Journal of Experimental Psychology: Learning, Memory, and Cognition

Action Memory and Metamemory

Neil W. Mulligan, Jonathan A. Susser, and Daniel J. Horschler

Online First Publication, May 12, 2022. http://dx.doi.org/10.1037/xlm0001136

CITATION
Mulligan, N. W., Susser, J. A., & Horschler, D. J. (2022, May 12). Action Memory and Metamemory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. http://dx.doi.org/10.1037/xlm0001136

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001136

Action Memory and Metamemory

Neil W. Mulligan, Jonathan A. Susser, and Daniel J. Horschler
Department of Psychology & Neuroscience, University of North Carolina at Chapel Hill

Actions can enhance memory, exemplified by the enactment effect. In a typical experiment, participants hear a series of simple action phrases (e.g., *bounce the ball*), which they either carry out (subject-performed tasks, or SPTs), watch the experimenter carry out (experimenter-performed tasks, EPTs), or simply listen to (verbal tasks, VTs). Later memory is usually better for SPTs than for either EPTs or VTs. Although research on action memory is extensive, research on action and metamemory is sparse and produces contradictory results. Furthermore, the metamemory literature has largely ignored the effects of action. Some theoretical perspectives argue that actions produce a particularly effective and automatic form of encoding, and that such nonstrategic encoding should produce inaccurate memory predictions. Other theories argue that action memory relies on executive control processes, suggesting that memory predictions for actions should be just as good (or better) than for control conditions. In Experiments 1a and 1b, participants predicted (with judgements-of-learning, JOLs) whether they would later remember SPTs and EPTs. Resolution (the correlation between JOLs and later recall) was greater for EPTs than SPTs, and not significantly different than zero in the latter case. Experiment 3 produced the same results with SPTs and VTs: resolution was greater for VTs and not significant for SPTs. The results are consistent with nonstrategic accounts of the enactment effect, and also highlight the importance of examining metamemory for actions given that actions can alter metamemory relative to verbal (VT) and other non-action (EPT) conditions. In addition, the presence of JOLs attenuates the enactment effect, a reactive effect of JOLs similar to that found with other encoding effects.

Keywords: action memory, metamemory, the enactment effect, the SPT effect, JOL reactivity

The study of memory has traditionally been dominated by verbal materials, even though everyday memory often entails retrieval of actions we have taken or observed. To remedy this imbalance, researchers in the 1980s initiated studies of action memory (e.g., Cohen, 1981; Engelkamp & Krumnacker, 1980), a line of inquiry which has generally found that actions (and gestures) enhance memory (Engelkamp, 1998; Ianì & Bucciarelli, 2017; Mulligan, 2014). Perhaps the most heavily investigated phenomenon in this domain is the enactment effect. In a typical experiment, participants hear a series of simple action phrases (e.g., *bounce the ball*), which they either carry out (subject-performed tasks, or SPTs), watch the experimenter carry out (experimenter-performed tasks, EPTs), or simply listen to (verbal tasks, VTs). Memory is typically better for SPTs than either of the other two conditions—the enactment effect (also referred to as the SPT effect)—indicating that carrying out the action enhanced memory (Engelkamp, 1998; Mulligan, 2014; Roediger & Zaromb, 2010).

A traditional and persisting view of the enactment effect is that carrying out an action elicits a particularly efficient and nonstrategic (or automatic) form of encoding (e.g., Wojcik et al., 2011), a view proposed early in its strongest form by Cohen (1981, 1983, 1985, 1989). The proposal that the enactment effect is driven by automatic forms of encoding was predicated on early results indicating that different populations (older vs. younger adults; children vs. adults; and mentally-retarded vs. control subjects) exhibited similar memory for SPTs but differed in memory for verbal materials (e.g., Bäckman & Nilsson, 1984; Cohen & Bean, 1983; Cohen & Stewart, 1982; Lind & Bowler, 2009). The usually worse memory in these populations (e.g., older adults; children; individuals with mental retardation) is often attributed to differential use of controlled, strategic encoding processes, making the equality in memory for SPTs consistent with the nonstrategic encoding thesis. Likewise, a number of studies indicated that, relative to verbal materials, SPTs are less affected by traditional encoding manipulations such as levels of processing, generation, and item elaboration (Cohen & Bryant, 1991; Helstrup, 1987; Nilsson & Cohen, 1988; Nilsson & Craik, 1990; Zimmer & Engelkamp, 1999). The notion is that the encoding of verbal material has room to improve when guided by better encoding strategies, but that actions by default produce robust encoding and so are not aided (at least as much) by strategies or controlled processes that can be fruitfully applied to verbal materials. Although some of these early results have been

---

Neil W. Mulligan https://orcid.org/0000-0002-8307-0594

Daniel J. Horschler https://orcid.org/0000-0003-0440-8217

Daniel J. Horschler is now at the Department of Psychology, Yale University.

Data and materials: https://osf.io/zs8du/?view_only=cc96d316f9b64d1db7c1604305a012a1.

Correspondence concerning this article should be addressed to Neil W. Mulligan, Department of Psychology & Neuroscience, University of North Carolina at Chapel Hill, CB #3270, Chapel Hill, NC 27599-3270, United States. Email: nmulligan@unc.edu

reevaluated, as discussed next, the idea that actions are more automatically encoded has persisted and found support in recent research (e.g., Ianì & Bucciarelli, 2017; Sahakyan & Foster, 2009; Schatz et al., 2011; Wang et al., 2021; Wojcik et al., 2011; Zhang & Zuber, 2020).

Is action memory actually unique? With regard to memory performance, not in the strong sense envisioned by Cohen (1981, 1983), perhaps. Some of the early population dissociations have been reevaluated as in the case of aging, which initially appeared to show preserved memory for SPTs in older age (Bäckman & Nilsson, 1984), but with subsequent research, using refined methods and greater power, finding equivalent age effects for VTs and SPTs (Feyereisen, 2009; Rönnlund et al., 2003; see Foley & Ratner, 2001 for similar results with children). With regard to encoding factors, the results are mixed, with some encoding factors demonstrating differences between actions and verbal materials and others exhibiting similarity. For example, an early study indicated that memory for SPTs was less affected by presentation time than verbal materials (Cohen, 1985), interpreted as SPTs exhibiting little effect of controlled rehearsal processes. Later research using more appropriate control conditions found that increased presentation time enhanced later memory for SPTs, VTs, and EPTs to a similar extent (Peterson & Mulligan, 2015). In contrast, the primacy effect, typically attributed to controlled rehearsal processes, is often found to be robust with VTs and EPTs, but smaller (or nonexistent) with SPTs (e.g., Cohen, 1981; Schatz et al., 2011; Seiler & Engelkamp, 2003). Consistent with this are the encoding manipulations mentioned earlier (level of processing, generation, item elaboration) that have larger effects on verbal materials than on action memory. Finally, dividing attention during encoding is more disruptive to memory for EPTs and VTs than for SPTs (Bäckman et al., 1991, 1993; Wang et al., 2021; cf. Engelkamp & Zimmer, 1996). Thus, reviews of action memory (e.g., Mulligan, 2014; Roediger & Zaromb, 2010) conclude that, despite some similarities in memory for actions and other materials, memory for SPTs is less reliant on strategic and controlled rehearsal processes than is memory for VTs or EPTs.

With respect to metamemory, these issues are quite open. Despite substantial research and increasing clarity on the ways in which memory for performed actions is similar and dissimilar to memory for observed actions and verbal material, research on metamemory for actions is unclear and sparse. There is a vast amount of research on metamemory using a wide variety of verbal materials (word lists, word pairs, sentences, text, etc..) but only a few studies on action memory, and these studies have had little impact on the broader metamemory literature. Indeed, several major reviews of metamemory research do not mention action memory or the studies reviewed next. This is true of older reviews (Metcalfe & Shimamura, 1994; Nelson, 1990) as well as more recent treatments (Dunlosky & Tauber, 2016). This is a potentially important omission given claims that metamemory for actions differ from metamemory for verbal and other types of material.

