



Are mnemonic failures and benefits two sides of the same coin?: Investigating the real-world consequences of individual differences in memory integration

Nicole L. Varga^{1,2} · Trent Gaugler³ · Jennifer Talarico⁴

Published online: 7 January 2019
© The Psychonomic Society, Inc. 2019

Abstract

Theories of reconstructive memory have long been influenced by investigations of false recognition errors, in which old/new judgements are compromised by spontaneous activation of associated but nonpresented concepts. Recent evidence similarly suggests that reconstructive memory processes (so-called memory integration) also support positive learning behaviors, such as inferential reasoning. Despite prevailing hypotheses, the question of whether a common integration process underlies these seemingly disparate mnemonic outcomes is not well understood. To address this question, young adults, recruited from two institutions, completed the Deese–Roediger–McDermott (Deese, *Journal of Experimental Psychology*, 58, 17–22, 1959; Roediger & McDermott, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814, 1995) and Bransford and Franks (*Cognitive Psychology*, 2, 331–350, 1971) false recognition paradigms, as well as an inferential paradigm (Varga & Bauer, *Memory & Cognition*, 45, 1014–1027, 2017b), all of which depend on integration of related information in memory. Across two experiments, the well-established tasks were adapted such that successful memory integration resulted in the same negative outcome (i.e., false recognition; Experiment 1) or positive outcome (i.e., inferential reasoning; Experiment 2). By capturing variability in item-to-item responding within and among tasks for each person, a common memory integration process was found to elicit positive and negative consequences in paradigms that required the combination of individual units to construct a composite understanding, but only when memory for directly learned and novel, integrated items were modeled together. Furthermore, linking task-related behavior to academic performance revealed that a greater propensity to integrate factual information (but not arbitrary materials) was related to higher SAT scores. Together, these results provide evidence for domain-general and domain-specific reconstructive mechanisms and their role in supporting educational success beyond the laboratory.

Keywords Memory integration · Reconstructive memory · Semantic memory · Recognition memory · Episodic memory

Memories are not exact replicas of the past, but instead are dynamic reconstructions of experiences. This reconstructive

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13421-018-0887-4>) contains supplementary material, which is available to authorized users.

✉ Nicole L. Varga
nvarga@austin.utexas.edu

Jennifer Talarico
talaricj@lafayette.edu

¹ Psychology, Emory University, Atlanta, GA, USA

² Present address: University of Texas, Austin, TX, USA

³ Mathematics, Lafayette College, Easton, PA, USA

⁴ Psychology, Lafayette College, Easton, PA, USA

property serves a necessary function in everyday life, enabling the flexible use of existing memories in new contexts (Bauer & Varga, 2017; Schlichting & Preston, 2015). For instance, in one episode, an individual may encounter a woman driving a blue car at the grocery store. In a later episode, the individual may see the same car at the bank, but driven by a man. If the individual notices the commonality between the experiences (the car), he or she may link the man and woman in memory. Furthermore, by forming connections between these events, the individual may extend beyond each of these episodes to derive new knowledge never directly specified, such that the man and woman are in a relationship. The capacity to link overlapping experiences in memory—referred to as *memory integration*—is critical to deriving new knowledge (Varga & Bauer, 2017b). In addition to supporting knowledge extension, memory integration has also been implicated in negative

outcomes—namely, false recollections (Marsh, Cantor, & Brashier, 2016; Schlichting & Preston, 2015). For instance, if two episodes have been linked in memory, they may lose their unique features, causing an individual to later confuse them as one experience. To continue with the example above, if the same individual later learns that the car was leaving the bank after robbing it, he or she may falsely remember that the man and woman were at the bank together. Despite the prevailing view that inferential reasoning and so-called false memories constitute two sides of the same coin—the negative and positive consequences of a domain-general memory integration process—this hypothesis has not yet been tested.

In the present research, we tested the theoretical proposal that engagement of a common integration process may result in both false recognition and successful knowledge extension. Importantly, this suggests that one's propensity to integrate related experiences should lead to opposing behavioral outcomes depending on whether memory for specific experiences or knowledge of relations among those experiences is assessed. To address this question, we modeled between-task associations of performance on well-established paradigms that frame memory integration as either a negative (i.e., false memory; Experiment 1) or a positive (i.e., correct inference; Experiment 2) outcome. We also assessed whether individual differences in the propensity to engage integration processes within and across tasks generalized beyond the laboratory to educational success as measured through the Scholastic Aptitude Test (SAT) and college grade point average (GPA). Together, the findings elucidate how memory integration functions as a generalized reconstructive process, or rather, how reconstructive processes are uniquely dependent on specific task demands.

Several seminal paradigms have identified the properties of memory that support its feats and fallacy, with prevailing theoretical accounts emphasizing the distributed nature of remembering. In other words, a single experience can be represented as a collection of distributed smaller features (McClelland & Rumelhart, 1985; Schacter, Norman, Koutstaal, 1998). As such, retrieval of prior memories is achieved under conditions that enable successful pattern completion, in which reactivation of a subset of the features pertaining to a particular memory triggers spreading activation of other constituent features (McClelland, McNaughton, & O'Reilly, 1995; O'Reilly & Rudy, 2001). To continue the previous example, reexperiencing the blue car at the bank may trigger reactivation of the prior, related memory (e.g., the woman driving it). If memories are dynamically reconstructed during retrieval as this view suggests, then memory errors, distortions, and illusions should commonly occur, especially when individuals are probed about specific episodes that overlap (e.g., retrieval of both the woman and the man when questioned by authorities about who was in the car at the bank; see Schacter et al., 1998, for discussion). Consistent with this notion, in contexts in which differentiation of individually experienced events is challenging, accurate memory retrieval is compromised. The

novel inferences that result from the distributed, flexible properties of memory have therefore traditionally been studied through paradigms that frame the outcome of this process as a mnemonic failure rather than a benefit.

The errors that result from distributed memory networks and reconstructive processing have been documented extensively in laboratory settings (Bartlett, 1932; Schacter et al., 1998). Indeed, the idea that all memory entails distributed, constructive processing gained in strength following demonstrations that distortions could be elicited with simple list-learning paradigms. For example, in the Deese–Roediger–McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) paradigm, participants are presented with individual words (*bed, rest, awake*) that are highly interassociated to a critical, nonpresented lure item (*sleep*). They then are asked to judge whether studied items and lures are old (i.e., studied) or new (i.e., nonstudied). Incorrect “old” responses to nonpresented lures (*sleep*) are comparable with correct hit rates for studied items (*bed*). This false-recognition effect has been explained by the constructive memory processes of spreading activation among related concepts stored in semantic memory (e.g., Meade, Watson, Balota, & Roediger, 2007; Roediger, Balota, & Watson, 2001) and/or through explicit generation of a gist representation of the nonpresented item (Brainerd & Reyna, 1998; Schacter et al., 1998). Support for these constructivist mechanisms comes from the finding that the more the studied items are associated with the critical, nonpresented concept (i.e., backward associative strength), the more robust the false recognition (e.g., Deese, 1959; Roediger, Watson, McDermott, & Gallo, 2001). As such, the DRM paradigm capitalizes on the strength of preexisting associative memory structures, such that studied concepts that are already strongly integrated with the critical lure in memory elicit higher false recognition rates to the nonpresented items (Roediger, Watson, et al., 2001). Indeed, the finding that false recognition judgments on the DRM are accompanied with high confidence suggests that these integrated concepts are spontaneously activated, thereby making errors difficult to monitor at test (Gallo, 2010).

