Supplemental Materials

Experiment 1 Method and Results

Stimulus validation and scoring. The stimuli enabled integration across a number of subject domains and logical relations. For example, from the biological sciences, participants learned that "Hematopoiesis is the cellular formation of blood" and that "The skeleton is the site of the production of blood" which could lend itself to self-derivation of the novel integration fact that "Hematopoiesis is a process that occurs in the skeleton". A major purpose of Experiment 1 was to ensure that the stimuli could not be systematically inferred after exposure to a single stem fact from a target pair (1-stem condition) as compared to when both stems were provided (2-stem condition). Scoring was conducted in an identical manner between conditions, such that self-derivation was considered successful if participants provided the correct sentence-final word of each integration fact (e.g., skeleton) or if they produced an accurate conceptual synonym (e.g., bones). In cases in which participants failed to generate an answer, they were permitted to say "I don't know" or "Skip" (thus, unsuccessful trials include both incorrect and unanswered questions). Participants' responses were hand-scored online.

For some fact pairs, we found that participants sometimes provided an accurate label derived from the stem facts but that did not necessitate integration. For example, participants could integrate the stem fact that "Modern day Belgium is the location of Waterloo" with the stem fact that "Napoleon's final defeat was at Waterloo" to self-derive the fact that "Napoleon lost his last battle in the country of Belgium." Instead of stating "Belgium" when prompted for knowledge of the integration fact, participants might answer "Waterloo," which though true, could be answered based on exposure to either of the individual stem facts. In these situations, the experimenter asked a disambiguating follow-up question (e.g., Can you tell me another word that would also accurately complete this sentence?). In cases in which the experimenter failed to do so (and hence, participants might have successfully integrated but failed to spontaneously provide the unique sentence-final label), we omitted the trials from analysis (1.29% of the total possible 1-stem and 2-stem trials; no more than 3 trials per participant).

In addition to examining between-group self-derivation performance in the 1-stem and 2-stem conditions, we also examined each fact pair individually. For instance, when the stem fact about either “Hematopoiesis is the cellular formation of blood” or “The skeleton is the site of the production of blood” was previously presented (but not both), participants provided the accurate label “skeleton” on only 6% of the trials (1-stem condition), as compared to 53% of the trials when both stems were previously provided (2-stem condition). A similar pattern of results was obtained for the “Belgium” fact pair, such that participants successfully self-derived the integration fact on 73% of the 2-stem trials, which stood in stark contrast to the 0% exhibited in the 1-stem condition. With the exception of one fact pair, for all fact pairs included in the stimulus set, self-derivation was at least twice as great in the 2-stem condition than the 1-stem condition (with 70% of the stimulus pairs far exceeding this criterion). In the case of the one exception, self-derivation was nearly identical in the 1-stem and 2-stem conditions. As a result, trials assessing performance on this fact were excluded from all reported analyses. A proportion score was calculated for each participant and condition, which accounted for the exclusion of the omitted fact and any missing data due to ambiguous answers.

Prior knowledge of the stem facts. The proportion of successfully self-derived integration facts was calculated within participants for cases in which neither stem fact was previously known, only the first stem was known, only the second stem was known, and both stems were known. Supplemental Table 1 (Panel A) lists the means, standard deviations, and number of trials per condition. A one-way repeated-measures analysis of variance (ANOVA) revealed that the effect of prior knowledge on self-derivation was not significant, $F(3, 21) = 1.32$, $p = 0.29$, $p\eta^2 = 0.16$. However, because there was a maximum of 15 possible trials in the 2-stem condition, and the 15 trials had one of four possible prior-knowledge outcomes, 75% of participants had missing data for at least one possible outcome and were thus excluded (reflected by the *N* of 8 reported in Supplemental Table 1, Panel A). To deal with this issue, we additionally examined the proportion of successful self-derivation for cases where neither stem was previously known compared to when at least one stem was previously known (only two participants were excluded from this two level prior-knowledge outcome analysis due to missing data). As reflected in Supplemental Table 1 (Panel A), there was a significant effect of prior knowledge, $F(1, 28) = 14.43$, $p = 0.001$, $p\eta^2 = 0.34$, such that participants were more likely to self-derive the integration fact when they knew at least one stem fact prior to participating in the study.