Cohen (1983; Cohen et al., 1991) argued that because actions are substantially encoded nonstrategically, we have little insight into their variable encoding whereas verbal material, more subject to controlled and strategic rehearsal, give rise to at least some awareness of their differential memorability. Consequently, predictions about memorability for VTs (and EPTs) should be at least somewhat accurate whereas predictions for SPTs should exhibit less resolution. According to this view, metamemory for actions and verbal materials should produce important differences with regard to monitoring.

There have only been a handful of studies on metamemory for actions, most by Cohen (1983, 1988; Cohen & Bryant, 1991; Cohen et al., 1991), which indicate that participants have little ability to predict which SPTs will later be recalled but generally have at least some ability to predict later recall for verbal materials. However, this research was largely conducted prior to the development of current standards for evaluating metamemorial monitoring (e.g., judgments-of-learning [JOLs], resolution, calibration; Rhodes, 2016), and the methods used produce some limitations in interpreting the results. Furthermore, direct comparisons between the SPT and control conditions were not always reported.

In a typical study (Cohen, 1988; Experiment 1), one group of participants was presented with a list of individual words and another group was presented with a list of action phrases for enactment (the SPT condition). A prediction about memorability was made for each word or action on a 4-point scale: 1 *(this word/task is too weak in my memory for me to recall it)* to 4 *(this word/task is strong enough in my memory so I will be able to recall it)*. A free recall test was then given for the words or actions. To assess the accuracy of the predictions (tantamount to a measure of relative accuracy, or resolution), the average rating for items recalled minus the average rating for items not recalled was computed for each participant; values greater than zero indicate at least some degree of resolution whereas values of zero (or lower) indicate no predictive accuracy. This number was significantly greater than zero for the word condition but not for the SPT conditions. However, this measure of resolution was not statistically compared between the word and SPT conditions, so it is unclear if resolution was significantly greater for words than actions. Cohen et al. (1991) likewise presented one group with a list of words and another with a list of SPTs, each followed by a memory prediction on the 4-point scale, and again, resolution was greater than 0 in the word condition but not in the SPT condition. But again, resolution was not statistically compared across the two conditions. Cohen & Bryant (1991) examined metamemory for SPTs from study trials with either short (5 s) or long duration (30 s), and found that neither duration produced significant resolution. However, this experiment did not have a verbal or EPT control condition for comparison.

Only one of Cohen's studies had a VT or EPT control condition, Experiment 2 of Cohen (1983). In this experiment, participants were presented with a study list of either auditory words, visual words, VTs, EPTs or SPTs, with each item followed by a recall/no-recall prediction. To assess resolution, the proportion of items actually recalled was computed separately for the two prediction categories, and a ratio of the two values computed, such that values above 1 indicated greater recall for items predicted to be recalled. This proportion significantly varied across the list conditions, indicating that resolution varied with the materials. However, there were no direct statistical comparisons between list conditions. The numerical value of this measure was highest for the word lists, next for VTs, then EPTs, and lowest for SPTs, but the lack of statistical comparison and the proximity of the values for the VT, EPT, and SPT conditions renders the results inconclusive.

One final study (McDonald-Miszczak et al., 1996) contrasted a VT and SPT condition with older and younger adults across three study-test blocks. In each block, the participants studied the same list of VT or SPT items, making a memory prediction (on a 7-point scale) after each item, followed by a free recall test. This experiment assessed resolution by computing the now commonly used gamma correlation between ratings and recall (e.g., Nelson, 1984), finding that the gamma correlation averaged over blocks did not significantly differ between the VT and SPT conditions, and was significantly above zero in both cases. Thus, in contrast to Cohen's results, the SPT condition produced significant resolution that did not differ from a verbal control condition.

The handful of studies on metamemory for actions produce conflicting implications, with the studies by Cohen suggesting that metamemory, particularly the accuracy of memory predictions, is particularly poor for actions, and the results of McDonald-Miszczak et al. (1996), suggesting that resolution might be equally good for actions and verbal materials. However, there are several limitations to these studies that render the question open. First is the issue of the best control condition. One of the Cohen studies (Cohen & Bryant, 1991) had no nonaction control condition. Two other Cohen studies (Cohen, 1988; Cohen et al., 1991) used lists of individual words as the comparison condition. This condition differs in several ways from the SPT condition in addition to the presence of actions: the nature of the list items (individual words vs. action phrases); the identity of the materials (the words in the word lists and the action phrases were not the same); the list lengths; and so forth Any of the differences could drive a difference in metamemory performance. A better control condition uses the same action phrases as the SPT condition—the traditional VT condition, as used in McDonald-Miszczak et al. (1996). An even better control condition uses the same action phrases and has the participant observe rather than carry out the action—the EPT condition. This control condition better isolates the action itself (that is the motoric component of enactment; Engelkamp, 1998). For example, for the action phrase *bounce the ball*, the VT condition controls for the verbal information experienced but the EPT condition also controls for nonaction perceptual information that is experienced in the SPT condition, such as the visual appearance of the object and its motion, and the sound of the ball as it bounces. This renders a better control for isolating the effects of enactment, per se. Only Cohen (1983; Experiment 2) included an EPT control condition, along with SPT, VT and word-list conditions, but did not report contrasts between individual conditions.

Second, all of the foregoing studies used between-subjects or pure-list designs, in which a study list was composed entirely of a single type of item (e.g., a pure list of SPT items, a pure list of VT items). This design contrasts with a mixed-list design in which the various item types are randomly intermixed in the study list. The enactment effect is one of several encoding effects that are moderated by the type of design, such that an enactment effect is typically found for mixed-list designs on a later free recall test whereas pure-list designs produce a reduced, null, or even negative enactment effect (e.g., Engelkamp & Dehn, 2000; Peterson & Mulligan, 2010, 2015; Steffens, 1999; Steffens et al., 2015; see McDaniel & Bugg, 2008; Mulligan & Lozito, 2004, for reviews). Consistent with this, Cohen (1983) found no difference in recall for SPTs and EPTs, and McDonald-Miszczak et al. (1996) likewise found no recall difference for SPTs and VTs. The pure-list design used in these prior studies may have minimized the usual enactment effect in memory. The current experiments use a mixed-list design with free recall as the memory test, to increase the likelihood that an enactment effect will be found in memory.[1]

Third, the McDonald-Miszczak et al. (1996) study used three study-test blocks to assess memory. In such a design, the JOLs made after the first block are typically influenced by the participant's memory for performance on the prior test (the memory-for-prior-test [MPT] heuristic, Finn & Metcalfe, 2007, 2008). That is, after the first block, JOLs may be largely based on memory for prior test performance rather than on the ability of the subject to discern differences in ongoing encoding (or reencoding). This is problematic in the present case because the question of interest focuses on whether actions permit insight into differentially effective encoding. The introduction of nonencoding bases for JOLs reduces our ability to answer this question and might explain why McDonald-Miszczak et al. (1996) found no difference in resolution between their SPT and VT conditions: the two conditions might both be largely based on nonencoding factors (e.g., the MPT heuristic) especially on the later blocks.

Finally, the studies by Cohen and colleagues did not assess metamemory in ways that have since become standard, such as using 0–100 JOL predictions and measuring resolution with the gamma correlation between JOLs and recall performance, both of which would facilitate comparisons with research in the current metamemory literature. The use of the typical JOL measure would also provide the secondary benefit of permitting assessment of absolute accuracy, or calibration (the difference between predicted [JOLs] and actual memory performance).

The Current Experiments

The current study examines metamemorial monitoring by contrasting SPTs and EPTs (Experiments 1a, 1b, and 2), and SPTs and VTs (Experiment 3). The main issue relates to resolution—the extent to which JOLs are predictive of actual memory performance. The view of Cohen (1983, 1988, for example) was that actions produce a particularly efficient and nonstrategic form of encoding and thus provide little insight into the differentially effective encoding, but that other forms of information are more subject to strategic, controlled rehearsal processes that provide at least some insight into differential encoding. This view is echoed in more recent research, which emphasizes that nonstrategic encoding plays a greater role for actions than EPTs or VTs (e.g., Sahakyan & Foster, 2009; Schatz et al., 2011; Wang et al., 2021; Wojcik et al., 2011; Zhang & Zuber, 2020). Likewise, the motor-encoding view of Engelkamp, Zimmer and colleagues argues for the relatively automatic and effective encoding of motor information, proposed as the basis of the enactment effect (e.g., Engelkamp, 1998; Seiler & Engelkamp, 2003; Zimmer & Engelkamp, 1999). These views imply that predictions about action memory should have less relation to actual memory than predictions for other materials.