Another form of mnemonic error that results from reconstructive, integrative processing is composite recollections, in which a single retrieved memory encompasses the combination of two or more events (e.g., Schooler & Tanaka, 1991). Unlike the DRM paradigm, which relies on activation of a prior memory experienced outside the experimental task, composite recollection paradigms examine the productive combination of novel premises learned within the paradigm. For instance, Bransford and Franks (1971) showed that individuals spontaneously integrated discrete sentences to abstract a wholistic idea never directly specified. When individuals learned that *the ants were in the kitchen, the jelly was on the table, and the ants ate the sweet jelly*, they reliably endorsed the novel, nonpresented sentence *the ants in the kitchen ate the*

sweet jelly which was on the table as “old.” Moreover, recognition confidence ratings on the Bransford and Franks (BF) task were highest for never-presented, four-unit sentences relative to novel sentences comprising fewer semantic units, presumably because four-unit items conveyed the complete, unified semantic idea that was spontaneously derived through integration during learning. Thus, when using full sentences (BF) or individual words (DRM), false recognition might result from a common integration process that activates previously associated units in long-term memory or supports generation of novel links among newly learned concepts.

More recently developed paradigms have examined the linkage of discrete memories with respect to how it supports positive task behaviors, including the derivation of new semantic knowledge (Bauer & Jackson, 2015; Varga & Bauer, 2017b), successful inferential reasoning (Preston, Shrager, Dudukovic, & Gabrieli, 2004), and decision-making (Kumaran, Summerfield, Hassabis, & Maguire, 2009). Similar to the literature on reconstructive memory errors, theories of memory integration have also relied on distributed memory models to account for how these positive behaviors emerge. Indeed, recent evidence from the self-derivation through integration (SDI) paradigm suggests that when an individual learns a new fact (*blood is produced in the skeleton*) that overlaps with prior knowledge (*hematopoiesis is the formation of blood*), the shared feature (*blood*) triggers rapid retrieval of the prior, related content (Varga & Bauer, 2017a). Through this process, indirectly experienced features (*hematopoiesis; skeleton*) are simultaneously activated and linked. Moreover, the extent to which integration mechanisms are spontaneously engaged during learning relates to how well individuals later productively extend their knowledge to derive new understandings and make inferential judgments at the time of test (e.g., *hematopoiesis occurs in the skeleton*; Varga & Bauer, 2017a; Zeithamova, Dominick, & Preston, 2012; Zeithamova & Preston, 2010; though see Kumaran & McClelland, 2012, for an alternative account). Thus, taken together, the theoretical models proffered to explain false recognition and inferential reasoning similarly emphasize the spontaneous activation of distributed, associated units in memory. It is therefore plausible that paradigms that frame the reconstructive capacity of memory as a negative or a positive outcome might engage a common process.

In two experiments, we tested the hypothesis that false recognition (as assessed through the DRM and BF) and productive knowledge extension (as assessed through the SDI) rely on the same underlying integration process. If a common integration process is engaged to encode separate-yet-related words (DRM), sentences (BF), and facts (SDI), then individuals should perform similarly across these tasks. In light of vast individual differences in the capacity for memory integration (Shohamy & Wagner, 2008; Varga & Bauer, 2017b), associations among tasks were modeled *within* individuals,

thus enhancing sensitivity to detect similar patterns of behavior across tasks. Moreover, to minimize the differential contribution of processes recruited during task-specific judgments (monitoring old/new source vs. logical truth/fallacy of novel, integrated premises), we modified the test demands such that the decision and resulting outcome (positive or negative) was consistent for each task within an experiment. That is, in Experiment 1 we adapted the SDI paradigm such that memory integration would result in a mnemonic failure (i.e., individuals were asked to judge whether the novel integration facts were old or new). In Experiment 2, we adapted the paradigms to frame the outcomes of interest as a benefit, emphasizing the productive meaning abstracted (i.e., whether the information that could be inferred was true or false). If mnemonic errors and benefits are, indeed, two sides of the same reconstructive memory coin, then it is critical to demonstrate that the processes that produce comparable negative consequences across tasks also produce positive consequences.

Identification of commonalities or differences in patterns of behavior among well-established paradigms has important implications for our understanding of basic, constructive memory processes, as well as for our understanding of the opposing behaviors memory integration is purported to support. Finally, to further examine whether a domain-general integration process generalizes beyond the laboratory paradigms, we also characterized the relation between task-based processes and academic success in the form of college GPA and SAT scores. To the degree that academic success relies on integration and extension of newly acquired information as well as accurate dissociation of past experience from similar, but inaccurate, events, then the academic measures should relate to the laboratory tasks in both experiments.

Experiment 1

Method

Participants

Participants were 144 adults recruited from undergraduate subject pools at their respective institutions (63 from Emory, 50 females; 81 from Lafayette, 61 females) who received course credit for participation. An additional seven participants took part in the study but were excluded due to technological failure ($n = 6$) and participation in a prior, related study that compromised the measures assessed here ($n = 1$). The sample was mostly non-Hispanic ($n = 58$ and 73 at Emory and Lafayette, respectively) and White ($n = 27$ and 63 at Emory and Lafayette, respectively). Comparisons of participant characteristics between institutions are provided in the [Supplemental Materials](#). The protocol and procedures for both

experiments were approved by the respective Institutional Review Boards.

Materials and procedure

Prior to data collection, all procedures were preregistered at AsPredicted (<https://aspredicted.org/rv9wb.pdf>). Participants were tested individually on the DRM, BF, and an adapted SDI paradigm, each of which included associated study items and an old/new recognition test that captured false alarms to critical, nonpresented items. Participants were instructed that they would be asked to read, respond to, and remember single words, grammatical sentences, and true factual statements, and that we were interested in whether performance was related across the three different tasks. After providing informed consent, participants completed each of the three paradigms on an individual computer presented via a Qualtrics (2016) questionnaire. As depicted in Fig. 1a, the study and test phases of each paradigm were completed consecutively. Whereas the encoding task differed across paradigms (see Fig. 1a), the test phase was identical such that participants were always asked to judge whether the individual words (DRM) or sentences (BF and SDI) were old or new. Paradigm order was randomized across participants. Moreover, within each phase (study/test) of the paradigms, the order of items was counterbalanced such that half the participants saw one pseudorandomized order and the other half saw its reverse. The survey presentation concluded with participant characteristic questions. The materials and procedure for each of the three paradigms are outlined below.

Deese–Roediger–McDermott paradigm (DRM) During study, participants viewed 48 items drawn from the four lists with the highest false recognition rates to the critical lures in Roediger, Watson, and colleagues (2001; i.e., *cold, rough, window, smell*). The 12 most strongly backwards associated items from each of the four lists were employed. For example, one list consisted of the words *hot, frigid, chilly, frost, heat, winter, ice, shiver, warm, snow, freeze, and arctic*, which were associated to the nonpresented lure *cold*. Items were presented for 1 second, and participants were instructed to read each word carefully before the next item automatically appeared (see Fig. 1a). No two items from the same list were presented consecutively.

During test, participants saw 28 items and were asked to determine if each was old or new (see Fig. 1a). Of those, eight were old items (the first and sixth most strongly backwards associated items; e.g., *chilly* and *hot*, respectively), four were the critical, nonpresented lures (one from each list; e.g., *cold*), eight were related new items (the two least backwards associated items from Roediger, Watson, et al., 2001; e.g., *air* and *weather*), and eight were unrelated new items (e.g.,

king, whistle). As above, no two items from the same list were presented consecutively.

Bransford and Franks paradigm (BF) The materials were adapted from the original Bransford and Franks (1971) investigation. During study, 24 sentences were presented, with six coming from each of four complete (though never explicitly presented) conceptual ideas (*ants, hut, man, breeze*). Of the six sentences from each idea, there were two each that included one, two, and three units of information—for example, *The ants were in the kitchen, The ants ate the sweet jelly, and The ants in the kitchen ate the jelly which was on the table*, respectively. Individual sentences could be combined to form a composite idea (*The ants in the kitchen ate the sweet jelly which was on the table*). As shown in Fig. 1a, participants were asked to identify the subject or object of each studied sentence (evenly distributed among the four concepts). No two items from the same conceptual idea or unit length were presented consecutively. Moreover, no more than three subject (or object) questions were presented consecutively.

