did not differ between names and occupations, participants were less likely to recall names compared with occupations. Thus participants’ JOLs were insensitive to the difficulty of learning and remembering proper names, leading to poor calibration for names while better calibration was evident for occupations. We note that this pattern obtained even when steps were taken to improve memory performance.

Participants in Experiments 1 and 2 were unable to anticipate that proper nouns are more challenging to remember than common nouns (i.e., occupations). A critical question remains: Can participants’ predictions become sensitive to the difficulty of learning names compared with occupations? We attempted to examine this in Experiment 3 by conferring participants the opportunity to adjust memory predictions in light of a prior experience with name learning.

**EXPERIMENT 3**

Experiment 3 allowed participants the chance to learn from prior experience to determine if they can become aware of the difficulty of learning names. Specifically, participants were given two sets of study–test trials (with novel materials at each study opportunity) that allowed them to adjust predictions across trials. In contrast to Experiment 2, where the same information was studied twice and tested once, participants in Experiment 3 studied and were tested on one set of information and then again with a new set of materials, permitting them the opportunity to learn from a prior study/test experience.

We expected to replicate prior differences in calibration at Time 1 (i.e., the first set of study–test trials) such that JOLs would not differ between names and occupations but that participants would exhibit better recall for occupations.
than names. However, a different pattern of results was expected for Time 2 (i.e., the second set of study–test trials). We predicted that participants would be able to learn from their experience of Time 1 and adjust memory predictions accordingly to reflect differences in recall between names and occupations. Thus, at Time 2, participants’ predictions were expected to be well calibrated for both names and occupations.

Method

Participants. A total of 40 students (\(M\) age = 18.90, \(SD = 1.24\)) from Colorado State University participated for credit in psychology courses.

Materials and procedure. Materials from Experiment 2 were used, along with an additional set of 20 stimuli drawn from the same sources as the previous experiments. Nouns were counterbalanced such that each word appeared equally often in the Time 1 study–test trial and the Time 2 study–test trial and also as female surnames, male surnames, female occupations, and male occupations. The procedure was identical to that used in Experiment 1 with the exception that, immediately after finishing the test phase for first study–test cycle, participants studied, made JOLs and were tested on a new set of 20 face–noun pairs at the Time 2 study–test cycle.

Results

Our a priori interest was in the impact of prior experience on JOLs. Accordingly, we examined Time 1 and Time 2 calibration separately.

Calibration Time 1. Predicted and actual recall performance at Time 1 is depicted in the left panel of Figure 3. These data were analysed in a 2 (Item Type: Name, Occupation) \(\times\) 2 (Measure: JOL, Recall) repeated-measures ANOVA. Reliable main effects of Measure, \(F(1, 39) = 64.38, \eta^2_p = .62\), and Item Type, \(F(1, 39) = 7.50, \eta^2_p = .16\), were qualified by a reliable Measure \(\times\) Item Type interaction, \(F(1, 39) = 4.53, \eta^2_p = .10\). Follow-up tests demonstrated that JOLs for names and occupations did not differ, \(t(39) = 2.84, Cohen’s d = .51\). Thus, while recall differed, predictions did not differ at Time 1.

Calibration Time 2. Predicted and actual recall performance at Time 2 is depicted in the right panel of Figure 3. These data were examined in a 2 (Item Type: Name, Occupation) \(\times\) 2 (Measure: JOL, Recall) repeated-measures ANOVA. The main effects of Measure, \(F(1, 39) = 26.95, \eta^2_p = .41\), and Item Type were reliable, \(F(1, 39) = 17.45, p < .001, \eta^2_p = .31\). However, unlike the pattern of data apparent for Time 1, Measure did not interact with Item Type, \(F(1, 39) = 1.67, p = .20, \eta^2_p = .04\), at Time 2. Follow-up tests demonstrated that JOLs for names were reliably lower than JOLs for occupations, \(t(39) = 3.52, Cohen’s d = 1.13\). Additionally, recall was significantly lower for names compared with occupations, \(t(39) = 3.60, Cohen’s d = 1.15\).

Discussion

Results of Experiment 3 at Time 1 paralleled those of the previous experiments. Specifically, JOLs did not differ between names and occupations, whereas memory for occupations was better than memory for proper names. Critically, at Time 2 participants provided lower JOLs for
names than occupations. Thus Experiment 3 indicates that participants are able to learn from experience to adjust memory predictions for proper and common nouns. However, while participants’ JOLs for names were better calibrated at Time 2, this occurred without a concomitant improvement in memory for names from Time 1 ($M = 17.38$, $SD = 17.84$) to Time 2 ($M = 18.38$, $SD = 18.36$), $t < 1$.