---

[1] Although there is also the possibility that making memory predictions on an item-by-item basis may also reduce or eliminate the enactment effect, an issue discussed later.

Another view of the enactment effect is the episodic-integration account (Feyereisen, 2009; Hainselin et al., 2014; Kormi-Nouri, 1995; Kormi-Nouri & Nilsson, 2001), an account that disputes the centrality of motor information in the enactment effect. This account instead argues that the enactment effect is due to enhanced episodic binding among the various constituents of the enacted event, especially among the verbal-semantic attributes of the enacted phrase. This binding process has been characterized as an executive control process (e.g., De Lucia et al., 2019). Thus, in contrast to other theories about the enactment effect, the episodic-integration account attributes the enactment advantage to strategic, controlled processes, which in turn implies that the SPT condition should exhibit substantial resolution, at least as much resolution as control conditions (EPT or VT conditions) that engage the central (controlled) binding process either less effectively or to a lesser degree.

The primary focus is on resolution but as secondary matters, the experiments also allow the assessment of (a) average JOLs and (b) absolute accuracy, or calibration. Average JOLs determine the extent to which participants predict an enactment effect in memory. As for calibration, the current experiments determine if enactment enhances or detracts from this aspect of metamemorial accuracy. Calibration has not been clearly assessed in the prior research.

A final issue relates to the enactment effect in memory. All of the present experiments used a mixed-list design because pure-list designs (as used in the prior research) typically reduce or eliminate the enactment effect in free recall. A mixed-list design maximizes the opportunity to observe the enactment effect, and the current experiments are based on materials and procedures that have demonstrated higher recall for the SPT than EPT or VT conditions (Hornstein & Mulligan, 2004; Mulligan & Hornstein, 2003; Peterson & Mulligan, 2015). However, the presence of item-by-item JOLs complicates the straightforward prediction of an enactment effect because JOLs can influence memory performance (i.e., the reactivity of JOLs; Double et al., 2018). In particular, the presence of JOLs can sometimes reduce or eliminate otherwise robust memory effects. For example, the usual effect of perceptual interference on memory was eliminated when participants made item-by-item JOLs during encoding (Besken & Mulligan, 2013). The same result has been reported with the generation effect (Begg et al., 1991; Matvey et al., 2001). Similarly, JOLs attenuate, if not eliminate, the levels-of-processing effect (Tekin & Roediger, 2020). Of course, JOLs do not always reduce memory effects (e.g., relatedness effects in cued recall actually increase with JOLs; Janes et al., 2018). It is unknown if JOLs attenuate the enactment effect but if so, it may be necessary to perform a control experiment without JOLs to verify that the usual enactment effect occurs under the current conditions.

Experiments 1a and 1b

Experiment 1a began the investigation with a comparison of SPT and EPT conditions. Experiment 1b was conducted as an exact replication to assess the replicability of Experiment 1a and to clarify one aspect of the results of that experiment (see Footnote 3).

Method

Participants

Thirty-two undergraduates from UNC at Chapel Hill participated in Experiment 1a in exchange for partial course credit in an introductory psychology course.[2] A separate sample of 32 UNC undergraduates participated in Experiment 1b. Otherwise, the experiments were identical. All experiments received approval from the UNC IRB.

Design and Materials

Encoding condition (SPT vs. EPT) was manipulated within subjects. Thirty-six action phrases were drawn from those used in Peterson & Mulligan (2015) and Hornstein & Mulligan (2004). Each phrase had a unique verb and object that was not duplicated in any other action phrase. Four of the phrases served as primacy (two) and recency (two) buffers on the study lists and were not scored on the final test. The remaining 32 phrases were the critical items. These phrases were randomly divided into Sets A and B, and used to create two study lists, one in which Set A was assigned to the SPT condition and Set B to the EPT condition, and another list with the assignments reversed. The phrases from the two sets were randomly intermixed in the study lists subject to the constraint that no more than two items in the same condition appeared in sequence. The primacy and recency buffers consisted of one SPT and one EPT item each.

Procedure

At the beginning of the experiment, participants were informed that they would hear a list of action phrases, half of which they would perform and the other half they would see the experimenter perform. Participants were also informed that they would be asked to predict whether they would remember each action on the later test. The participant and the experimenter sat on opposite sides of a table facing each other. Part of the table was screened from the participant's view, behind which were the objects required for the actions. Each study trial consisted of the following. First, an action phrase was presented over computer speakers. Next, the experimenter removed the relevant object from behind one of the screens and placed it on the table. The experimenter then said either "Enact" or "Observe" depending on whether the item was in the SPT or EPT condition. For SPT trials, the participant carried out the action. For EPT trials, the participant watched as the experimenter carried out the action. Each study trial lasted 6 seconds, which pilot testing showed was sufficient time to carry out the actions. After the trial was complete, the experimenter placed the object back behind the screen, out of the participant's view. The participant then provided a JOL, indicating how confident he or she was about being able to recall the action phrase on the later test on a scale of 0 (*Not confident at all*) to 100 (*Extremely*

---

[2] Experiment 1a was conducted several years prior to submission of this article, before a priori power calculations were routinely reported. The sample size was chosen to be similar to sample sizes used in our previous experiments in which the SPT and EPT conditions were compared within-subjects (Hornstein & Mulligan, 2004; Mulligan & Hornstein, 2003; Peterson & Mulligan, 2015). Post-hoc power computations were based on size of the enactment effect in these experiments, which averaged $d_z = .73$. The power to detect an enactment effect of this size is .99 for $n = 32$.

confident). The participant spoke the rating aloud and the experimenter entered it into the computer. The JOL rating was self-paced.

After the study phase, the participant was given a 3-minute distractor task consisting of arithmetic problems. Following this, the recall test was administered. The participant was given a blank sheet of paper and asked to write down as many of the action phrases as could be recalled. The participant was asked to try to recall the entire action phrase when possible but to recall any parts of the phrase if it could not be recalled in its entirety. The test lasted 5 minutes.

The data and materials for all experiments are available here: https://osf.io/zs8du/?view_only=cc96d316f9b64d1db7c1604305a012a1.

Results

JOLs and Recall Performance

The results from both experiments are presented in Table 1. During the study phase, JOLs for SPTs were significantly greater than those for EPTs in both Experiment 1a, $t(31) = 3.69$, $p = .001$, $d_Z = .65$, and Experiment 1b, $t(31) = 3.05$, $p = .005$, $d_Z = .54$. Performance on the recall test was measured in two ways, according to both a lenient and strict criterion. For the lenient scoring, the item was scored as correct if either the action or the object of the action phrase was recalled. For the strict scoring, both the action and the object of the action phrase needed to be recalled together. The final score was computed as the percent of SPT or EPT items recalled. The lenient and strict scoring produced the same results (in this and the following experiments) with one exception noted below. Consequently, only the strict results are reported. The recall scores did not significantly differ between SPTs and EPTs for either Experiment 1a, $t(31) = 1.23$, $p = .23$, or Experiment 1b, $t(31) = .87$, $p = .39$. Thus, no significant enactment effect was found in recall.

Resolution

To assess the relationship between JOLs and recall, gamma correlations were computed for each subject separately for SPTs and EPTs. Gamma cannot be computed if either the JOLs or the recall status is the same for all items in a condition. In Experiment 1a, 31 participants had a valid gammas for the SPT condition, 31 for the EPT condition, and 30 subjects had a valid gamma for both. In Experiment 1b, 31 subjects had a valid gammas for the SPT condition, 32 for the EPT condition, and 31 subjects had a valid gamma for both. Thus, the *df*s vary slightly in the subsequent analyses.