During test, participants were presented with 44 sentences and asked to judge whether they were old or new (see Fig. 1a). Of those, 12 were old (one one-unit, two-unit, and three-unit sentence from each conceptual idea). The remaining items were new, including the four critical four-unit composite sentences that could be derived from integration of the individual study items. Four distractor four-unit sentences unrelated to the concepts presented during study were also presented (e.g., *The brown dog chased the striped ball in the yard*). The remaining new items were evenly distributed among one-unit, two-unit, and three-unit length sentences, half of which were related to the original four ideas and half of which were unrelated. No more than three related/unrelated, three same unit-length sentences, or two old items were presented consecutively.

Modified self-derivation through integration paradigm (SDI)

During study, participants read 20 pairs of related facts (40 individual “stem” facts; e.g., *Cyanide is found in pips; Apple seeds are called pips*). Similar to the protocol employed by Bauer and Jackson (2015), participants were asked to judge whether a typical student would know this fact before arriving at college (see Fig. 1a). Related facts were separated by a lag of four to eight intervening sentences.

During test, participants were presented with 40 sentences and asked to judge whether they were old or new (see Fig. 1a). Ten sentences were old and pertained to one of the two facts from a related pair (e.g., either *Cyanide is found in pips* or *Apple seeds are called pips*). Ten statements were integrated from previously presented facts, but were “new” because participants had not seen those exact sentences during study (e.g., *Apple seeds contain cyanide*). The remaining 20 items were new distractor facts unrelated to items seen during study (e.g.,

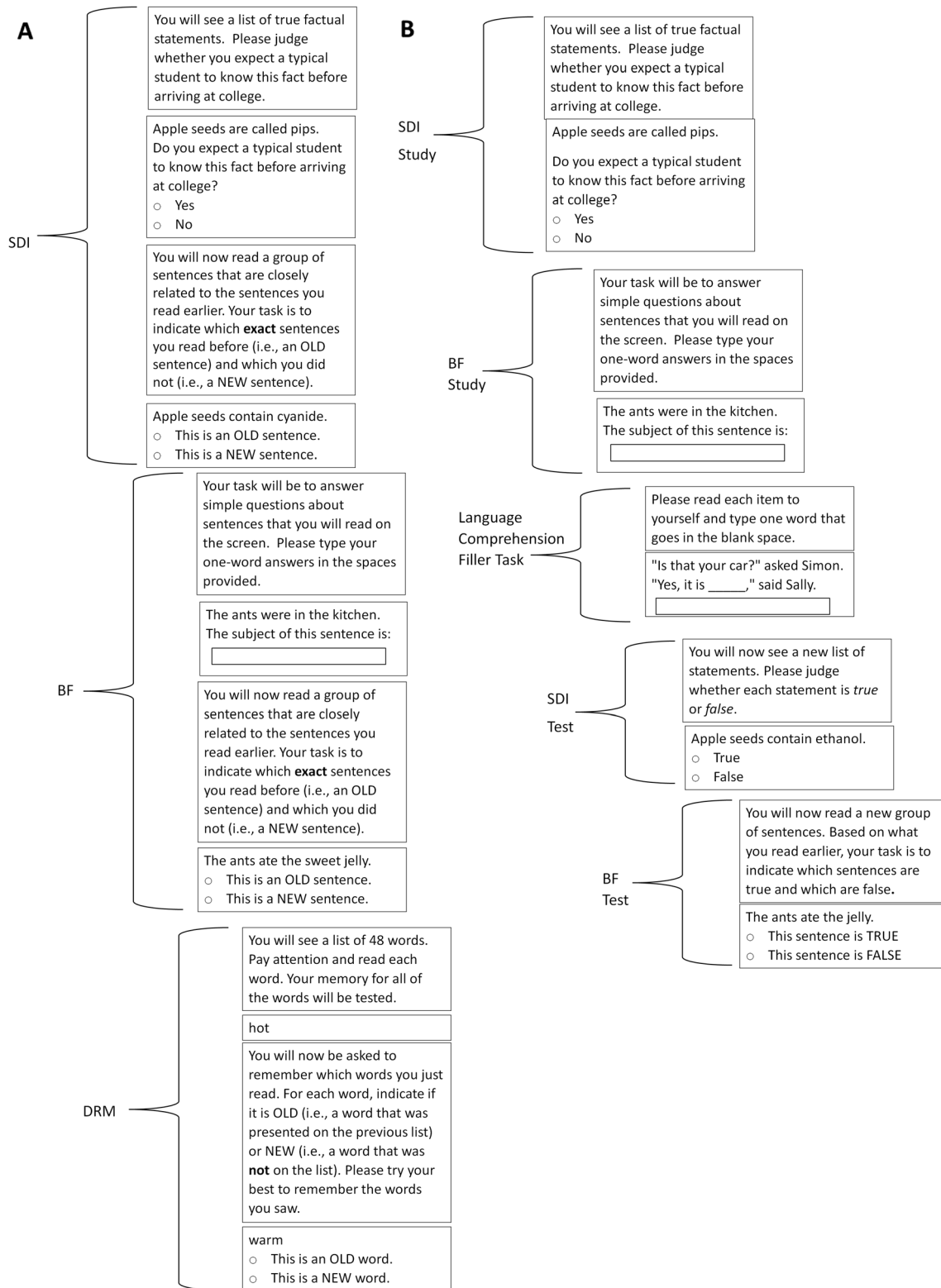


Fig. 1 Schematic of procedure for Experiments 1 (a) and 2 (b)

The world's largest biome is the taiga). No more than three related, three novel distractor items, or two “old” items were presented consecutively. Note that no participant saw both the one-stem (“old”) and integrated two-stem (“new”) version of

the same concept during test. Instead, there were two versions of the test, such that the one-stem facts in one order appeared as the corresponding new integration facts in the other order (and vice versa) across the sample.

Academic measures College GPA was calculated at the end of the semester of participation. The highest SAT score at the time of enrolment was obtained from the college registrars. In the case that an individual only took the ACT (21% of participants), the College Board concordance table (College Board Research, 2009) was used to convert the score into the SAT equivalent.

Data analytic approach

The primary question we sought to address concerned whether, *within* a participant, responses to the individual false-alarm items (i.e., “old” responses to the four novel critical lures in DRM, four novel composite ideas in BF, and 10 novel integration sentences in SDI) were associated across paradigms, as well as with measures of educational performance. To assess this, we modeled the data using the PROC GLIMMIX procedure in SAS Studio 3.6 (SAS Institute, Cary, NC), jointly modeling the multivariate clustered responses to the critical lure items (18 total) and the continuous measures of GPA and SAT. The procedure allowed for flexible modeling of mixed generalized linear models. More specifically, instead of modeling the mean response directly, we modeled a function of the mean response as a typical regression. We fit our joint model with two separate link functions for the two distinct types of response variables, using the logit link for the sets of binary old/new responses to the critical lures (18 total) and the identity link for the academic measures (two total). Each response was also modeled as a function of the respondent’s undergraduate institution, counterbalancing order, and task (i.e., BF, DRM, SDI, SAT, and GPA). We used the Kenward–Roger method for obtaining the degrees of freedom (Alnosaier 2007; Kenward & Roger, 1997) and included a random effect for task.