Experiment 2 Method and Results

Visual-auditory learning task. Investigations of integration of separate yet related paired associates (AB, BC) also frequently employ measures of studied, nonrelated associations (DE) to control for variance explained by memory for directly learned items (e.g., Kumaran et al., 2009; Preston et al., 2004; Shohamy & Wagner, 2008; Zeithamova & Preston, 2010). Because a major goal of the present work was to test retention of self-derived knowledge over a delay, we could not conduct a within-subject assessment of memory for the directly learned stem facts because it would have compromised the test for long-term retention of the integrated knowledge. Yet because it is important to control for the effects of memory for directly learned items (as opposed to integration, per se), at Session 2 we assessed performance on a number of standardized tasks, including long-term retrieval of directly learned paired associates.

The Visual-Auditory Learning (VAL) subtest of the *WJ-III COG* (Test 2) served as a measure of associative memory (*Median* reliability = 0.86 from 5-19 years; 0.91 from 20-90 years) (Woodcock et al., 2001). In this task, participants are shown a series of rebuses (pictographic symbols of words) and later asked to recall the visual-auditory associations from long-term memory. Specifically, once the rebuses are learned, participants are presented with several rebuses forming a sentence and are asked to speak the associated words aloud. Because this task requires the ability to store information and to retrieve it later in the process of thinking, Test 2 is argued to reflect the broad Cattell-Horn-Carroll (CHC) factor of long-term retrieval. Importantly, this factor is not to be confused with one's long-term memory store which constitutes the *contents* of knowledge. Instead, this task taps the general processing abilities of storing and retrieving information from long-term memory. Participants received one point for each incorrectly answered item, defined as a failure to identify the correct word or to do so within 5 seconds of viewing a rebus. The correct word was provided if participants failed to state it within the 5-second time limit. Because scoring was conducted on-line, an independent coder listened to all audio recordings to ensure that the 5-second pause was reliably scored. If participants were allotted more than 5 seconds, the item was subsequently scored as incorrect. If participants were corrected too soon, the item was counted as a missing trial (0.59% of trials). A proportion score was then derived by dividing the total number of errors by the number of valid trials. Due to reliance on the number of errors, this score should negatively correlate with self-derivation performance.

Self-derivation scoring. As in Experiment 1, in cases in which participants provided an ambiguous answer and the experimenter failed to provide a disambiguating follow-up question, we omitted the trials from analysis (0.83% of the total possible trials; no more than 2 trials per participant). Trials assessing performance on the fact pair that did not produce greater 2-stem compared to 1-stem performance were also excluded from all reported analyses (see above).

Partial correlations controlling for memory for directly learned items. As expected, the Pearson correlation coefficient between the proportion of errors on the VAL task and proportion of successfully self-derived integration facts revealed a significant negative relation, $r(99) = -0.30$, $p = .002$, such that individuals who exhibited better self-derivation performance also exhibited better memory for directly learned associates. Yet when the effect of memory for directly learned items was partialled out, the significant negative relation between self-derivation and response time on all trials, $pr(98) = -0.21$, $p = .04$, as well as with response time on only correct trials, $pr(98) = -0.22$, $p = .03$, was still observed. Although the positive relation between self-derivation and response time on incorrect trials was still apparent, the correlation failed to reach statistical significance after controlling for directly learned items, $pr(98) = 0.18$, $p = .068$.

Similar to reaction time measures, the previously reported Spearman correlations between self-derivation and explicit awareness also remained when controlling for memory for directly learned items. That is, when VAL performance was partialled out, there was still a significant positive correlation between self-derivation and explicit awareness, $pr_s(69) = 0.40$, $p < .001$, as well as between explicit awareness and response time on incorrect trials, $pr_s(69) = 0.28$, $p = .02$. The non-significant relation between explicit awareness and response time on the total corpus of open-ended integration questions, $pr_s(69) = 0.13$, $p = .27$, and on successful trials, $pr_s(69) = 0.03$, $p = .83$, also remained. Thus, the previously reported correlations between self-derivation performance and other individual difference measures (i.e., response times and explicit awareness) remained broadly consistent after controlling for the independent assessment of memory for directly learned paired associates.