Experiment 3 suggested that participants were able to learn that names are more difficult to remember than occupations without any impact on memory performance. Based on these data alone, one might conclude that a lack of awareness of the difficulty of name learning has little or no bearing on memory performance. However, an important assumption of research in metacognition is that monitoring influences control processes related to the self-regulation of learning (e.g., Nelson, 1996; Nelson & Narens, 1994). From this perspective, an awareness that names are difficult to learn (monitoring) may serve to influence subsequent encoding of names (control processes). Experiment 4 was designed to examine this issue.

**EXPERIMENT 4**

Participants in Experiment 4 were permitted to self-pace study for each face–noun pair. In addition, as in Experiment 3, participants were given two separate sets of study–test trials, with novel face–noun pairs for each trial. As in prior experiments, it was expected that JOLs would not differ between names and occupations at Time 1 while recall of occupations would be superior to recall of names. Our particular interest was in study time allocation for names compared with occupations across trials as well as any potential changes in memory performance. Optimally, participants should devote more study time to those items that are least likely to be remembered. For example, the discrepancy-reduction model (e.g., Thiede & Dunlosky, 1999) suggests that participants seek to reduce the discrepancy between the desired goal (e.g., to learn the given information) and the current state of learning (e.g., how much of the information is learned) in order to perform optimally (but see Ariel, Dunlosky, & Bailey, 2009; Metcalfe & Kornell, 2005, for alternative perspectives). Thus names and occupations are regarded as equally memorable, and participants should allocate roughly equivalent amounts of study to names and occupations at Time 1.

We anticipated a different pattern of data at Time 2. In particular, results from Experiment 3 demonstrated that participants provided lower JOLs for names at Time 2 compared with Time 1 indicating that, with experience, they began to appreciate the difficulty of remembering names. If such findings are replicated in Experiment 4, participants should seek to allocate more study time to names compared with occupations at Time 2. One potential outcome of a change in study time allocation is that memory for names may improve from Time 1 to Time 2.

To summarise, participants in Experiment 4 studied and were tested on two different sets of face–noun pairs. Study time was self-paced, permitting participants to allocate as much study time as they deemed necessary to each item. If participants learn from a prior experience that names are more difficult to remember than occupations, they may devote more study time to names than occupations. Such a change in study time allocation may potentially improve participants’ recall of names.

**Method**

**Participants.** A total of 48 students ($M$ age = 18.73, $SD = 1.18$) from Colorado State University participated for credit in psychology courses.

**Materials and procedure.** The materials and procedure were identical to Experiment 3 with one key exception. Specifically, participants in Experiment 4 were allowed to self-pace study with instructions to study each pair for as long as they deemed necessary (but no longer), such that they would be able to remember the name or occupation when given only the face. In order to measure self-paced study time participants pressed the space bar to advance to the JOL screen and then reported their JOL aloud to an experimenter. All other aspects of the procedure were identical to Experiment 3.

**Results**

**Self-paced study.** Mean self-paced study time data are reported in Table 1 and were examined

\footnote{Participants reported JOLs aloud in order to eliminate switching between using the keyboard and writing.}
in a 2 (Item Type: Name, Occupation) \times 2 (Time: 1, 2) repeated-measures ANOVA.\(^4\) Main effects of Item Type, \(F(1, 47) = 10.29, \eta^2_p = .18\), and Time, \(F(1, 47) = 10.09, \eta^2_p = .18\), were qualified by a reliable Item Type \times Time interaction, \(F(1, 47) = 8.45, \eta^2_p = .15\). Follow-up tests demonstrated that while self-paced study at Time 1 did not differ between names and occupations \((t < 1)\), participants allocated significantly more time for names than occupations at Time 2, \(t(47) = 4.05, \text{Cohen’s } d = 1.18\). Thus, with a prior experience, participants sought to allocate more study time to names than occupations. It should also be noted that participants spent reliably more time studying names at Time 2 compared with Time 1, \(t(47) = 3.45, \text{Cohen’s } d = 1.01\).

Similar to Experiment 3, our primary interest was in the impact of experience on memory predictions. Accordingly, we examined Time 1 and Time 2 calibration separately.\(^5\)

**Calibration Time 1.** Predicted and actual recall performance at Time 1 is depicted in the left panel of Figure 4. These data were analysed in a 2 (Item Type: Name, Occupation) \times 2 (Measure: JOL, Recall) repeated-measures ANOVA. Main effects of Measure, \(F(1, 47) = 49.45, \eta^2_p = .51\), and Item Type were evident, \(F(1, 47) = 18.00, \eta^2_p = .28\), and qualified by a reliable Measure \times Item Type interaction, \(F(1, 47) = 11.50, \eta^2_p = .20\).