Table 1
Results of Experiments 1a and 1b: Means (SD)

	Experiment 1a		Experiment 1b	
Measure	SPT	EPT	SPT	EPT
JOL	79.7 (16.0)	76.4 (18.6)	75.3 (20.7)	70.3 (20.1)
Recall	32.0 (11.7)	30.7 (14.0)	37.5 (11.4)	35.2 (13.2)
Gamma	−.07 (0.48)	.20 (0.53)	.02 (0.50)	.26 (0.45)
Calibration	47.6 (20.6)	45.8 (24.2)	37.8 (23.2)	35.2 (24.1)

Note. JOL = judgment of learning; SPT = subject-performed task; EPT = experimenter-performed task.

The gamma correlations were significantly larger for EPTs than SPTs for both Experiment 1a, $t(29) = 2.75$, $p = .010$, $d_Z = .50$, and for Experiment 1b, $t(30) = 2.13$, $p = .041$, $d_Z = .38$.[3] In addition, gamma correlations were significantly greater than zero for EPTs (Experiment 1a: $t(30) = 2.22$, $p = .034$, $d_Z = .40$; Experiment 1b: $t(31) = 3.45$, $p = .002$, $d_Z = .61$) but not for SPTs, (Experiment 1a: $t(30) = -.80$, $p = .43$; Experiment 1b: $t(30) = .17$, $p = .87$).

Calibration

To assess calibration, the difference between the average JOL and proportion recalled was computed separately for the SPT and EPT condition for each participant. Scores nearer zero indicate better calibration. For the strict scoring, calibration was not significantly different for the SPT and EPT conditions in Experiment 1a, $t(31) = .14$, $p = .89$, or in Experiment 1b, $t(31) = .90$, $p = .37$. Participants were markedly overconfident (calibration scores were substantially positive) but the encoding conditions did not differ in calibration.[4]

Combined Analysis of Experiments 1a and 1b

For more powerful analyses, the data from Experiments 1a and 1b were combined and reanalyzed with experiment as a factor. For recall, the SPT and EPT conditions did not significantly differ, $F(1, 62) = .86$, $p = .34$. The gamma correlations were significantly greater in the EPT than SPT condition, $F(1, 59) = 8.15$, $MS_e = .140$, $p = .006$, $\eta^2_p = .164$. For all analyses, there was no indication of an interaction between encoding condition and experiment (all $F\text{s} < 1$).

Discussion

The primary results focus on resolution. Gamma correlations were higher for EPTs than SPTs, and gamma correlations were significantly greater than zero for EPTs but not for SPTs. This was found in Experiment 1a and in the replication Experiment 1b. Furthermore, in the combined analysis, this consistency was demonstrated by the lack of interaction with experiment—indicating that the results were qualitatively and quantitatively replicated.

Next are two issues of secondary importance. First, participants' average JOLs indicate an expectation that SPTs would be more memorable than EPTs. This extends the results of McDonald-Misczcak et al. (1996) who found higher JOLs for SPTs than VTs. This also appears consist with the results of Cohen's (1983) prediction ratings, which appear somewhat higher for SPTs than EPTs or VTs, although no statistical comparisons were reported. Ironically, in the present case as in both of these prior studies, the enactment effect was not actually observed on the recall test. This

---

[3] The one difference between the lenient and strict recall scores occurred for the gamma correlations in Experiment 1a. When gamma was computed between JOLs and the lenient scores, the correlations were not significantly greater for EPTs ($M = 0.18$, $SD = 0.43$) than SPTs ($M = 0.03$, $SD = 0.42$), $t(29) = 1.43$, $p = .16$. For Experiment 1b, however, gammas with the lenient scores were significantly greater for EPTs ($M = 0.20$, $SD = 0.46$) than SPTs ($M = -0.04$, $SD = 0.33$), $t(30) = 2.72$, $p = .011$, $d_Z = .49$, consistent with the gammas for strict scores.

[4] The lenient recall scores are necessarily higher than the strict scores, but relative to these scores, participants' JOLs were likewise substantially overconfident (in this and the subsequent Experiment 3).

may mean that the JOL ratings were (at least partly) based on a general belief about the efficacy of enactment encoding rather than an accurate projection about the actual subsequent memory performance under the current conditions. Second, calibration was equal across conditions, showing that absolute accuracy of JOLs was similar for SPTs and EPTs.

Finally, the recall scores did not exhibit an enactment effect. At the outset, it was clear that the use of a mixed-list design should make the enactment effect more likely but that the addition of item-by-item JOLs had the possibility of attenuating the effect, as has been found with other encoding effects (Begg et al., 1991; Besken & Mulligan, 2013; Matvey et al., 2001; Tekin & Roediger, 2020). In the present case, the use of JOLs may have rendered the usual enactment effect nonsignificant. In order to determine if this is actually the case, Experiment 2 repeated the current experiment without JOLs to verify that the enactment effect would otherwise occur under the present conditions.

Experiment 2

Method

Participants

Twenty-two undergraduates from UNC at Chapel Hill participated in exchange for partial course credit in an introductory psychology course. A power analysis was based on our previous experiments comparing SPT and EPT conditions within-subjects (Hornstein & Mulligan, 2004; Mulligan & Hornstein, 2003; Peterson & Mulligan, 2015), which yielded an average enactment effect on recall of $d_z = .73$. The sample size required to detect this effect with power of .90 ($\alpha = .05$, two-tailed) is 22.

Design, Materials, and Procedure

Experiment 2 was identical to Experiments 1a and 1b with the exception that the JOLs were removed from the study trials.

Results and Discussion

As expected, the enactment effect was found: recall was significantly greater for SPTs (M = 46.0, SD = 19.7) than EPTs (M = 31.8, SD = 20.8), $t(21) = 2.70$, $p = .013$, $d_Z = .58$.

A cross-experiment comparison shows that the significant enactment effect in Experiment 2 is indeed significantly larger than the nonsignificant effect in the prior experiments. The recall data from Experiments 1a and 1b were combined and compared to the recall data from Experiment 2. The analysis revealed a significant interaction between encoding condition and experiment (1a and 1b vs. 2), $F(1, 84) = 7.32$, $MS_e = .017$, $p = .008$, $\eta^2_p = .080$. The interaction indicates that the enactment effect is significantly larger in Experiment 2 than in the combined data from Experiments 1a and 1b, consistent with the notion that the item-by-item JOLs attenuated (and largely eliminated) an otherwise robust enactment effect.

The results are clear: when the item-by-item JOLs were removed, a robust enactment effect emerged in the recall test. This indicates that the typical enactment advantage is found with the present materials and procedures, and that the presence of JOLs in Experiments 1a and 1b dampened the effect. JOLs have been shown in past research to attenuate otherwise robust encoding effects. It appears that the enactment effect (at least as measured by the SPT–EPT difference) can likewise be reduced by JOLs.

Experiment 3

SPTs produce worse resolution than EPTs, implying that actions impair the relative accuracy of metacognition. Experiment 3 compares SPTs with VTs to further examine this issue. Although there are good arguments that EPTs are the better control condition for examining the effects of enactment on memory (and metamemory; e.g., Engelkamp, 1998), there are several reasons for extending the analysis to the VT control condition. First, numerous studies of action memory make use of VTs as the control condition (e.g., Ma et al., 2021; McDonald-Miszczak et al., 1996; Sahakyan & Foster, 2009; Smith et al., 2010; Zhang & Wang, 2020; Zhang & Zuber, 2020; Zimmer & Engelkamp, 1999). To increase the applicability of the current analysis to the broader literature on the enactment effect, it is important to contrast SPTs with both of the commonly used control conditions. Second, using a VT control condition more directly gets at Cohen's original claims that metamemory for actions differs from metamemory for verbal materials. Third, the experiment provides an additional opportunity to see if the poor resolution of SPTs replicates. Finally, the issue of JOL reactance may be less critical with VTs as the control condition because the SPT–VT difference in memory is more robust and consistent than the SPT–EPT difference (e.g., De Lucia et al., 2019; Feyereisen, 2009). Thus, even if JOLs dampen the enactment effect, it is possible that the SPT-VT difference would still be detectable.