Chi square analyses of relations between awareness, self-derivation, and response time. Because explicit awareness of the task structure can also be treated as a dichotomous variable (scored 0 or 1), a chi-square test of independence was performed to further examine its relation to self-derivation and response time during test. To do so, we used a median split to categorize participants as "high" or "low" self-derivation performers (*Median* = 0.52) as well as "fast" or "slow" responders, separately for the total corpus of trials (*Median* = 7061 ms), for correct trials (*Median* = 3971 ms), and for incorrect trials (*Median* = 12002 ms). Consistent with the correlation analyses reported above, explicit awareness was related to self-derivation performance, X^2 (1, N = 69) = 8.92, p = .003. As reflected in Supplemental Figure 1, participants who were aware of the opportunity to integrate were more likely to be "high" performers relative to those who were unaware. As reflected in Panels A and B of Supplemental Figure 2, there was no relation between explicit awareness and response time on the total corpus of trials, X^2 (1, N = 72) = 0.45, p = .51, or on correct trials, X^2 (1, N = 72) = 0.006, p = .94. In contrast, as reflected in Supplemental Figure 2 (Panel C), explicit awareness was associated with response speed on incorrect trials, X^2 (1, N = 72) = 11.39, p = .001, such that participants who were aware of the opportunity to integrate were more likely to spend longer on incorrect trials relative to those who were unaware. The finding that explicit awareness was only associated with response speed on incorrect trials converges with the pattern of results reported based on correlational analyses.

Lag effects. As reflected in Supplemental Figure 3, a one-way repeated measures ANOVA revealed that successful self-derivation performance at Session 1 did not differ as a function of lag, F(2, 232) = 0.66, p = .52, $p\eta^2$ = 0.006. Yet it is possible that the presence of lag effects would be contingent upon whether participants were explicitly aware of the opportunity to integrate. To assess this possibility, we also examined self-derivation across lags separately for participants who were and were not aware of the relational structure of the task. When these additional ANOVAs were conducted, no significant main effect of lag was observed for participants who were explicitly aware, $F(2, 94) = 0.63$, $p = 0.54$, $p\eta^2 = 0.01$, nor for those who were unaware of the task structure, $F(2, 48) = 0.72$, $p = 0.49$, $p\eta^2 = 0.03$ (see Supplemental Figure 3). Thus, the degree of lag between to-be-integrated stem facts had no discernable impact on subsequent knowledge extension, even when controlling for the potential effect of explicit awareness.

Prior knowledge of the stem facts. The proportion of successfully self-derived integration facts was calculated within participants for cases in which neither stem fact was previously known, only the first stem was known, only the second stem was known, and both stems were known. A one-way repeated-measures ANOVA revealed a significant effect of prior knowledge, $F(2.68, 225.12) = 40.12$, $p < .001$, $p\eta^2 = 0.32$. As reflected in Supplemental Table 1 (Panel B), follow-up post hoc tests using Bonferonni corrected adjustments for multiple comparisons indicated that participants were more likely to self-derive the integration fact when they reported prior knowledge of both stem facts, which differed from performance in all other conditions ($ps < .001$). Moreover, participants were least likely to self-derive the integration fact when they possessed no prior knowledge of the stem facts, which statistically differed from performance in all other conditions ($ps < .001$). Conversely, self-derivation did not vary as a function of whether the first stem or second stem was reported to be known ($p = 1.00$). As noted in Supplemental Table 1 (Panel B), only 85 participants in the sample contributed data to all four prior-knowledge outcomes. Therefore, we also examined self-derivation according to whether neither stem was previously known or at least one stem was known (see Supplemental Table 1, Panel B for number of subjects and trials entered into analyses). As expected, there was a significant effect of prior knowledge, $F(1, 115) = 194.43$, $p < 0.001$, $p\eta^2 = 0.63$, such that self-derivation of the integration facts was enhanced when participants reported knowing at least one stem fact relative to no stem facts.