Follow-up tests demonstrated that while JOLs for names and occupations did not differ, \(t(47) = 1.17, p = .25\), Cohen’s \(d = .34\), recall of names was significantly lower than recall of occupations, \(t(47) = 4.39, \text{Cohen’s } d = 1.28\).

**Calibration Time 2.** Predicted and actual recall performance at Time 2 is depicted in the right panel of Figure 4. These data were analysed in a 2 (Item Type: Name, Occupation) \times 2 (Measure: JOL, Recall) repeated-measures ANOVA. Main effects of Measure, \(F(1, 47) = 4.29, \eta^2_p = .08\), and Item Type, \(F(1, 47) = 15.53, \eta^2_p = .25\), were qualified by a reliable Measure \times Item Type interaction, \(F(1, 47) = 5.48, \eta^2_p = .10\). Follow-up tests demonstrated that JOLs for names were reliably lower than JOLs for occupations, \(t(47) = 2.36, \text{Cohen’s } d = .69\). Additionally, recall was significantly lower for names compared with occupations, \(t(47) = 3.48, \text{Cohen’s } d = 1.02\).

**Name recall.** A primary question of interest for Experiment 4 was to determine whether permitting control over study time allocation would improve recall. While memory improved for both names and occupations, we focus primarily on proper names as proper name learning is of foremost interest in this paper. Specifically, Time 2 name recall reliably exceeded Time 1 name recall, \(t(47) = 3.21, \text{Cohen’s } d = .94\). Thus permitting participants to self-pace study time led to an improvement in proper name learning, a finding that may inform attempts to improve memory for proper names.\(^6\)

**Discussion**

Experiment 4 produced several notable findings. First, calibration at Time 1 for Experiment 4 replicated prior experiments such that predictions of memory performance did not vary between

---

\(^4\) Analyses were additionally conducted using median self-paced study time.

\(^5\) Omnibus analyses of Item Type (Name, Occupation), Measure (JOL, Recall), and Time (1, 2) revealed main effects of Item Type, \(F(1, 47) = 29.83, \eta^2_p = .39\), and Measure, \(F(1, 47) = 27.44, \eta^2_p = .37\), but not Time \((F < 1)\). These findings were qualified by several reliable interactions. Time interacted with Measure, \(F(1, 47) = 33.73, \eta^2_p = .42\), such that Time 1 JOLs were greater than Time 2 JOLs, \(t(47) = 4.30, \text{Cohen’s } d = 1.25\), while Time 2 recall exceeded Time 1 recall, \(t(47) = 3.46, \text{Cohen’s } d = 1.01\). In addition, Item Type interacted with Measure, \(F(1, 47) = 14.67, \eta^2_p = .24\), such that occupation JOLs were marginally greater than name JOLs, \(t(47) = 1.96, p = .06\), Cohen’s \(d = .57\), while occupation recall was reliably greater than name recall, \(t(47) = 5.53, \text{Cohen’s } d = 1.61\). No other interactions were supported.

---

\(^6\) Recall of names in Experiments 3 and 4 was additionally analysed with an omnibus analysis comparing performance at Time 1 and Time 2. There was a main effect of Time, \(F(1, 86) = 5.61, \eta^2_p = .06\), such that name recall was greater at Time 2 than Time 1. There was also a main effect of Experiment, \(F(1, 86) = 7.14, \eta^2_p = .08\), such that name recall was greater in Experiment 4 compared to Experiment 3. Importantly, a marginal Time \times Experiment interaction was present, \(F(1, 86) = 3.60, p = .06, \eta^2_p = .04\). Specifically, name recall at Time 1 did not differ between Experiment 3 \((M = 17.38, SD = 17.84)\) and Experiment 4 \((M = 23.75, SD = 20.47); t(86) = 1.54, p = .13\). However, Time 2 name recall was greater for Experiment 4 \((M = 32.79, SD = 24.60)\) compared with Experiment 3 \((M = 18.38, SD = 18.36); t(86) = 3.06, \text{Cohen’s } d = .66\).
names and occupations while occupations were better remembered than names. Experiment 4 also demonstrated that poor calibration at Time 1 influenced control decisions. Specifically, participants spent equivalent amounts of time studying names and occupations at Time 1. Thus monitoring impairments led to errors in study choices (cf. Rhodes & Castel, 2009), which may perpetuate proper name learning difficulty.