Method

Participants

Sixty undergraduates from UNC at Chapel Hill participated in exchange for course credit. The effect size of enactment on resolution in Experiments 1a and 1b averaged $d_z = .43$. The sample size required to detect this effect with power of .90 ($\alpha = .05$, two-tailed) is 59, increased to 60 to satisfy counterbalancing requirements.

Design, Materials, and Procedure

Encoding condition (SPT vs. VT) was manipulated within subjects. The materials and procedure were the same as Experiments 1a and 1b with the exception that the EPT study trials were replaced with VT trials in which the participant heard the action phrase but did not see the experimenter carry out the action. Before the study phase, participants were informed that they would hear action phrases and that in some cases they would be asked to enact the phrase and in other cases to listen to the phrase without enacting. The SPT trials proceeded as in Experiments 1a and 1b. The VT trials started with the presentation of the action phrase over the computer's speakers. Next, the experimenter retrieved the relevant object from behind the screen and placed it on the table. The experimenter then instructed the participant that this was a Listen trial. Participants then had 6 seconds to encode the phrase in the presence of the object. Next, the experimenter placed the object back behind the screen, out of the participant's view. The participant then gave a JOL, and the next trial began.

Results

JOLs and Recall Performance

The results are in Table 2. During the study phase, JOLs for SPTs were significantly greater than those for VTs, $t(59) = 5.25$, $p < .001$, $d_Z = .68$. Recall was also greater for SPTs than VTs, $t(59) = 4.48$, $p < 001$, $d_Z = .58$, demonstrating the enactment effect (as measured by the SPT–VT difference).

Resolution

The gamma correlation could not be computed for one participant in the VT condition. The gamma correlations were significantly greater for VTs than SPTs, $t(58) = 3.01$, $p = .004$, $d_Z = .39$, and were significantly greater than zero in the VT condition $t(58) = 4.55$, $p < .001$, $d_Z = .59$, but not in the SPT condition, $t(59) = .26$, $p = .80$.

Calibration

Calibration scores were computed separately for the SPT and VT condition for each participant. Calibration was not significantly different in the SPT and VT conditions, $t(59) = 1.61$, $p = .11$. As in the Experiments 1a and 1b, participants were substantially overconfident but the encoding conditions did not significantly differ in calibration.

Discussion

The most important result is that resolution is greater for VTs than SPTs, consistent with the results contrasting SPTs and EPTs. Regardless of whether the control condition is VT or EPT, actions appear to disrupt resolution. And like Experiments 1a and 1b, there is no evidence of above-chance resolution for SPTs, further indicating that enactment eliminates the (limited) metacognitive insight participants otherwise might have into the encoding of the action phrases.

In contrast, an important difference with Experiments 1a and 1b is that a significant enactment effect was found in recall. The enactment effect is often larger and more consistent when VTs are the control condition (e.g., De Lucia et al., 2019; Feyereisen, 2009). So it may be the case that the more robust SPT–VT difference is still detectable even when item-by-item JOLs are used whereas the SPT–EPT difference may be more vulnerable to this attenuation. Another possibility is that the SPT–VT difference is simply unaffected by the presence of JOLs during encoding. A direct comparison of the effect with and without JOLs would be necessary to determine which is the case. For present purposes, however, this is secondary; primary is the simple finding of the SPT–VT difference in recall, which demonstrates that the effect can be found with the current methods, and precludes any necessity of demonstrating the effect in an experiment without JOLs.

With regard to average JOLs and calibration, the results are consistent with Experiments 1a and 1b. The average JOLs show that participants expect to remember SPTs better than VTs, just as they expected better recall of SPTs than EPTs. The calibration results are likewise consistent. Participants generate the same overconfidence as in the earlier experiments, and this overconfidence is similar for SPTs and VTs.

Table 2
Results of Experiment 3: Means (SD)

Measure	SPT	VT
JOL	73.3 (18.8)	67.1 (19.5)
Recall	41.5 (13.6)	31.8 (11.7)
Gamma	−.01 (0.43)	.22 (0.46)
Calibration	31.7 (21.9)	35.3 (21.1)

Note. JOL = judgment of learning; SPT = subject-performed task; VT = verbal task.

General Discussion

The present experiments contrasted SPTs with EPTs and VTs to determine if action affects metamemory. Some conceptions of the enactment effect have implications for metamemory, which have not been adequately investigated, and research on metamemory has largely ignored action memory. The primary results show that SPTs produce worse resolution than EPTs and VTs, with gamma correlations that are insignificant in the former condition. This indicates that participants have little insight into the quality of memory encoding for actions whereas they have at least some insight into encoding for comparable verbal and observed events. Despite the difference in relative accuracy, the degree of absolute accuracy (calibration) is roughly equivalent for SPTs and EPTs, and for SPTs and VTs. Finally, the average JOL was consistently higher for SPTs than for EPTs or VTs.

In early studies of action memory, Cohen (1983, 1989; Cohen et al., 1991) proposed that action encoding is relatively nonstrategic or automatic, a view echoed in recent research (e.g., Ianì & Bucciarelli, 2017; Sahakyan & Foster, 2009; Schatz et al., 2011; Wang et al., 2021; Wojcik et al., 2011; Zhang & Zuber, 2020). This view implies that we should have less insight into the quality of encoding for SPTs compared to VTs and EPTs, which are more subject to controlled and strategic encoding processes. The motoric view of Engelkamp and colleagues (e.g., Engelkamp, 1998; Seiler & Engelkamp, 2003) likewise implies that the operative motoric representations are a relatively automatic consequence of enactment, and thus dovetails with broader nonstrategic accounts of the enactment effect. The primary results of the current experiments are consistent with this view: participants had no measurable ability to predict which SPTs will be recalled, and had higher (if still modest) ability to predict recall of EPTs and VTs.

In contrast, the current results appear inconsistent with the episodic-integration account (Feyereisen, 2009; Kormi-Nouri, 1995; Kormi-Nouri & Nilsson, 2001), which argues that the enactment effect results from enhanced episodic binding among the constituents of the enacted event, especially among the verbal-semantic attributes of the enacted phrase. The locus of the enactment advantage in such executive control processes (e.g., De Lucia et al., 2019) should render the encoding open to metacognitive insight and, at least to some degree, predictive accuracy. If the SPT condition heightens the involvement of an executive control process, then enactment should offer the same or higher degree of monitoring ability as the EPT or VT conditions that engage these control processes less effectively or fully. The current results are contrary to this view—there is no evidence that SPTs produce equivalent or higher levels of resolution compared to EPTs and VTs.

It might be asked whether the results compel the conclusion that SPT encoding is relatively nonstrategic (relative, that is, to VT and EPT encoding). Not in the sense that all possible contrary accounts have been ruled out, only in the limited sense that of the extant accounts of the enactment effect, the nonstrategic view fits the results and the episodic-integration account appears not to. A successful alternative to the nonstrategic view would have to propose that the enactment effect is due to greater strategic encoding of SPTs, and then account for why participants neglect this information when making JOLs for SPTs but not when making JOLs for VTs or EPTs. To our knowledge, no account with these characteristics has been proposed.

We have assessed the prediction first articulated by Cohen (1983; Cohen et al., 1991) that action-based encoding produces limited insight into the memorability of the event whereas the other types of encoding produce greater, if still limited, insight. This prediction predated much of the current research on metamemorial monitoring, focusing primarily on the use of JOLs (Rhodes, 2016). Furthermore, this prediction makes assumptions about the nature of memory predictions that are important to assess. In particular, the view articulated by Cohen assumes that memory predictions are sensitive to variation in controlled, strategic rehearsal processes, and that an encoding condition dominated by nonstrategic processing will be missing this basis of accurate prediction. At the outset, it should be noted that this is a reasonable assumption under the traditional view that we have awareness of and control over strategic processes, which should provide at least some insight into their effectiveness, whereas nonstrategic processes are not subject (or less subject) to awareness and control, and variation thereof gives rise to limited insight.

