Distribution Agreement

Rachael Gilbert | 05/07/2021

___________________________ | ______________________________

Name | Today's Date

Wait A Minute, What Isn't Right Here?
Investigating How We Identify, Learn and Extend True and Untrue Information

by

Rachael E. Gilbert

Dr. Patricia Bauer

Adviser

Department of Psychology

Dr. Patricia Bauer

Adviser

Dr. Joseph Manns

Committee Member

Dr. Eve Mullen

Committee Member

2021

Wait A Minute, What Isn't Right Here?
Investigating How We Identify, Learn and Extend True and Untrue Information

by

Rachael E. Gilbert

Dr. Patricia Bauer

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Psychology

2021

Abstract

Wait A Minute, What Isn't Right Here?
Investigating How We Identify, Learn and Extend True and Untrue Information

By Rachael E. Gilbert

Compiling and extending information is a focus in the literature surrounding the perpetuation of untrue information and the literature that focuses on self-derivation, or the generation, of novel information. Yet, few studies have put these two fields in conversation with one another and focused specifically on the manner in which untrue information is incorporated into an individual's knowledge base and used to draw novel conclusions and inform future decisions. In a study with university undergraduate students, I examined differences in participants' ability to derive new information when they encountered untrue information. I also tested differences in self-derivation performance before and after an opportunity to self-correct when untrue information was consistent as opposed to inconsistent with a participant's prior knowledge. Lower self-derivation performance was present in all instances in which participants encountered untrue information, but performance was lowest in instances where participants encountered information that was both untrue and consistent with their prior knowledge base. These patterns of findings suggest that the presence of untrue information can impair an individual's self-derivation performance, and that inconsistency of untrue information is a key part of how an individual identifies and corrects the untrue information they encounter.

Wait A Minute, What Isn't Right Here?
Investigating How We Identify, Learn and Extend True and Untrue Information

by

Rachael E. Gilbert

Dr. Patricia Bauer

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Psychology

2021

Acknowledgements

Deepest thanks to both my adviser, Dr. Patricia Bauer for supporting me in the completion of this thesis through the Psychology Department. Many thanks also to my in-lab adviser and supervisor, Dr. Jessica Dugan for working tirelessly with me throughout this process. I would also like to take time to thank both my committee members, Dr. Eve Mullen and Dr. Joseph Manns for their participation in my committee and their insightful feedback. I would also like to thank them for their role in supporting and encouraging my intellectual curiosity and academic success throughout my time at Emory University.

Table of Contents

Wait A Minute, What Isn't Right Here?

Investigating How We Identify, Learn and Extend True and Untrue Information

Rachael E. Gilbert

Department of Psychology, Emory University

Abstract

Compiling and extending information is a focus in the literature surrounding the perpetuation of untrue information and the literature that focuses on self-derivation, or the generation, of novel information. Yet, few studies have put these two fields in conversation with one another and focused specifically on the manner in which untrue information is incorporated into an individual's knowledge base and used to draw novel conclusions and inform future decisions. In a study with university undergraduate students, I examined differences in participants' ability to derive new information when they encountered untrue information. I also tested differences in self-derivation performance before and after an opportunity to self-correct when untrue information was consistent as opposed to inconsistent with a participant's prior knowledge. Lower self-derivation performance was present in all instances in which participants encountered untrue information, but performance was lowest in instances where participants encountered information that was both untrue and consistent with their prior knowledge base. These patterns of findings suggest that the presence of untrue information can impair an individual's self-derivation performance, and that inconsistency of untrue information is a key part of how an individual identifies and corrects the untrue information they encounter.

Wait A Minute, What Isn't Right Here?