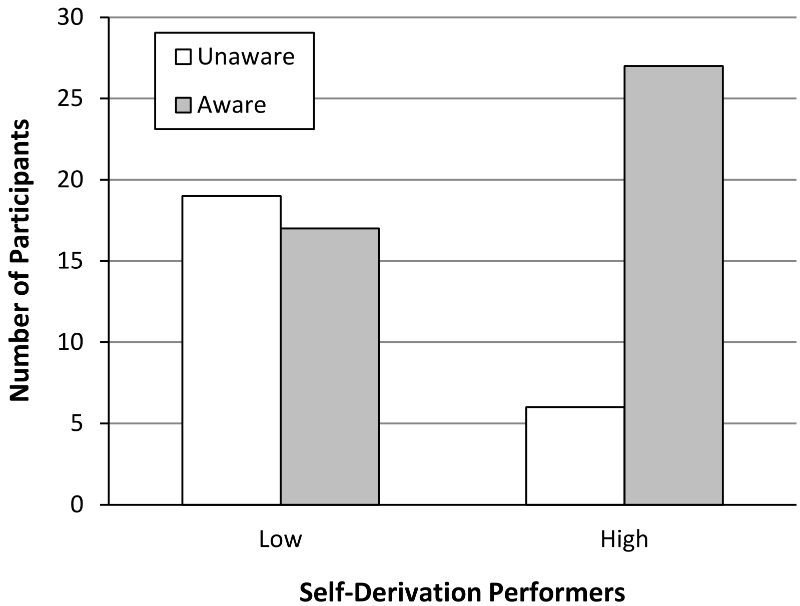
Supplemental Table 1.

Means and standard deviations for proportion of successfully self-derived integration facts as a function of the number of stem facts previously known in Experiment 1 (Panel A) and Experiment 2 (Panel B).

Experiment 1: Panel A				
Analysis	Stems Known	*M*	*SD*	*M* Trials
Four Outcomes (N = 8)	Zero	0.32	0.34	8.57
	First	0.66	0.40	3.00
	Second	0.56	0.50	1.14
	Both	0.50	0.46	1.86
Two Outcomes (N = 29)	Zero	0.39	0.28	10.34
	At Least One	0.58	0.35	4.07
Experiment 2: Panel B				
Analysis	Stems Known	*M*	*SD*	*M* Trials
Four Outcomes (N = 85)	Zero	0.44	0.23	17.66
	First	0.66	0.34	3.92
	Second	0.67	0.32	3.92
	Both	0.84	0.26	3.24
Two Outcomes (N = 116)	Zero	0.42	0.23	19.07
	At Least One	0.70	0.24	9.68

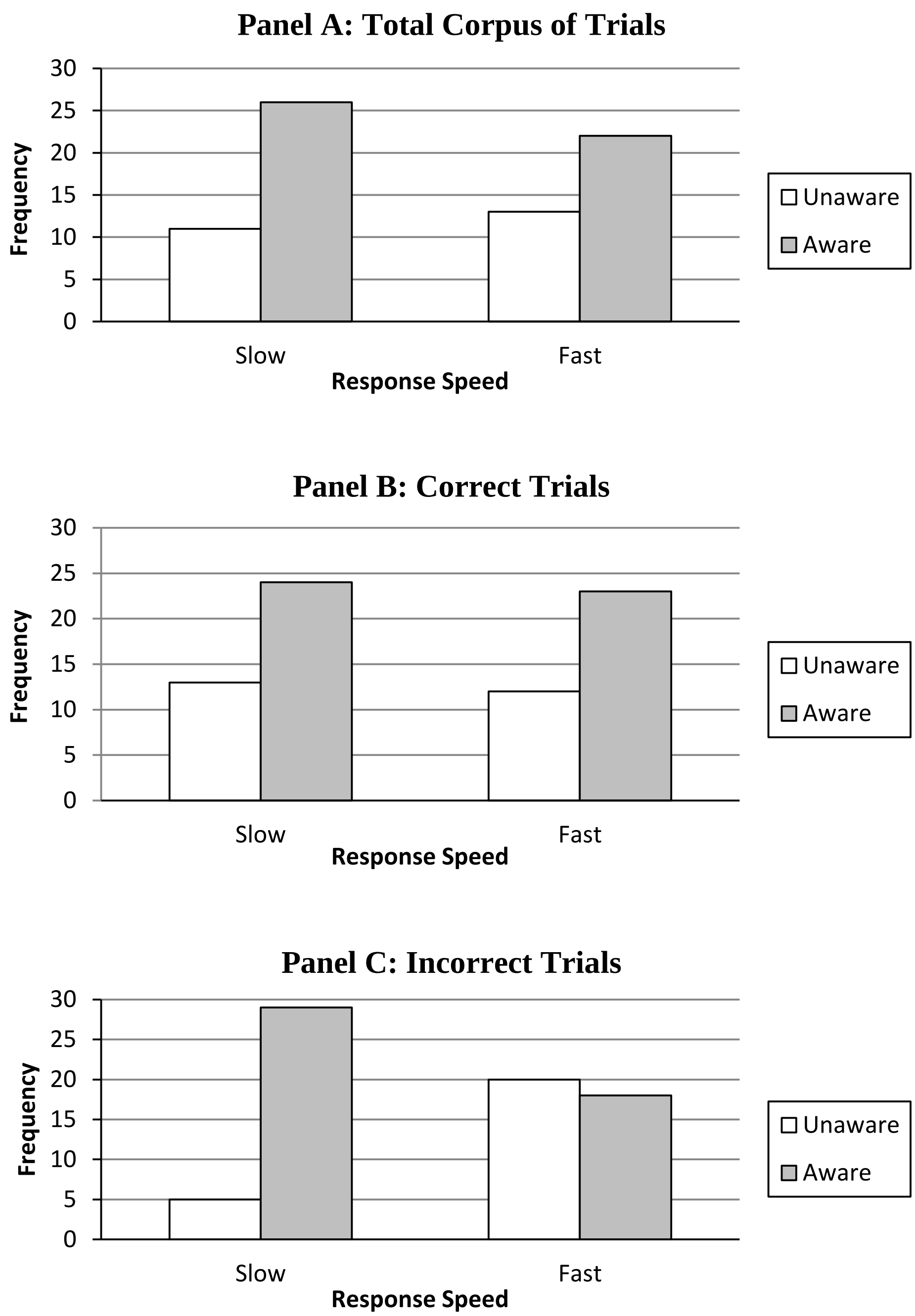
Note. *M* Trials indicates the average number of trials entered into analyses per condition