Despite the plausibility of this assumption, is it supported by research on JOLs? Of course, current research on metamemory has often shown that JOLs are affected by sources of information (cues) that are not related to actual memory performance. For example, various manipulations of fluent processing impact JOLs but have no (or opposite) effects on actual memory performance (e.g., Benjamin et al., 1998; Besken & Mulligan, 2013; Susser & Mulligan, 2015). But JOLs are also affected by information that is diagnostic of later recall, such as relatedness in paired-associative learning (e.g., see Rhodes, 2016; for review). Most relevant for current purposes is evidence that controlled and strategic encoding processes contribute to JOLs and enhance their accuracy (i.e., resolution). For example, Hertzog et al. (2010) showed that JOLs are sensitive to relatedness and also to strategy use such that more effective strategy use led to both higher JOLs and higher subsequent recall, resulting in enhanced resolution. Bröder and Undorf (2019) likewise found that encoding strategies can impact JOLs and recall. Friedman and Castel (2011) examined JOLs for Remember (R) and Forget (F) items in item-method directed forgetting. The usual interpretation of item-method directed forgetting is that people choose to rehearse R items and not F items, producing an R advantage in later recall that is driven by controlled, strategic encoding processes (e.g., Sahakyan & Foster, 2009). In Friedman and Castel's (2011) study, JOLs were higher for R than F items, indicating that participants were aware of the difference in controlled rehearsal and used it to make JOLs. The same results were obtained when a recall-value manipulation was used in which some items were assigned positive values during encoding (e.g., +5, indicating that these items should be recalled later) or negative values (e.g., −5, indicating that these items should not be recalled later, and should instead be forgotten). In addition, resolution was generally greater than zero for R (or positive value) items but not for F (or negative value) items. This is consistent with the notion that F items are not rehearsed (in the controlled sense), so variation in the effectiveness of their encoding is difficult to track, but R items are subject to controlled rehearsal, so variation in their encoding has at least some impact on JOLs and resolution (see Ariel & Dunlosky, 2011, for related evidence).

In sum, the hypothesis tested in the present experiments relies on the assumption that controlled, strategic rehearsal has at least some impact on JOLs and contributes to resolution, an assumption that is theoretically reasonable and appears supported by metamemory research. Of course, even in the EPT and VT conditions, the gamma correlations were still quite modest, indicating that JOLs were mostly driven by cues that are not diagnostic of recall and only modestly reflect diagnostic factors. While in the SPT condition, the JOLs appear wholly dominated by nondiagnostic information.

In the current experiments, SPTs produced little resolution. However, consideration of metamemory research indicates that there may be circumstances under which SPTs will show significant, and perhaps substantial resolution, particularly if additional, nonencoding information is brought to bear on JOLs. When participants experience multiple study-test blocks with the same material, resolution typically improves over blocks (Rhodes, 2016). This is likely because on later study phases, participants may think back about what was recalled earlier, giving higher JOLs to information that was previously recalled and lower JOLs for items that were not recalled—the MPT heuristic (Finn & Metcalfe, 2007, 2008). Over multiple blocks, new information about recall performance can be brought to bear in later study phases that was not available during the initial phase—that is, a new source of nonencoding information can now impact JOLs. Presumably, memory for prior recall could impact JOLs for SPTs as well as for other forms of information. If so, then resolution should increase across blocks for SPTs as well as for VT or EPT conditions, potentially attaining high levels of resolution.

This possibility is consistent with the results of McDonald-Miszczak et al. (1996) as well as a new study (Kubik et al., 2022) published after the present research was complete. Like McDonald-Miszczak et al. (1996), Kubik et al. examined metamemory for SPTs and VTs over three study-test blocks.[5] Both studies found that resolution for both SPTs and VTs increased over study-test blocks, with SPTs (as well as VTs) reaching relatively high levels of resolution (e.g., average gammas above .60 in Kubik et al., 2022). Furthermore, Kubik et al. reported higher gamma correlations for the VT than SPT condition, consistent with the results of Experiment 3. It should be noted, however, that this conflicts with the results of McDonald-Miszczak et al. (1996), who reported no significant differences in gamma correlations between conditions. However, the gamma correlations in McDonald-Miszczak et al. were numerically lower for SPTs than VTs in six of six experimental cells, so perhaps the differences between the studies are

---

[5] Kubik et al. (2022) also examined the effects of retrieval practice, an issue not relevant for current purposes.

not as stark as they might seem. Regardless, follow-up research on this issue should be done with EPTs as the control condition to better isolate the effect of action in the SPT condition.

Turning to the enactment effect in memory, Experiments 2 and 3 produced the expected effect. This is important given that earlier experiments on action memory and metamemory used pure-list designs. These designs often do not produce an enactment effect with free recall even with the typical experimental procedure, which does not use JOLs. Indeed, these prior studies failed to demonstrate an enactment effect in memory (Cohen, 1983; McDonald-Miszczak et al., 1996).

Even so, the current results indicate that the enactment effect (at least as measured by the SPT–EPT difference) is attenuated by the presence of JOLs. This is pertinent in two ways. First, it indicates that even with the use of a mixed-list design, it may not be possible to simultaneously observe an enactment effect in memory (that is, the SPT–EPT difference), and measure JOLs. However, for this to compromise the results of Experiments 1a and 1b, the presence of JOLs would have to not only eliminate the enactment effect in memory, but also induce a spurious difference in resolution by either selectively increasing resolution for EPTs or selectively reducing resolution for SPTs. There is no reason to suspect that is the case, and Experiment 3 argues against this in showing that the SPT–VT difference in resolution co-occurs with the usual SPT advantage in memory performance. Second, the lack of an enactment effect in Experiments 1a and 1b constitutes another example of the reactivity of JOLs. Recent research on this issue often focuses on whether JOLs increase or decrease memory performance relative to a no-JOL condition (e.g., Double et al., 2018; Myers et al., 2020). But another manifestation of reactance is when JOLs change the size or nature of a memory effect. This has been documented with the effect of perceptual interference, generation, and levels of processing, all of which are decreased by JOLs (Begg et al., 1991; Besken & Mulligan, 2013; Matvey et al., 2001; Tekin & Roediger, 2020). The present results indicate that the enactment effect joins this list.

Because our primary focus is not on JOL reactivity, we do not provide a complete analysis of this issue here. However, this is an important issue for future research because the reduction (or elimination) of the enactment effect by JOLs may be contrary to the nonstrategic view. That is, if the SPT advantage is rooted in nonstrategic encoding, one might wonder why the introduction of JOLs should reduce or eliminate that advantage. There are possibilities consistent with the nonstrategic view—perhaps JOLs bring recall levels in the EPT condition up to the level of SPT recall—but other possibilities seem less consistent—for instance, that JOLs reduce SPT recall. Given that we did not experimentally manipulate the presence of JOLs within a single experiment, our data do not provide this level of detail. It should be noted, however, that cross-experiment comparison of Experiments 1a and 1b with Experiment 2 imply that JOLs reduce SPT recall more than increasing EPT recall, a picture that appears inconsistent with the nonstrategic view. A clearer answer to this issue will require a direct, experimental analysis.

The present experiments focused primarily on resolution and did not directly assess theories of metamemory. However, the results regarding average JOLs merit some discussion. We consistently found that JOLs were higher for SPTs than for EPTs or VTs. A dominant framework for interpreting monitoring and JOLs is the cue-utilization account (Koriat, 1997; see Rhodes, 2016). This framework differentiates among three types of cues that may influence JOLs: intrinsic, extrinsic, and mnemonic cues. Intrinsic cues are properties of the to-be-remembered items, including the perceptual or semantic characteristics of the stimulus. Extrinsic cues refer to the conditions of learning, including such factors as the processes carried out on the stimulus, the relations among stimuli (e.g., serial position), and the characteristic of retrieval (e.g., the retention interval, nature of retrieval cues). Finally, mnemonic cues refer to the experiences of the learner that may be thought to predict later recall, such as the ease or fluency of processing a stimulus, or the memory for having previously recalled a stimulus. With regard to the intrinsic–extrinsic distinction, the cue-utilization account argues that intrinsic cues are typically heavily weighted in making JOLs and that extrinsic cues are relatively neglected. In the present case, the presence or absence of action may be thought of as an intrinsic property of the item because it is salient and clearly discerned. In terms of the cue-utilization account, this is one reason to expect the enactment manipulation to affect JOLs.