To calculate estimates of task variability for each individual, the covariance structure was blocked by participant ID. As depicted in Fig. 2, the various dependencies within each individual’s suite of responses were modeled with a unique 20 × 20 covariance matrix. In doing so, it was possible to calculate a common covariance estimate of the propensity to give similar item-to-item responses within a task (i.e., one parameter estimate for each of matrices A–E) and between tasks (i.e., one parameter estimate for each of the 10 nonshaded matrices) at the individual level. Moreover, the covariance structure type was unstructured for the task effect, which permitted differing, unconstrained levels of correlation between all possible sets of pairwise task combinations: three recognition paradigm combinations (e.g., BF × DRM), one academic-measure combination (e.g., SAT × GPA), and six paradigm × academic combinations (e.g., BF × SAT; BF × GPA).

We chose to focus only on the critical novel items from the DRM and BF, relative to the other novel, related items included in the recognition tests, because they have previously been

	SDI Questions (10)	BF Questions (4)	DRM Questions (4)	GPA Measure (1)	SAT Score (1)
SDI Questions (10)	A	AB	AC	AD	AE
BF Questions (4)	AB	B	BC	BD	BE
DRM Questions (4)	AC	BC	C	CD	CE
GPA Measure (1)	AD	BD	CD	D	DE
SAT Score (1)	AE	BE	CE	DE	E

Fig. 2 Schematic of an individual-specific 20 × 20 covariance structure. Submatrices A–E (shaded along the diagonal) represent the separate compound symmetric structures for each of the five within-task responses (e.g., 10 × 10 for SDI, 4 × 4 for BF, etc.) in Experiment 1. Each of the remaining, nonshaded submatrices located outside of the main diagonal provide a single covariance between propensities to give similar responses for all 10 possible combinations of tasks (e.g., 4 × 10 and transposed 10 × 4 submatrix AB represents the covariance between the BF and SDI tasks) *Note.* SDI = self-derivation through integration; BF = Bransford and Franks; DRM = Deese–Roediger–McDermott

shown to elicit the highest rates of false recognition and thus provide a stronger test of whether a common integration process was engaged among paradigms. Critically, the potential power issues associated with the lower number of items included was counteracted by the number of individuals sampled (see [Supplemental Materials](#) for power simulations). In addition to examining false alarms to the critical lures alone, we also examined whether rejection of related yet nonstudied information (critical new items) was associated with memory for item-specific information (studied old items). Evidence from the DRM paradigm shows that reductions in false recognition are associated with increases in veridical memory for studied items, suggesting that better encoding of individual items protects against incorrect endorsement of related lures (Roediger, Watson, et al., 2001). Hence, if a common process is engaged during encoding of individual (yet related) words (DRM), sentences (BF), and factual statements (SDI), then cross-task associations might also be observed when interactions between responses to studied items and related lures are considered. We accounted for this possibility by including three interaction terms that differentially modeled the effect of the hit rates for old items on the BF, DRM, and SDI paradigms by the task type.

Results

The task-based data and program code used for analyses are available at <https://osf.io/te84d/>. Group-level descriptive statistics for each paradigm are reported in Table 1 (see [Supplemental Results](#) for across-paradigm performance comparisons and simulation data suggesting the nonsignificant effect of near ceiling and floor performance on the power to detect between-task associations). The primary series of analyses concerned whether item-to-item responses were associated *within* individuals. Thus, we first assessed whether the propensities to false alarm to the critical lure items within the same paradigm were significantly correlated with one another. As reflected in the main diagonal of Table 2, all of the covariance parameters were significant and exhibited positive confidence intervals, which indicates that the propensity to false alarm to items within a paradigm (e.g., *cold*, *rough*, *window*, *smell* for the DRM) were positively correlated. That is, performance on the subset of critical lure items within each task was highly reliable.

We next assessed whether an individual's propensity to false alarm to an item in one paradigm was associated with the propensity to false alarm to an item in a different paradigm. These results are summarized above the main diagonal in Table 2. As all three of the between-paradigm confidence intervals included zero, there was no evidence of significant correlations at the .05 level. We then used the covariance parameter estimates to evaluate whether an individual's propensity to give false-alarm responses within each paradigm was associated with the GPA and SAT measures (see Table 3). The covariances between the propensity to false alarm and academic performance were not statistically significant for any of the paradigm/academic comparisons.

Finally, we explored the degree to which the propensity to give false-alarm responses to items in a paradigm was associated with the true hit rate for old items in that same paradigm.

Table 2 Estimated covariances of propensities to false alarm on critical lures within and between paradigms in Experiment 1

	SDI	BF	DRM
SDI	1.75 (1.24, 2.64)	.50 (−.08, 1.08)	.04 (−.34, .43)
BF		.51 (.19, 3.33)	−.34 (−.76, .07)
DRM			.62 (.35, 1.41)

Note. Parenthetical values reflect 95% Wald confidence intervals. SDI = self-derivation through integration; BF = Bransford and Franks; DRM = Deese–Roediger–McDermott

The only paradigm for which this association was statistically significant was BF, $F(1, 96.4) = 18.58, p < .001$. The estimated association was positive, indicating that as the true hit rate increased for BF old items, the propensity to false alarm on BF critical new items increased as well. For both DRM and SDI, the test for this association was not statistically significant ($p > .3$). We then explored the associations of true hit rates and false alarm rates *between* paradigms, as well as associations of true hit rates and performance on the SAT and GPA measures. For each of the three true hit rate predictors included in the model (BF, DRM, SDI), the estimated relations and p values are reported in Table 4. The only across-task interaction to reach the conventional level of statistical significance with a Bonferroni correction was between BF and SDI, such that hit rates on BF were positively associated with the propensity to false alarm on the SDI (uncorrected p values also reported in Table 4).

Discussion

A domain-general view of memory integration suggests that individuals should exhibit similar performance on tasks thought to engage this process, irrespective of specific task demands. Contrary to this view, the results of Experiment 1

Table 1 Group-level performance on old (hits) and critical new (false alarms) items across samples and paradigms in Experiment 1

	Paradigm	Old		New critical	
		<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range
Full Sample (<i>N</i> = 144)	DRM	.77(.19)	.38–1.00	.63(.30)	.00–1.00
	BF	.82(.15)	.42–1.00	.88(.19)	.00–1.00
	SDI	.91(.14)	.10–1.00	.16(.22)	.00–1.00
Emory (<i>n</i> = 63)	DRM	.79(.20)	.38–1.00	.65(.30)	.00–1.00
	BF	.83(.16)	.42–1.00	.85(.23)	.00–1.00
	SDI	.91(.15)	.10–1.00	.15(.19)	.00–.80
Lafayette (<i>n</i> = 81)	DRM	.76(.18)	.38–1.00	.61(.29)	.00–1.00
	BF	.82(.14)	.42–1.00	.90(.16)	.50–1.00
	SDI	.90(.13)	.10–1.00	.18(.24)	.00–.90

Note. DRM = Deese–Roediger–McDermott; BF = Bransford and Franks; SDI = self-derivation through integration

Table 3 Estimated covariances between propensities to false alarm on critical lures from each recognition paradigm with continuous measures of GPA and SAT in Experiment 1

	GPA	SAT/ACT
SDI	-.01 (-.12, 0.11)	-28.97 (-60.93, 2.99)
BF	-.06 (-.19, 0.06)	-23.58 (-57.26, 10.10)
DRM	-.03 (-.12, 0.07)	10.60 (-14.33, 35.54)

Note. Parenthetical values reflect 95% Wald confidence intervals. SDI = self-derivation through integration; BF = Bransford and Franks; DRM = Deese–Roediger–McDermott

indicate that an individual’s propensity to false alarm to nonpresented yet related words (DRM), combined linguistic units (BF), and integrated facts (SDI) were not significantly associated. Nevertheless, the propensity to accurately remember studied linguistic units on the BF was associated with the propensity to incorrectly judge the novel integration facts on the BF and SDI as “old.” Thus, although false-alarm behaviors were not directly associated across tasks, the processes engaged during encoding and/or retrieval of studied, linguistic items (BF) interacted with failures of the same participant to identify integrated sentences (BF and SDI) as novel.