Investigating How We Identify, Learn and Extend True and Untrue Information

The term "fake news" has been increasingly used over the past several years to describe the phenomenon in which information presented to the public may be inaccurate, biased or misrepresentative of the truth. This phenomenon is a growing worry for public figures, political organizations and individuals alike. This concern is a valid one, and it extends far beyond the political sphere, as we use the information that we learn to make decisions, solve problems and even build and extend our knowledge base. Indeed, in addition to learning new information that is directly presented, humans are also capable of putting separate pieces of information together through memory integration to produce novel information and conclusions, a process known as self-derivation (Bauer & San Souci, 2010; Bauer & Varga, 2017; Varga & Bauer, 2017). Self-derivation is essential to understanding how "fake news," and untrue information in general, are perpetuated and spread. We encounter a multiplicity of information in our daily lives, only some of which is true. Much of the untrue information that we encounter may appear true on its face and is only revealed to be untrue when integrated or compared with other sources of information. In the present research study, I sought to understand how college-aged adults integrate multiple sources of true and untrue information, derive novel conclusions from such information, and then evaluate the truth of the resulting conclusions.

To fully understand the process by which we compare and compile multiple sources of information it is essential to understand self-derivation through integration. Self-derivation is the process by which individuals combine two or more related pieces of information together and produce novel information (Bauer et. al 2012). For example, in a standard self-derivation task, adults are provided with the following related facts: "Apple seeds are called pips" and "Pips contain cyanide." Through the process of self-derivation, participants have demonstrated production of the novel integration fact: "Apple seeds contain cyanide," despite never being directly given that information in the context of the study (Varga & Bauer, 2017). Self-derivation and integration are essential to learning and building a knowledge base, as they allow individuals to extend the knowledge that is presented to them and then draw novel conclusions from the information that they receive.

Prior research suggests that successful self-derivation begins relatively early, around 4 years of age (Bauer & San Souci, 2010), but increases over time, with children demonstrating increased performance on self-derivation tasks as they get older (Bauer & Larkina, 2016). Studies completed with adults also evidence successful self-derivation (Varga & Bauer, 2017). Adults demonstrate a wide range of variability of performance in tasks of self-derivation, with undergraduates at the same institution scoring anywhere between 3-93% percent correct on self-derivation tasks (Varga & Bauer, 2017). Furthermore, variable performance on tasks of self-derivation within a given age range has been associated with an individual's overall reading comprehension, working memory and verbal comprehension (Varga and Bauer, 2017).

Much of the current literature on self-derivation through integration has focused on identifying factors that predict performance on self-derivation tasks in children and adults and understanding the significant differences in performance within age groups. In all cases, the information used as stimuli in the research has been true facts. However, the information that individuals encounter and are tasked with integrating in the world outside the laboratory is not always true. Thus, it is essential to understand the process by which individuals integrate, evaluate, and derive new knowledge when they encounter information that is inconsistent and/or untrue.

A substantial amount of work has been done to understand the process by which we recognize and evaluate information to be untrue (Bottoms & Marsh, 2010; Brashier, Eliseev, & Marsh, 2020; Cantor & Marsh, 2017; Erikson & Mattson, 1981). Prior research has demonstrated that individuals often fail to identify untrue information when first reading statements that contain errors, even when they have previously demonstrated the prior knowledge needed to identify the errors (Bottoms & Marsh, 2010; Brashier, Eliseev, & Marsh, 2020; Cantor & Marsh, 2017). This “knowledge neglect” has been labeled the Moses illusion, so named for one of the stimuli in the first study which examined the phenomenon (Erikson & Mattson, 1981). The untrue question, “How many of each animal did Moses take on the ark?,” is often overlooked by participants, even if they had previously demonstrated knowledge that Noah, rather than Moses, was the Biblical character who took animals on the ark (Bottoms & Marsh, 2010; Cantor & Marsh, 2017).