At Time 2 calibration was improved such that JOLs became more diagnostic of memory performance for both names and occupations. Importantly, the improvement in calibration led to changes in the allocation of study time such that, at Time 2, participants spent more time studying names than occupations. In turn, this change in study-time allocation improved memory for names, a finding that has implications for theories of proper name learning. We examine these and other issues in the General Discussion.

**GENERAL DISCUSSION**

The current set of experiments is the first, to our knowledge, to systematically examine participants' metacognitive awareness of name learning vis-à-vis predictions of memory performance for proper names compared with occupations. Experiment 1 demonstrated that while names were recalled less frequently than occupations, memory predictions did not differ between names and occupations. Experiment 2 replicated this effect (i.e., no difference in predictions for names compared with occupations) under conditions that improved memory performance. Thus Experiments 1 and 2 demonstrated that memory predictions were insensitive to the difficulty of learning proper names.

Experiment 3 provided two study-test trials in order to determine whether participants can learn from experience to improve metacognitive judgements. Indeed, results demonstrated that JOLs for names became better calibrated after a study–test opportunity. Finally, participants in Experiment 4 were permitted to self-pace study across two study–test trials in an effort to provide control over learning. Results showed that when participants were unaware of the difficulty of learning names compared with occupations they allocated the same amount of study time to both types of nouns (i.e., Time 1). However, with experience (i.e., Time 2), participants were able to improve calibration and consequently allocated more study time to names than occupations. Crucially, this change in study-time allocation also improved memory for names. We note that it was not a prior study-test experience alone that led to the improvement in memory for names. For example, participants in Experiment 3 were provided with two study-test trials with names and occupations and exhibited no improvement in memory for names from Time 1 to Time 2. Rather, an improvement in memory for names was only evident when participants were permitted to self-pace learning (i.e. Experiment 4).

Data from the current experiments are consistent with prior metacognitive research demonstrating discrepancies between memory performance and metamemory awareness (e.g., Koriat & Bjork, 2005; Koriat et al., 2004; Rhodes & Castel, 2008, 2009; Tauber & Rhodes, 2010). One potential explanation for such discrepancies is provided by Koriat's (1997) cue-utilisation framework. According to this framework, memory predictions are inferential and can be based on three types of information: intrinsic, extrinsic, or mnemonic cues. Intrinsic cues include information about the specific items to-be-remembered (e.g., item difficulty, or proper versus common nouns), extrinsic cues include information about the study experience (e.g., retention interval), and mnemonic cues are internal indices of learning. Mnemonic cues are typically informed by practice or experience and can thus modify existing personal theories of metacognition. From this perspective, participants will provide memory predictions that match memory performance to the extent that they are sensitive to the diagnostic cues provided during study.
In contrast, a direct access approach (e.g., Arbuckle & Cuddy, 1969; Dunlosky & Nelson, 1994; Hart, 1965) suggests that memory predictions are based on the strength of a memory episode such that strong memory traces elicit high predictions, while weak memory traces elicit lower predictions. From this perspective, participants’ memory predictions should reflect later memory performance because learning a proper name is more challenging than learning a common noun, resulting in a weaker memory trace and lower memory predictions. Our data suggest that while noun type (i.e., proper vs common) directly influences memory performance, participants only became sensitive to this influence with prior experience. A direct access account suggests that predictions should have been sensitive to later memory performance if participants are able to directly access the strength of items during encoding, contrary to the current data. Thus our data lend support for inferential accounts of metacognition (e.g., Koriat, 1997).

The present study replicates the oft-reported finding that proper names are more challenging to learn and remember than common nouns (e.g., Baressi et al., 1998; Bredart et al., 1995; Cohen & Faulkner, 1986; James, 2004; James et al., 2008; McWeeny et al., 1987; Peressotti et al., 2003). However, theories of proper name learning (e.g., IAC: Burton & Bruce, 1992; Valentine, Brennen, & Bredart, 1996; NST: Burke et al., 2000) do not make specific provisions for metacognitive deficits in proper name learning. The current data, particularly in Experiment 4, indicate that improvements in calibration influence choices made during study and thus influence memory for names. Such data suggest that theories of proper name learning should posit some role for monitoring and control processes. From a more applied perspective, a great deal of research has demonstrated that older adults experience more difficulty learning proper names (e.g., Baressi et al., 1998; James, 2004; James et al., 2008) than young adults. Additional research should focus on metacognitive training as a potential tool to improve proper name learning, particularly for older adults. At the same time, older adults frequently cite memory for proper names as one of their most dominant concerns (Cohen & Faulkner, 1984). Given this, it may be that older adults would provide more accurate metacognitive judgements and, as such, metacognitive training may benefit younger adults more than older adults.

REFERENCES