Another major purpose of the present experiment was to test whether the reconstructive processes that these laboratory paradigms capture extend to measures of academic performance. Examination of within-subject relations between task behavior and academic measures revealed no association between the propensity to false alarm and SAT or GPA. Therefore, individual differences in the ability to judge the source of integrated information do not seem to be associated with more multifaceted cognitive outcomes.

So far, there is evidence for some commonality in processing between the BF and SDI paradigms (at least when item-specific memory is included in the model). Hence, it is possible that this across-task association reflects a commonality in how to-be-integrated BF and SDI sentences were initially processed, which made subsequent monitoring of the integrated units more challenging. To shed further light on this issue, in Experiment 2 we further probed commonalities between these tasks when memory integration instead supported a mnemonic benefit, namely, derivation of new knowledge. This additional analysis is important for establishing whether the commonalities found here are common to all possible outcomes of memory integration, or unique to contexts in which it induces errors. Because the DRM relies on activation of concepts already established in long-term memory rather than newly learned information, it was not possible to adapt it into a true/false paradigm and was thus not employed in Experiment 2.

Experiment 2

Method

Participants

Participants were 151 adults (70 from Emory, 51 females; 81 from Lafayette, 55 females) recruited from the same subject pools and similarly compensated as in Experiment 1. No participant who had completed Experiment 1 was included in Experiment 2. The sample was mostly non-Hispanic ($n = 57$

Table 4 Estimated relations of true hit rates with both propensities to false alarm (FA) to critical lures from other paradigms as well as with continuous measures of GPA and SAT in Experiment 1

Paradigm	Association with	Estimate	DF	<i>t</i> value	Unadjusted <i>p</i> value	Bonferroni-adjusted <i>p</i> value
BF (hits)	DRM (FA)	1.30	132.1	1.75	.08	.33
	SDI (FA)	3.07	165.7	2.98	.003	.01
	SAT	51.05	2858	.79	.43	1
	GPA	.52	145.1	2.28	.02	.10
DRM (hits)	BF (FA)	1.70	113.3	2.21	.03	.12
	SDI (FA)	.46	140.2	.59	.56	1
	SAT	-10.56	2858	-2.1	.84	1
	GPA	.01	145.2	.07	.94	1
SDI (hits)	BF (FA)	1.52	76.61	1.65	.10	.41
	DRM (FA)	0.96	127.9	1.22	.22	.89
	SAT	76.36	2858	1.11	.27	1
	GPA	.05	114.7	.06	.96	1

Note. BF = Bransford and Franks; DRM = Deese–Roediger–McDermott; SDI = self-derivation through integration

and 73 at Emory and Lafayette, respectively) and White ($n = 36$ and 70 at Emory and Lafayette, respectively). See the [Supplemental Materials](#) for comparisons of participant characteristics across institutions.

Materials and procedure

Prior to data collection, all procedures were preregistered at AsPredicted (<https://aspredicted.org/gw9zw.pdf>). Participants were tested individually on two tasks, variations of the BF and SDI paradigms from Experiment 1, but that framed mnemonic reconstruction as a positive outcome. Instead of an old/new judgment, participants were asked if the items were true or false in the sense of factual content. If participants were able to spontaneously construct integrative understandings by flexibly combining the information presented during study, they should be able to evaluate the accuracy of never-before-presented but logically congruent statements. As in Experiment 1, individuals were instructed to read, respond to, and remember grammatical sentences and true factual statements (see Fig. 1b for complete instructions).

After providing informed consent, participants completed the tasks on an individual computer via a Qualtrics (2016) questionnaire, which was structured like the one used in Experiment 1 (i.e., instruction timing, thank-you text following each portion, and demographic questions presented at the end). Unlike the study-test phase blocking employed in Experiment 1, to minimize ceiling effects, participants completed both study phases before completing the corresponding test phases (see Fig. 1b for a schematic overview). Participants completed a 5–10-minute language-comprehension filler task between the study and test phases. As in Experiment 1, paradigm order was randomized across participants. Moreover, within each phase (study/test) of each paradigm, the order of items was counterbalanced such that half the participants saw one pseudorandomized order and the other half saw its reverse. The materials and procedures for each paradigm are outlined below.

Modified Bransford and Franks paradigm (BF) The procedure and materials during study were identical to those employed in Experiment 1. However, during test, participants saw 32 sentences, none of which were presented during study. Half of those sentences were conceptually true and consisted of composites of sentences presented at study (one sentence each of one-unit, two-unit, three-unit, and four-unit lengths from each of the four conceptual ideas). Unlike Experiment 1, in which the critical, nonpresented four-unit sentences should be judged as “new” (*The ants in the kitchen ate the sweet jelly which was on the table*), here, the same sentences should be judged as “true.” The false sentences shared the same structure in that they were created by combining components from the original four conceptual ideas, yet they combined units

across rather than within ideas. For example, at study participants encoded *The hut was tiny* (hut idea) and *The old man was resting on the couch* (man idea) and were later asked whether *The hut was old* was a true or false statement. This served an important control to ensure that participants were tracking the meaning of the sentences and not combining all information indiscriminately. No more than two units from the same concept, two true or false items, or two sentences of the same unit length were presented consecutively.

Modified self-derivation through memory integration paradigm (SDI) During study, participants read 100 total sentences, including the same 20 pairs of related “stem” facts from Experiment 1 and 60 additional stand-alone, nonintegrable distractor facts (e.g., *Ethanol results in fewer greenhouse gas emissions than gasoline*). Despite the additional facts, the encoding task, temporal spacing parameters, and counterbalancing criteria were identical to those in Experiment 1 (see Fig. 1b).

During test, participants were presented with 20 factual statements and asked to evaluate whether they were factually true or false (see Fig. 1b). Half of the statements were true, constructed through integration of pairs of previously presented stem facts (e.g., *Apple seeds contain cyanide*). The other half were also derived through integration of related stem facts; however, a key word was substituted with a term from a distractor sentence to create a conceptually plausible but false statement (e.g., *Apple seeds contain ethanol*), thus acting as a similar control as that employed in the BF true/false task. There were two versions of the test, with each integration sentence being presented in “true” and “false” form across the versions. No more than three true or false items were presented consecutively.

Academic measures The academic measures were the same as in Experiment 1. Because 34% of participants took the ACT, their scores were converted into the SAT equivalent.

Data analytic approach

The central question we sought to address was whether, *within* a single participant, responses to the same critical items from Experiment 1 were associated between paradigms and with measures of educational success when the judgment assessed was veracity of information (true/false) rather than the source of that understanding (old/new). We employed the same modeling strategy as in Experiment 1, jointly modeling the multivariate response consisting of the clustered responses to the critical true items conserved across experiments (four BF, 10 SDI) and the continuous measures of GPA and SAT, with the various dependences within each individual’s suite of responses modeled with a unique 16×16 covariance matrix. Importantly, the true items included were the same critical lure items modeled in Experiment 1, thereby enabling examination

of the opposing outcome on the same set of stimuli. Each response was also modeled as a function of undergraduate institution, counterbalancing order, task (i.e., BF, SDI, SAT, and GPA), and two interaction terms that differentially modeled the effect of propensities to correctly reject false items for the BF and SDI instruments by task type (constituting an analogous approach to the examination of hits in Experiment 1). Thus, the model implemented here provides a complementary test of the claim that a common integration process supports behavior across the BF and SDI tasks and that individual differences in these processes have implications for academic success.