Additional research has also demonstrated that certain factors, including whether participants are “experts” in the stimuli (i.e., they have studied that field at a university level) predicts identification of untrue information (Cantor & Marsh, 2017). This research has indicated that individuals are more likely to identify, or “flag,” information as untrue when they are “experts” in the subject that is being assessed. Despite this increase in identification of untrue information, expertise does not eliminate knowledge neglect altogether, and “experts” still routinely overlook untrue information (Cantor & Marsh, 2017). Further, research surrounding knowledge neglect has also indicated that participants are more likely to identify errors when directly asked to evaluate the truth of the passages and/or statements that they read, as opposed to when they are not directly asked to evaluate the truth of the passages/statements that they read (Brashier, Eliseev, & Marsh, 2020; Cantor & Marsh, 2017). This research highlights essential questions both about how individuals evaluate the information that is presented to them and how they evaluate the sources from which they receive information. Though studies like these have begun to help us understand the process by which we evaluate untrue information, there are still significant gaps in our understanding. Specifically, it is not yet known how individuals integrate information that is untrue, and likely incongruent with prior knowledge, with true information to derive new knowledge. This focus is of value as it more clearly reflects how individuals would encounter untrue information and "fake news" in their daily lives. We do not encounter untrue information "in a vacuum." Rather, we are presented with a steady stream of true and untrue information on a daily basis that we encode, evaluate and integrate into our knowledge base, and further, use to derive novel understandings.

Through the current study, I seek to answer three main questions presented by the current status of the literature. First, I seek to extend the current research on self-derivation and investigate whether participants are capable of self-derivation with untrue information. Second, if there is evidence of self-derivation for untrue information, I seek to understand how self-derivation performance differs based on variations in the timing and consistency of untrue information presented. Last, I seek to understand how individuals evaluate the truth of the integration statements they generate, and if these evaluations differ based upon the timing and consistency of the untrue information that was presented.

In the current study, participants encountered separate but related pairs of facts, which, when integrated, would produce a novel integration fact that participants were not directly presented. The pairs of facts that were presented to each participant contained a combination of true and untrue information. Some of the untrue information was consistent with the prior knowledge that participants had been presented, while other pieces of untrue information were inconsistent with the prior knowledge that they had been presented. Participants were then asked to self-derive novel information from the paired facts in both open-ended and forced choice formats. Afterwards they were asked to evaluate how "true" they viewed the statement they just generated to be. After completing the first task, participants were then presented with the corrected versions of all the untrue information that they had encountered throughout the course of the study. Afterwards, they completed a second task in which they were once again asked to self-derive and evaluate the truth of the statements that they generated.

In line with previous studies, I predict that participants will be more likely to flag information as untrue if it is inconsistent with the prior knowledge that they have received as opposed to when it is untrue, but consistent information. Therefore, if a participant does not encounter inconsistent information during the encoding period for untrue information, then they will be less likely to identify the self-derived integration fact as untrue than in conditions when they receive inconsistent information. I also predict if a participant encounters inconsistent and untrue information, then they will be more likely to correct themselves at a later opportunity than when they encounter untrue, consistent information. I also hypothesize that within the inconsistent condition, if the participant encounters the inconsistent information after being presented with the two true paired facts, then they will be more likely to demonstrate successful self-derivation than if they encountered inconsistent information in between the two paired facts. Extending this, I also hypothesize that if an individual encounters inconsistent information in between the two true paired facts, as opposed to afterwards, then they will be more likely to "flag" the integration fact that they produce as untrue.

Method

Participants

In total, 60 participants were recruited from a pool of available individuals within SONA. A total of 15 participants were excluded because they did not complete all parts of the study, meaning that the final sample size was 45. All participants were current Emory undergraduate students enrolled in a class/program that required participation in research for credit. 60% of participants identified as female, 35.56% of participants identified as male, 2.22% of participants identified as nonbinary, and 2.22% of participants did not provide information regarding gender identity. 42.22% of participants identified as Caucasian/White, 4.44% identified as African American/Black, 48.89% identified as Asian and 4.44% identified as Latinx. Data were also collected on class standing: 42.22% of participants were freshman, 31.11% of participants were sophomores, 20% were juniors and 6.67% were seniors. Participants were also asked to identify their primary area of study, 42.22% identified themselves as STEM majors (Biology, Chemistry, Physics, Economics, Neuroscience and Behavior Biology, Computer Science, Environmental Science, Human Health), 20% identified themselves as business/finance majors (Currently (or planning to be) enrolled in the business school), 33.33% identified as social science majors (Psychology, Sociology, Anthropology, Political Science, Philosophy Politics and Law, Linguistics), 2.22% identified as liberal arts/fine arts majors (Art History, History, Religious Studies, Classics, English, Creative Writing, Philosophy, Dance, Film, Visual Arts, Music, Theatre, Women and Gender Studies), and 2.22% did not answer demographic questions regarding major. This information was collected for descriptive purposes only and was not included in analysis.