The enactment manipulation might also evoke differences in the mnemonic experience of the learner. Taking action feels quite different than not taking action in a number of ways (e.g., motorically, tactually, in the need for motor planning). There are also differences in terms of the degree of self-involvement in the SPT versus control conditions (e.g., Feyereisen, 2009; Kormi-Nouri & Nilsson, 2001) that may elicit a difference in the mnemonic experience. This indicates, that in terms of the cue-utilization account, it is an open question as to whether the intrinsic differences between actions and nonactions or the experiential differences (or both) drive differences in average JOL. Regardless, it should be noted that the SPT advantage with respect to JOLs is robust, happening for within-subject designs as used in the present experiments, and in between-subjects designs used in earlier research (McDonald-Miszczak et al., 1996). This is notable because some variables that have robust effects on JOLs in within-subject designs do not have an effect when the conditions are manipulated between subjects (e.g., Susser et al., 2013).

Another important avenue for subsequent research is the relationship between actions and metamemorial control. Research on metamemory distinguishes between monitoring (the evaluation of states of knowledge) and control (using these evaluations to guide learning behaviors). As an aside, this distinction should be kept separate from the issue of strategic (or controlled) versus automatic encoding processes that often guide theorizing about memory as opposed to metamemory (though see Peng & Tullis, 2021). The current studies focused only on monitoring and resolution. However, given that monitoring is an important input to metamemorial control (e.g., for study time decisions, restudy decisions, etc.; e.g., Nelson, 1990), the current results suggest that the poorer resolution in the SPT condition should feedforward, possibly producing less effective metamemorial control in the SPT condition than in the VT or EPT condition, an implication that should be assessed in subsequent research.

In closing, the study of action memory was initially prompted by questions of whether memory for actions was a distinct form of episodic memory (Zimmer et al., 2001). Subsequent research found sufficient similarities between memory for actions and memory for other materials to reject the strong claim of uniqueness but sufficient differences to support the nonstrategic view of action memory (e.g., Sahakyan & Foster, 2009; Schatz et al., 2011; Wang et al., 2021; Wojcik et al., 2011; Zhang & Zuber, 2020; see Mulligan, 2014; Roediger & Zaromb, 2010, for reviews). With regard to metamemory, the study of actions has been largely missing (cf. Susser & Mulligan, 2015; Susser et al., 2017; Yang et al., 2009). The current experiments (and the earlier Cohen studies) indicate that metamemory for actions bears some important differences from metamemory for other materials, in terms of relative accuracy (resolution; supporting the nonstrategic view of action memory) and average JOLs, along with at least one similarity with respect to absolute accuracy (calibration). The differences especially call for additional analysis of action metamemory.

References

Ariel, R., & Dunlosky, J. (2011). The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition*, *39*(1), 171–184. https://doi.org/10.3758/s13421-010-0002-y

Bäckman, L., & Nilsson, L. (1984). Aging effects in free recall: An exception to the rule. *Human Learning: Journal of Practical Research & Applications*, *3*, 53–69.

Bäckman, L., Nilsson, L. G., Herlitz, A., Nyberg, L., & Stigsdotter, A. (1991). Decomposing the encoding of action events: A dual conception. *Scandinavian Journal of Psychology*, *32*(4), 289–299. https://doi.org/10.1111/j.1467-9450.1991.tb00880.x

Bäckman, L., Nilsson, L. G., & Kormi-Nouri, R. (1993). Attentional demands and recall of verbal and color information in action events. *Scandinavian Journal of Psychology*, *34*(3), 246–254. https://doi.org/10.1111/j.1467-9450.1993.tb01119.x

Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition*, *19*, 487–497.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68. https://doi.org/10.1037/0096-3445.127.1.55

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, *41*(6), 897–903. https://doi.org/10.3758/s13421-013-0307-8

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica*, *197*, 153–165. https://doi.org/10.1016/j.actpsy.2019.04.011

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, *22*(1), 267–281. https://doi.org/10.1111/j.1467-9450.1981.tb00402.x

Cohen, R. L. (1983). The effect of encoding variables on the free recall of words and action events. *Memory & Cognition*, *11*(6), 575–582. https://doi.org/10.3758/BF03198282

Cohen, R. L. (1985). On the generality of the laws of memory. In L. Nilsson & T. Archer (Eds.), *Perspectives on learning and memory* (pp. 247–277). Erlbaum.

Cohen, R. L. (1988). Metamemory for words and enacted instructions: Predicting which items will be recalled. *Memory & Cognition*, *16*(5), 452–460.

Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review*, *1*(1), 57–80. https://doi.org/10.1007/BF01326550

Cohen, R. L., & Bean, G. (1983). Memory in educable mentally retarded adults: Deficit in subject or experimenter? *Intelligence*, *7*(3), 287–298. https://doi.org/10.1016/0160-2896(83)90019-3

Cohen, R. L., & Bryant, S. (1991). The role of duration in memory and metamemory of enacted instructions (SPTs). *Psychological Research*, *53*(3), 183–187. https://doi.org/10.1007/BF00941385

Cohen, R. L., Sandler, S. P., & Keglevich, L. (1991). The failure of memory monitoring in a free recall task. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(4), 523.

Cohen, R. L., & Stewart, M. (1982). How to avoid developmental effects in free recall. *Scandinavian Journal of Psychology*, *23*(1), 9–15. https://doi.org/10.1111/j.1467-9450.1982.tb00408.x

De Lucia, N., Milan, G., Conson, M., Grossi, D., & Trojano, L. (2019). Enactment effect in patients with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *41*(9), 965–973. https://doi.org/10.1080/13803395.2019.1642306

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, *26*(6), 741–750. https://doi.org/10.1080/09658211.2017.1404111

Dunlosky, J., & Tauber, S. U. K. (Eds.). (2016). *The Oxford handbook of metamemory*. Oxford University Press.

Engelkamp, J. (1998). *Memory for actions*. Psychology Press/Taylor & Francis.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 671–682. https://doi.org/10.1037/0278-7393.26.3.671

Engelkamp, J., & Krumnacker, H. (1980). Image- and motor-processes in the retention of verbal materials. *Zeitschrift Für Experimentelle Und Angewandte Psychologie*, *27*(4), 511–533.

Engelkamp, J., & Zimmer, H. (1996). Organisation and recall in verbal tasks and in subject-performed tasks. *European Journal of Cognitive Psychology*, *8*, 257–274. https://doi.org/10.1080/095414496383086

Feyereisen, P. (2009). Enactment effects and integration processes in younger and older adults' memory for actions. *Memory*, *17*(4), 374–385. https://doi.org/10.1080/09658210902731851

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 238–244. https://doi.org/10.1037/0278-7393.33.1.238

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, *58*(1), 19–34. https://doi.org/10.1016/j.jml.2007.03.006

Foley, M. A., & Ratner, H. H. (2001). The role of action-based structures in activity memory. In H. D. Zimmer, M. Guynn, R. Cohen, J. Engelkamp, R. Kormi-Nouri, & M. Foley (Eds.), *Memory for action: A distinct form of episodic memory?* (pp. 112–135). Oxford University Press.