Results

The task-based data and program code used for analyses are available at <https://osf.io/te84d/>. Group-level descriptive statistics for each paradigm are reported in Table 5 (see Supplemental Results for between-paradigm performance comparisons), yet the key question concerns associations between tasks within individuals. Table 6 displays the associations for the propensities to correctly identify true premises within a paradigm (along the diagonal) and across paradigms (off the diagonal). The covariance parameters for within-task item associations were significant and exhibited positive confidence intervals, indicating that performance on one true item within a paradigm was positively related to the other true items within that task. We next assessed the main question of whether an individual’s propensity to identify a true item in the SDI was associated with the propensity to identify a true item in the BF. As displayed above the diagonal in Table 6, there was no statistically significant across-paradigm association with respect to the propensity to correctly identify a sentence as true.

We next examined the covariance parameter estimates capturing associations between propensities to correctly identify true sentences on the different paradigms with the GPA and SAT measures (see Table 7). The covariance between the

Table 5 Group-level accuracy on true and false items across paradigms and samples in Experiment 2

	Paradigm	True		False	
		<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range
Full Sample <i>N</i> = 151	BF	.90(.10)	.56–1.00	.83(.17)	.19–1.00
	SDI	.72(.19)	.20–1.00	.73(.18)	.20–1.00
Emory <i>n</i> = 70	BF	.92(.10)	.56–1.00	.85(.16)	.30–1.00
	SDI	.76(.19)	.20–1.00	.76(.17)	.30–1.00
Lafayette <i>n</i> = 81	BF	.89(.11)	.56–1.00	.81(.18)	.25–1.00
	SDI	.68(.19)	.20–1.00	.70(.18)	.20–1.00

Note. BF = Bransford and Franks; SDI = self-derivation through integration

Table 6 Estimated covariances of propensities to correctly identify true integration facts between paradigms in Experiment 2

	SDI	BF
SDI	.45 (.29, .81)	.26 (–.14, .66)
BF		.69 (.24, 6.48)

Note. Parenthetical values reflect 95% Wald confidence intervals. *Note.* SDI = self-derivation through integration; BF = Bransford and Franks

propensity to correctly identify true statements and SAT was significant for the SDI paradigm, but not for BF. The covariances of the propensity to correctly identify true statements with GPA were not significant for either paradigm.

We also assessed the degree to which the propensity to correctly identify true sentences within a paradigm was associated with the rate at which an individual correctly rejected false sentences within the same paradigm. This association was only statistically significant for BF, $F(1, 75.54) = 7.24, p = .009$, and the estimated relationship was positive, indicating that as the rate at which an individual correctly rejected false sentences increased, the propensity to correctly identify true sentences also increased. Finally, we explored the association of the rate at which individuals correctly rejected false sentences with the rate at which they correctly identified true sentences *between* paradigms, as well as with performance on the SAT and GPA measures. As depicted in Table 8, the only such interaction to reach the conventional level of statistical significance was between SDI and SAT, such that correctly rejecting false integration sentences on the SDI was positively associated with performance on the SAT. A marginal relation was also observed between BF and SDI, such that correctly rejecting false sentences on the BF was positively associated with identifying true sentences on the SDI (unadjusted p value = .06).

Discussion

The present experiment investigated whether individuals exhibit similar performance on mnemonic tasks that frame the products of memory integration as a positive consequence, in this case, as the derivation of novel yet true semantic concepts. A significant positive association was observed between

Table 7 Estimated covariances between propensities to correctly identify true integration facts from the two paradigms and continuous measures of GPA and SAT in Experiment 2

	GPA	SAT
SDI	.05 (–.03, .13)	29.02 (11.58, 46.46)
BF	.03 (–.15, .22)	23.86 (–21.58, 69.31)

Note. Parenthetical values reflect 95% Wald confidence intervals. *Note.* SDI = self-derivation through integration; BF = Bransford and Franks

Table 8 Estimated relations of correct rejections of false items with both propensities to correctly identify true items from other paradigms as well as with continuous measures of GPA and SAT in Experiment 2

Paradigm	Association with	Estimate	DF	<i>t</i> value	Unadjusted <i>p</i> value	Bonferroni-adjusted <i>p</i> value
BF (correct rejections)	SDI (true items)	.90	128.4	1.87	.06	.19
	SAT	−16.85	2402	−.30	.76	1
	GPA	-.24	159.6	−1.01	.31	.94
SDI (correct rejections)	BF (true items)	.62	124.8	.56	.58	1
	SAT	177.02	2402	3.24	.001	.004
	GPA	.30	160.8	1.33	.19	.56

Note. Parenthetical values reflect 95% Wald confidence intervals *Note.* SDI = self-derivation through integration; BF = Bransford and Franks

accurate identification of the true BF items and correct rejection of false BF items. The results of Experiment 2 also provided marginal evidence for within-person associations between items that required judgement of the veracity of integrated linguistic units (BF) and integrated factual statements (SDI), such that the propensity to reject false BF sentences was positively associated with the propensity to adopt true SDI sentences. However, caution must be exerted in interpreting this finding given that it only held when corrections for multiple comparisons were not applied. Finally, a higher propensity to successfully identify true integration facts and to correctly reject false integration facts on the SDI was positively related to SAT performance. Conversely, no association was found between the BF and SAT nor between either paradigm and GPA.

General discussion

In two experiments, we evaluated the prevailing proposal that a domain-general memory integration process may lead to false recognition and productive knowledge extension (e.g., Marsh et al., 2016; Schlichting & Preston, 2015). When we examined the propensity to inaccurately judge the source (Experiment 1) and to accurately judge the veracity (Experiment 2) of self-constructed words (DRM), linguistic units (BF), or factual statements (SDI), there was no evidence for across-task associations. However, when we additionally examined the interaction between performance on old versus new items (Experiment 1) and true versus false items (Experiment 2), between-task associations emerged between the BF and SDI (though this association failed to reach the conventional level of significance in Experiment 2). Furthermore, individual differences in the laboratory-based, factual self-derivation task (SDI) generalized to real-world academic success, such that those who better judged the novel, integration facts as “true” and “false” in Experiment 2 had higher SAT scores. In the discussion to follow, we consider the

unique and distinct processes that lead to performance within and across mnemonic paradigms, as well as the implications for our understanding of the processes that apply to success in academic endeavors.

The current research takes an important step toward furthering our understanding of the process(es) that underlie the productive extension and erroneous by-products of a reconstructive memory system. A maximally adaptive memory system exhibits domain-general properties that enable individuals to solve novel problems across a range of domains (Chiappe & MacDonald, 2005). Yet this adaptive quality may also make us prone to errors, particularly when we spontaneously derive novel inferences without awareness that these conceptions were not directly experienced (Marsh et al., 2016). In Experiment 1, the only between-task association emerged once memory for directly encoded items was modeled in conjunction with the propensities to false alarm to the critical lures. Specifically, the propensity to correctly judge previously studied linguistic sentences as “old” on the BF was related to the propensity to incorrectly judge both within-task novel linguistic sentences (BF) and between-task novel integration facts (SDI) as “old.” Thus, individuals who correctly recognized studied items on the BF task were the same individuals who incorrectly judged critical new items as “old” on both the BF and SDI tasks. It is important to emphasize that this negative hit/false alarm association on BF is in the opposite direction as that previously documented for the DRM (Roediger, Watson, et al., 2001). The negative hit/false alarm association sometimes observed in the DRM is thought to reflect monitoring processes engaged during retrieval, such that external features of experience that accompany retrieval of previously studied items supports rejection of internally generated lure items that lack this component. However, given the opposite pattern of association between BF hits and BF/SDI false alarms here, it is unlikely that this finding reflects the same type of source-monitoring/reality-monitoring typically attributed to the DRM task (e.g., Johnson, Hashtroudi, & Lindsay, 1993). Instead, it is possible that individuals

spontaneously integrated the BF sentences into wholistic semantic ideas at the time of initial encoding. Thus, when it came time to engage in source or reality monitoring during the recognition test, robust memory for the individual traces made it more difficult to reject semantically consonant (but never presented) sentences as “new.”