The sample size was selected to fit the time restraints of data analysis, as each participant produced a significant number of data points and ensuring that each of these points was analyzed in depth was determined to be a greater priority than generating a large number of participants. Participants were awarded 3 credits for their participation, which they were required to accumulate over the course of their enrollment in an introductory psychology course. All procedures were reviewed and approved by the school board of the university Institutional Review Board.

Materials

Stimuli

Prior Knowledge Stimuli

One hundred and two facts were developed to establish a base of prior knowledge from which participants could draw in subsequent phases. Prior knowledge facts were divided into three categories: relevant prior knowledge, irrelevant but related prior knowledge, and irrelevant and unrelated prior knowledge. Relevant prior knowledge consisted of information that could be used to generate the integration statements which would be assessed later in open ended and forced-choice formats (e.g., prior knowledge about Picasso when participants would later be asked to self-derive a fact about Picasso). Irrelevant but related prior knowledge consisted of items that would not appear on subsequent open-ended or forced-choice assessments but were related to the subject(s) that would be assessed (e.g., prior knowledge about art history when participants would later be asked to self-derive a fact about Picasso). Irrelevant and unrelated prior knowledge consisted of items that would not appear on subsequent open-ended or forced choice assessments and were not related to the subject(s) being assessed (e.g., prior knowledge about whales when participants would later be asked to self-derive a fact about Picasso). These were included to ensure that participants were receiving three total facts surrounding the subject of each integration fact across all conditions.

Encoding Phase 1

Encoding Phase 1 consisted of eighty stem facts (forty pairs), hereafter called stem facts. Pairs of facts were divided into three categories: Consistent-True facts (20 stem facts, 10 pairs), Consistent-Untrue facts (12 stem facts, 6 pairs), and Inconsistent facts (48 stem facts, 24 pairs). Consistent-True pairs contained two facts that were correct (e.g., the two stem facts "Alexander Graham Bell was an inventor from Scotland" and "Alexander Graham Bell invented the telephone" are both true). Consistent-Untrue pairs contained one fact that was incorrect but was consistent with other knowledge presented to them in the context of the study (e.g., the stem fact "A softer type of body hair is called lanugo" is true, the stem fact "Lanugo is common in people with diabetes" is untrue, and the untrue fact is consistent with the prior knowledge "A softer type of body hair is called lanugo"). Inconsistent pairs contained one fact that was untrue that was also inconsistent with information provided to them in Prior Knowledge Building (e.g., the stem fact "Duchamp's most well-known sculpture is called 'Fountain'" is true, the stem fact "A popular sculpture made from a shower is called 'Fountain'" is untrue, and the untrue fact is inconsistent with the prior knowledge "A popular sculpture made from a urinal is called 'Fountain'"). Differences in the number of pairs within each condition was due to original stimuli being thrown out because they could not be adapted to the current study.

Encoding Phase 2

Encoding Phase 2 consisted of eighty facts (forty pairs), all of which were correct. That is, the Consistent-True facts were repeated, and the facts given in the Consistent-Untrue and Inconsistent categories in Encoding Phase 1 were corrected.