Supplemental Figure 1



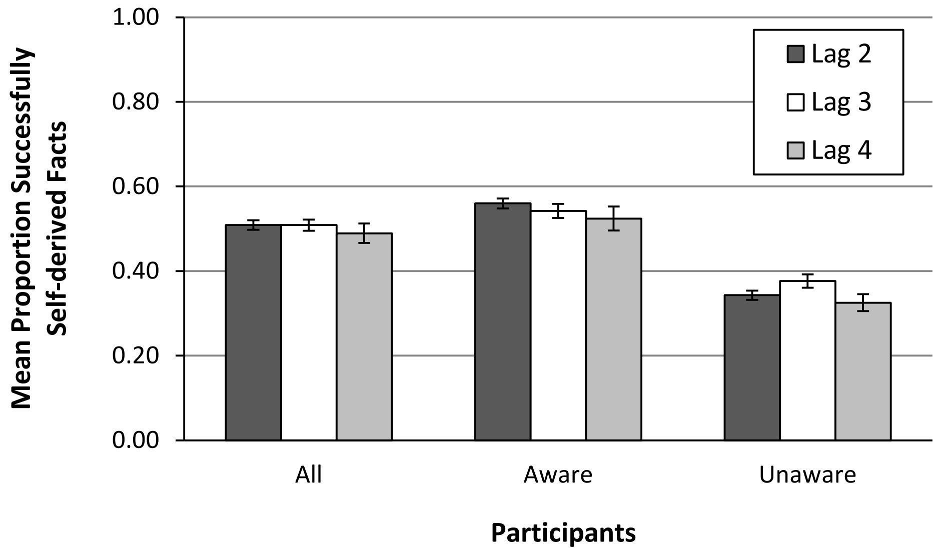
Supplemental Figure 1. Number of participants who were or were not aware of the relational structure of the learning task, categorized as a function of whether they were in the upper or lower half of the self-derivation performance distribution.

Supplemental Figure 2



Supplemental Figure 2. Number of participants who were or were not aware of the relational structure of the task, categorized as a function of whether they were slow or fast responders across all possible trials (Panel A), correct trials (Panel B), and incorrect trials (Panel C).

Supplemental Figure 3



Supplemental Figure 3. Mean proportion of self-derived integration facts by lag at Session 1 (Experiment 2) separately for all participants, those who were aware of the opportunity to integrate, and those who were unaware. Error bars represent standard error of the mean.