The suggestion that individuals who exhibited less accurate performance on the critical new items in BF and SDI in Experiment 1 were more likely to have spontaneously integrated at the time of encoding is consistent with findings from Bransford and Franks (1971). In the original study, individuals were most confident of “recognizing” novel sentences that contained more recombined studied units (the four-unit critical lures employed here) relative to fewer units (three-unit, two-unit, and one-unit sentences). This suggests that a spontaneously engaged integration process may drive this reconstructive error. False alarms to conjunction items (i.e., the wholistic linguistic units in the BF and integration facts in the SDI) can thus be considered an interference effect, in which the recombination of previously presented stimulus components makes it difficult to source the discrete units that originally contained them (Reinitz & Hannigan, 2004). Moreover, because the ability to link separate yet related information requires comprehension of the individual units, it is logical that better memory for the individual premises (i.e., hits to old BF items) would be associated with incorrectly judging novel yet integrated sentences as previously learned (on both the BF and SDI). Furthermore, this across-task association was likely specific to old items on the BF (rather than the SDI) because encoding in this task involved exposure to six individual sentences within each linguistic component (as opposed to only two from a pair in the SDI), making it inherently more difficult to maintain the unique features of these individual sentences and link them all together to construct the logically consistent, wholistic four-unit idea. Consistent with this interpretation, at the group level, the BF paradigm elicited fewer hits relative to the SDI paradigm. Hence, it is plausible that the same integrative encoding mechanism resulted in false alarms on the BF and SDI tasks, but that hits to BF items simply provided a more sensitive measure of this integration process due to its more difficult nature—individuals who integrated across the overlapping premises were also more likely to retain precise memories for the six individually experienced units.

The pattern of associations (and lack thereof) among tasks in Experiment 1 builds upon and extends our understanding of the cognitive processes engaged during related learning episodes. Recent evidence has suggested that variability in the capacity to self-derive new knowledge through memory integration is positively associated with working memory capacity (Varga, Esposito, & Bauer, 2019). Likewise, in a memory conjunction paradigm analogous to the BF linguistic task in which individuals studied sequentially presented words (stargaze, catfish) and were then tested for false recognition of the

recombined units (stargaze), errors were substantially reduced when working memory was blocked (Reinitz & Hannigan, 2004). This parallel pattern of prior results, coupled with the between-task BF/SDI association evidenced in the present research, provides further support for the idea that working memory underlies the construction of an integrated, composite representation during initial learning. Conversely, high working memory spans do not influence false recall of critical words in the DRM paradigm under incidental learning conditions (Watson, Bunting, Poole, & Conway, 2005). Therefore, the noticeable absence of an association between the BF and SDI paradigms with the DRM is consistent with the proposed role of working memory in supporting the construction of integrated memory representations. Furthermore, unlike the BF and SDI paradigms, which capture the construction of novel links among concepts, the DRM relies on the activation of a nonpresented word based on preexisting semantic knowledge. This suggests that whereas the BF and SDI require one to actively maintain and flexibly represent the unique and overlapping aspects of related sentences in memory (see Schlichting, Mumford, & Preston, 2015, for corroborating evidence from a comparable memory integration paradigm), the DRM relies on activation of relational links already formed in long-term memory.

Examining whether a domain-general memory integration process produces positive and negative consequences required paradigms that could be adapted to elicit errors in determining the source of integrated traces as well as accurate recognition of logically consistent inferences. Because prior exposure to the positive or negative task demand would have compromised our measure of the opposing outcome, it was necessary to test the different mnemonic outcomes in distinct samples. Thus, in Experiment 2, different individuals were exposed to the same related linguistic units (BF) and factual statements (SDI) that elicited false recognition in Experiment 1. When asked to judge the truth of novel, integrated sentences (Experiment 2), rather than whether it was old or new (Experiment 1), we observed high levels of accuracy in identifying sentences that were logically consistent (true) as well as inconsistent (false) with previously learned information. Indeed, the lowest average accuracy for any item category (true vs. false) was 72% correct. This pattern of results indicates that, under mirrored encoding conditions that utilized the same study and critical sentences, the spontaneous linkage of related information further supports the acquisition of novel, true wholistic conceptualizations. Importantly, although performance was high, our supplemental simulation data suggested that it did not interfere with our ability to detect between-task associations.

Similar to the false recognition versions of the BF and SDI paradigms employed in Experiment 1, there was marginal

evidence for a between-task association in the true/false versions implemented in Experiment 2. Specifically, the propensity to reject false BF sentences was positively associated with the propensity to identify true BF and true SDI sentences. Importantly, when judging the incorrect sentence that *The hut was old* in the BF task, it was not sufficient to rely solely on familiarity of the individual items because each false statement contained elements presented during learning: *The hut was tiny* and *The old man was resting on the couch*. Hence, indiscriminate combination of nonrelated concepts or reliance on familiarity-based responding would have led to errors on the false (but not true) items in this task, which could account for the null relation observed between the BF and SDI when true items were examined alone. Moreover, like Experiment 1, because the false items on the BF task were drawn from a relatively small set of conceptual units (i.e., the same four overarching concepts were recombined across all test items), accuracy on false items perhaps provided a more sensitive measure of one's capacity to encode, comprehend, and retrieve directly learned premises (at least relative to the SDI false items which were drawn from a pool of 20 different concepts that were more distinct and potentially less confusable). Thus, a greater propensity to reject false BF sentences should have implications for one's ability to demonstrate knowledge of true sentences abstracted through combination of individually related units (BF and SDI true items).

Taken together, the finding that no across-task associations were evident when critical items were examined alone rather than in conjunction with performance on directly experienced items is potentially revealing of the nature of the commonality between the SDI and BFI tasks. One possible explanation is that, in the absence of "old" (studied) or "false" (indiscriminately integrated sentences) items capable of capturing the strength of one's memory precision for directly acquired items, that responding to the conjunction items alone does not adequately capture the integration processes invoked at encoding. As noted above, false alarms (Experiment 1) and identification of true integrated ideas (Experiment 2) may be accomplished through applying an indiscriminate, familiarity-based strategy during the test phase. As such, information regarding how one's veridical memory for individual premises interacts with one's judgement of the productive extension of that information provides additional sensitivity to detect subtle mnemonic processes engaged during learning that later support subsequent test-phase responses. The importance of considering both item-specific and relational processing is further supported by simulation data which indicated that power to detect between-task associations was unlikely to be affected by measurement limitations, suggesting that the null associations among critical items were not simply due to the lower number of items sampled. Therefore, we suggest that by modeling memory for item-specific information, we were able to capture one's propensity to spontaneously integrate related

premises during learning (at least on the BF and SDI), which similarly influenced positive and negative task judgments of those conjunction items across both experiments.