Friedman, M. C., & Castel, A. D. (2011). Are we aware of our ability to forget? Metacognitive predictions of directed forgetting. *Memory & Cognition*, *39*(8), 1448–1456. https://doi.org/10.3758/s13421-011-0115-y

Hainselin, M., Quinette, P., Juskenaite, A., Desgranges, B., Martinaud, O., de La Sayette, V., Hannequin, D., Viader, F., & Eustache, F. (2014). Just do it! How performing an action enhances remembering in transient global amnesia. *Cortex*, *50*, 192–199. https://doi.org/10.1016/j.cortex.2013.10.007

Helstrup, T. (1987). One, two, or three memories? A problem-solving approach to memory for performed acts. *Acta Psychologica*, *66*(1), 37–68. https://doi.org/10.1016/0001-6918(87)90017-5

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age differences in the monitoring of learning: Cross-sectional evidence of spared resolution across the adult life span. *Developmental Psychology*, *46*(4), 939–948. https://doi.org/10.1037/a0019812

Hornstein, S. L., & Mulligan, N. W. (2004). Memory for actions: Enactment and source memory. *Psychonomic Bulletin & Review*, *11*(2), 367–372. https://doi.org/10.3758/BF03196584

Ianì, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, *96*, 110–121. https://doi.org/10.1016/j.jml.2017.05.004

Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review*, *25*(6), 2356–2364. https://doi.org/10.3758/s13423-018-1463-4

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Kormi-Nouri, R. (1995). The nature of memory for action events: An episodic integration view. *European Journal of Cognitive Psychology*, *7*(4), 337–363. https://doi.org/10.1080/09541449508403103

Kormi-Nouri, R., & Nilsson, L. G. (2001). The motor component is not crucial! In H. D. Zimmer, R. Cohen, M. J. Guynn, J. Engelkamp, R. Kormi-Nouri, & M. A. Foley (Eds.), *Memory for action: A distinct form of episodic memory?* (pp. 97–111). Oxford University Press.

Kubik, V., Jemstedt, A., Eshratabadi, H. M., Schwartz, B. L., & Jönsson, F. U. (2022). The underconfidence-with-practice effect in action memory: The contribution of retrieval practice to metacognitive monitoring. *Metacognition and Learning*. Advance online publication. https://doi.org/10.1007/s11409-021-09288-2

Lind, S. E., & Bowler, D. M. (2009). Recognition memory, self-other source memory, and theory-of-mind in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*(9), 1231–1239. https://doi.org/10.1007/s10803-009-0735-2

Ma, J., Wang, L., Chen, L., & Zhang, Y. (2021). Imagery processing in action memory–mental imagery is necessary to the subject-performed task effect. *Journal of Cognitive Psychology*, *33*(1), 12–23. https://doi.org/10.1080/20445911.2020.1862129

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory & Cognition*, *29*(2), 222–233. https://doi.org/10.3758/BF03194916

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*(2), 237–255. https://doi.org/10.3758/PBR.15.2.237

McDonald-Miszczak, L., Hubley, A. M., & Hultsch, D. F. (1996). Age differences in recall and predicting recall of action events and words. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *51*(2), P81–P90. https://doi.org/10.1093/geronb/51B.2.P81

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition: Knowing about knowing*. MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

Mulligan, N. W. (2014). Memory for pictures and actions. In T. J. Perfect & D. S. Lindsay (Eds.), *The Sage handbook of applied memory* (pp. 20–36). Sage. https://doi.org/10.4135/9781446294703.n2

Mulligan, N. W., & Hornstein, S. L. (2003). Memory for actions: Self-performed tasks and the reenactment effect. *Memory & Cognition*, *31*(3), 412–421. https://doi.org/10.3758/BF03194399

Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. In B. H. Ross (Ed.), *Psychology of learning and motivation* (pp. 175–214). Elsevier Academic Press.

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, *48*(5), 745–758. https://doi.org/10.3758/s13421-020-01025-5

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, *26*, 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5

Nilsson, L. G., & Cohen, R. L. (1988). Enrichment and generation in the recall of enacted and non-enacted instructions. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 427–432). Wiley.

Nilsson, L., & Craik, F. I. (1990). Additive and interactive effects in memory for subject-performed tasks. *European Journal of Cognitive Psychology*, *2*(4), 305–324. https://doi.org/10.1080/09541449008406210

Peng, Y., & Tullis, J. G. (2021). Dividing attention impairs metacognitive control more than monitoring. *Psychonomic Bulletin & Review*, *28*(6), 2064–2074. https://doi.org/10.3758/s13423-021-01950-9

Peterson, D. J., & Mulligan, N. W. (2010). Enactment and retrieval. *Memory & Cognition*, *38*(2), 233–243. https://doi.org/10.3758/MC.38.2.233

Peterson, D. J., & Mulligan, N. W. (2015). Action memory and encoding time: Evidence for a strategic view of action memory processing. *The American Journal of Psychology*, *128*(4), 419–429. https://doi.org/10.5406/amerjpsyc.128.4.0419

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). Oxford University Press.

Roediger, H. L., & Zaromb, F. M. (2010). Memory for actions: How different? In L. Bäckman & L. Nyberg (Eds.), *Aging, memory and the brain: Essays in honour of Lars-Göran Nilsson* (pp. 24–52). Psychology Press.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L. (2003). Recall of subject-performed tasks, verbal tasks, and cognitive activities across the adult life span: Parallel age-related deficits. *Aging, Neuropsychology and Cognition*, *10*(3), 182–201. https://doi.org/10.1076/anec.10.3.182.16449

Sahakyan, L., & Foster, N. L. (2009). Intentional forgetting of actions: Comparison of list-method and item-method directed forgetting. *Journal of Memory and Language*, *61*(1), 134–152. https://doi.org/10.1016/j.jml.2009.02.006

Schatz, T. R., Spranger, T., Kubik, V., & Knopf, M. (2011). Exploring the enactment effect from an information processing view: What can we learn from serial position analyses? *Scandinavian Journal of Psychology*, *52*(6), 509–515. https://doi.org/10.1111/j.1467-9450.2011.00893.x

Seiler, K. H., & Engelkamp, J. (2003). The role of item-specific information for the serial position curve in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 954–964. https://doi.org/10.1037/0278-7393.29.5.954

Smith, S. J., Souchay, C., & Conway, M. A. (2010). Improving metamemory in ageing and Parkinson's disease. *Age and Ageing*, *39*(1), 116–119. https://doi.org/10.1093/ageing/afp188

Steffens, M. C. (1999). The role of relational processing in memory for actions: A negative enactment effect in free recall. *The Quarterly Journal of Experimental Psychology, A: Human Experimental Psychology*, *52*(4), 877–903. https://doi.org/10.1080/713755860

Steffens, M. C., von Stülpnagel, R., & Schult, J. C. (2015). Memory recall after "learning by doing" and "learning by viewing": Boundary conditions of an enactment benefit. *Frontiers in Psychology*, *6*, 1907. https://doi.org/10.3389/fpsyg.2015.01907

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, *22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition*, *41*(7), 1000–1011. https://doi.org/10.3758/s13421-013-0323-8

Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language*, *95*, 116–123. https://doi.org/10.1016/j.jml.2017.03.002

Tekin, E., & Roediger, H. L. (2020). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift Für Psychologie*, *228*(4), 278–290. https://doi.org/10.1027/2151-2604/a000425

Wang, L., Li, Q., Wu, Q., Takahashi, S., & Wu, J. (2021). The categorical relational process mechanism in enactment learning: Effects of divided attention and categorical cues. *Journal of Cognitive Psychology*, *33*(2), 133–145. https://doi.org/10.1080/20445911.2021.1883032

Wojcik, D. Z., Allen, R. J., Brown, C., & Souchay, C. (2011). Memory for actions in autism spectrum disorder. *Memory*, *19*(6), 549–558. https://doi.org/10.1080/09658211.2011.590506

Yang, S. J., Gallo, D. A., & Beilock, S. L. (2009). Embodied memory judgments: A case of motor fluency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1359–1365. https://doi.org/10.1037/a0016547

Zhang, X., & Wang, L. (2020). Integrative representation of semantic and motor information underlies the enactment effect in memory for actions. *Memory*, *28*(8), 1078–1088. https://doi.org/10.1080/09658211.2020.1815789

Zhang, X., & Zuber, S. (2020). The effects of language and semantic repetition on the enactment effect of action memory. *Frontiers in Psychology*, *11*, 515. https://doi.org/10.3389/fpsyg.2020.00515

Zimmer, H. D., Cohen, R. L., Guynn, M., Engelkamp, J., Kormi-Nouri, R. & Foley, M. A. (Eds.). (2001). *Memory for action: A distinct form of episodic memory?* Oxford University Press.

Zimmer, H. D., & Engelkamp, J. (1999). Levels-of-processing effects in subject-performed tasks. *Memory & Cognition*, *27*(5), 907–914. https://doi.org/10.3758/BF03198543

Received July 2, 2021
Revision received February 16, 2022
Accepted February 18, 2022 ■