A final aim of the present research was to determine whether the reconstructive processes engaged across the false recognition (Experiment 1) and/or productive knowledge extension paradigms (Experiment 2) generalized beyond the laboratory to apply to academic behaviors. The only task that was associated with academic success was the factual self-derivation task (SDI), only when the consequence of memory integration resulted in a positive outcome (Experiment 2), and only with the standardized aptitude measure. That is, individuals who better judged the veracity of integrated, factual content derived in Experiment 2 performed better on the SAT. Notably, because evaluation standards often vary between individual classes (and institutions), our ability to observe significant associations between tasks and GPA may have been limited. Given that explicit self-derivation of integrated knowledge has been shown to be related to concurrent academic success in children (Esposito & Bauer, 2017) as well as to longitudinal academic success in children and adults (Varga et al., 2019), it is not surprising that recognition of logically consistent and inconsistent factual premises as measured here would confer similar benefits for academic success. Yet when individuals were exposed to the same to-be-integrated SDI facts and were instead asked to monitor the old/new status of the content rather than the conceptual meaning, no relation to academic success was observed. This finding suggests that the end-product of memory integration, and not the mnemonic processes that support memory integration alone, jointly contribute to predictions of academic success. The null correlations between academic performance with the BF and DRM further supports the proposal that the predictive utility of memory integration is specific to contexts in which the by-product of this learning ability persists in memory over the long term. Indeed, knowledge newly derived through integration exhibits high retention over 1-week periods in both children (Varga & Bauer, 2013; Varga, Stewart, & Bauer, 2016) and adults (Varga & Bauer, 2017b).

In summary, the present research employed an individual difference approach to evaluate whether a common integration process supports mnemonic feats and fallacies (e.g., Marsh et al., 2016; Schlichting & Preston, 2015). Critically, when individual differences are typically examined, between-subject associations constitute the modal approach, thereby limiting evaluation of whether the correlates and component processes of reconstructive memory are true for each and every individual (Kanning, Ebner-Priemer, & Schlicht, 2013). Through capturing variability in item-to-item responding within and among several validated reconstructive memory tasks for each person (and across two separate institutions), we provide a strong test of whether a domain-general

mechanism was engaged in different tasks. We found that a greater propensity to correctly recognize directly experienced items (Experiment 1) and to reject combinations of items from unrelated units (Experiment 2) on the BF task was associated with a greater propensity to false alarm to (Experiment 1) and accurately identify (Experiment 2) logically consistent yet nonpresented premises on the BF and SDI tasks. Moreover, we found that performance on the factual knowledge integration task (SDI) generalized to academic success when the outcome of integration was framed as a positive consequence that emphasized the semantic meaning rather than source of that knowledge. This finding is important because it suggests, at least with respect to the factual integration task, that the products that result from memory integration have implications for educational success, whereas, metamemory for how that integrated information was acquired is less critical to academic performance. Hence, the integration process engaged during tasks that require spontaneous linkage of several separate yet related premises appears to result in both negative and positive consequences. At the same time, the finding that cross-task similarities were unique to tasks that required integration of novel premises but not long-standing semantic associations suggests that reconstructive processes are also influenced by specific task demands (i.e., formation of integrated representations online vs. retrieval of previously integrated information). Nevertheless, methodological constraints precluded examining both the positive and negative outcomes within the same individual, or equating the materials employed across tasks (number of items, item difficulty/content, etc.). Thus, future research should further explore whether different collections of seemingly related tasks, such as those that similarly define integration as a positive outcome but employ arbitrary versus meaningful units, rely on the same underlying domain-general cognitive processes.

Acknowledgements This research was partially supported by funding from NIH HD67359. The authors extend their sincere appreciation to Patricia J. Bauer for helpful discussions, valuable feedback, and ongoing support throughout the completion of this work. The authors also thank Allison McHayle and Zeyu Xue for collecting data at Lafayette College, as well as Veronica Morgan for assistance with obtaining academic records at Emory University. Finally, the authors extend their appreciation to all the individuals who participated in the research, without whom this work would not have been possible.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Alnoasir, W. S. (2007). *Kenward-Roger approximate F test for fixed effects in mixed linear models*. Unpublished doctoral thesis, Oregon State University, Corvallis, Oregon.
- Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge: Cambridge University Press.
- Bauer, P. J., & Jackson, F. L. (2015). Semantic elaboration: ERPs reveal rapid transition from novel to known. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(1), 271–282. doi:<https://doi.org/10.1037/a0037405>
- Bauer, P. J., & Varga, N. L. (2017). Similarity and deviation in event segmentation and memory integration: Commentary on Richmond, Gold, & Zacks. *Journal of Applied Research in Memory and Cognition*, 6(2), 124–128.
- Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to “remember” than things that were. *Psychological Science*, 9(6), 484–489.
- Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331–350.
- Chiappe, D., & MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *The Journal of General Psychology*, 132(1), 5–40.
- College Board Research. (2009). *ACT and SAT concordance tables* (Office of Research and Development Research Notes, RN-40). Retrieved from <https://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-2009-40-act-sat-concordance-tables.pdf>
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58, 17–22.
- Esposito, A. G., & Bauer, P. J. (2017). Going beyond the lesson: Self-generating new factual knowledge in the classroom. *Journal of Experimental Child Psychology*, 153, 110–125.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, 38(7), 833–848.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3–28.
- Kanning, M. K., Ebner-Priemer, U. W., & Schlicht, W. M. (2013). How to investigate within-subject associations between physical activity and momentary affective states in everyday life: A position statement based on a literature overview. *Frontiers in Psychology*, 4(187), 1–16. doi:<https://doi.org/10.3389/fpsyg.2013.00187>
- Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53, 983–997.
- Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, 119(3), 573–616.
- Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. A. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron*, 63, 889–901. doi:<https://doi.org/10.1016/j.neuron.2009.07.030>
- Marsh, E. J., Cantor, A. D., & Brashier, N. M. (2016). Believing that humans swallow spiders in their sleep: False beliefs as side effects of the processes that support accurate knowledge. In B. Ross (Ed.), *Psychology of learning and motivation* Vol. 64, pp. 93–132. Cambridge: Academic Press.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.
- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114(2), 159–188.
- Meade, M. L., Watson, J. M., Balota, D. A., & Roediger, H. L., III. (2007). The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. *Journal of Memory and Language*, 56(3), 305–320.
- O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, 108(2), 311–345.

- Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, *14*(2), 148–152.
- Qualtrics [Computer software]. (2016). Provo, UT: Qualtrics. Available from <http://www.qualtrics.com>
- Reinitz, M. T., & Hannigan, S. L. (2004). False memories for compound words: Role of working memory. *Memory & Cognition*, *32*(3), 463–473.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814.
- Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385–407.
- Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association.
- Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, *49*(1), 289–318.
- Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*(8151), 1–10.
- Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8.
- Schooler, J. W., & Tanaka, J. W. (1991). Composites, compromises, and CHARM: What is the evidence for blend memory representations?. *Journal of Experimental Psychology: General*, *120*(1), 96–100. doi: <https://doi.org/10.1037/0096-3445.120.1.96>
- Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*(2), 378–389. doi: <https://doi.org/10.1016/j.neuron.2008.09.023>
- Varga, N. L., & Bauer, P. J. (2013). Effects of delays on 6-year-old children's self-generation and retention of knowledge through integration. *Journal of Experimental Child Psychology*, *115*(2), 326–341.
- Varga, N. L., & Bauer, P. J. (2017a). Using event-related potentials to inform the neurocognitive processes underlying knowledge extension through memory integration. *Journal of Cognitive Neuroscience*, *29*(11), 1932–1949.
- Varga, N. L., & Bauer, P. J. (2017b). Young adults self-derive and retain new factual knowledge through memory integration. *Memory & Cognition*, *45*, 1014–1027.
- Varga, N. L., Esposito, A. G., & Bauer, P. J. (2019). *Cognitive correlates of memory integration across development: Explaining variability in an educationally relevant phenomenon*. Manuscript under revision.
- Varga, N. L., Stewart, R. A., & Bauer, P. J. (2016). Integrating across episodes: Investigating the long-term accessibility of self-derived knowledge in 4-year-old children. *Journal of Experimental Child Psychology*, *145*, 48–63.
- Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 76–85.
- Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168–179.
- Zeithamova, D., & Preston, A. R. (2010). Flexible memories: Differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *The Journal of Neuroscience*, *30*(44), 14676–14684.