Procedure

Participants were recruited using an online participant-base called SONA. Participants were told that the study would take approximately ninety minutes and would be presented in two sessions approximately 24 hours apart from one another. Participants were asked to read a consent statement prior to participation, and digitally signed a copy in order to proceed. Participants also completed a demographic survey at the time of registration (see Appendix A). To be included in data analysis, participants were required to complete all parts of the study. Prior Knowledge was presented one calendar day (roughly 24 hours) prior to the administration of the rest of the study. Participants were sent a link to a Qualtrics survey which presented all 102 Prior Knowledge items one at a time. These 102 items were counterbalanced into four different randomized versions of Prior Knowledge. Participants were told to read each item carefully, and that once they clicked “next” they would not be allowed to go back to review previous facts. Data were collected on time spent reading each statement, as well as overall time spent studying all materials. Participants were then told that they would receive another link to a second Qualtrics survey the following day, and that the link would only work for one twenty-four-hour period.

The following day, participants first were presented Encoding Phase 1, in which they were presented each stem fact one at a time. These items were also counterbalanced into two different versions. In each version, paired stem facts were presented within nine pages of one another but were never directly presented one after the other. They were then instructed to take a fifteen-minute break and return to the session, where they completed the Open-Ended Phase 1 and Forced-Choice Phase 1 sections. In Open-Ended Phase 1, participants encountered forty fill-in the blank statements which required them to self-derive new knowledge by integrating the two stem facts they had encountered earlier. After each fill-in the blank, participants were also asked to assess the truthfulness of the statement they just completed on a scale from 1 to 100 (1 being not at all true, and 100 being entirely true). If participants ranked the truthfulness of the statement at 50 or below, they were then asked to supply a short reasoning for why they viewed the statement as untrue in a free-response format.

Forced Choice Phase 1 consisted of forty multiple-choice questions assessing successful self-derivation. Each forced choice question provided four multiple choice answers from which participants could choose: correct self-derivation, incorrect self-derivation, incorrect but related and unknown (see Appendix B for sample). Correct self-derivation answers were the true statements that were the integrated version of the stem facts presented to them in Prior Knowledge Building and Encoding Phase 1. Incorrect self-derivation answers were the integrated version of the incorrect stem facts that were presented to participants at Encoding Phase 1, incorrect but related answers were answers that were not presented at Encoding Phase 1 but held high surface similarity to the subject of each question, and unknown was an option for participants to respond, “I don’t know”. After each forced-choice page participants were asked to assess the truthfulness of the statement they just selected on a scale from 1 to 100 (1 being not at all true, and 100 being entirely true). As in the open-ended phase, if participants ranked the truthfulness of the statement at 50 or below, they were then asked to supply a short reasoning for why they viewed the statement as untrue in a free-response format.

Each of these sections presented one question at a time and asked participants to evaluate their answers on separate pages. Participants were required to enter some value in all open-ended and forced choice questions in order to proceed.

Participants were then instructed to take another fifteen-minute break and return to the session. After the break, they were presented Encoding Phase 2, which again presented each stem fact one at a time. They were instructed to take a final fifteen-minute break and return to the session for Open-Ended Phase 2 and Forced-Choice Phase 2. Like Open-Ended Phase 1, Open-

Ended Phase 2 consisted of forty fill-in the blank statements which assessed an individual's self-derivation performance. Participants were once again asked to evaluate the truth of the statements they completed on a scale of 1 to 100 and to provide reasoning if they evaluated the statement at 50 or below. Each section presented one question at a time and asked participants to evaluate their answers on separate pages. Forced-Choice Phase 2 was structured in the same manner as Forced-Choice Phase 1. Finally, participants were debriefed on the purpose of the study and given the contact information of the primary researcher if they had any further questions.

Data Reduction and Data Analytic Plan

Participants' responses were graded as either correct self-derivation (1) or incorrect (0). Incorrect responses were also graded on the dimension of incorrect self-derivation (1) or incorrect non-self-derivation (0). Incorrect self-derivation meant that the participant demonstrated self-derivation using two stem facts given to them in the context of the study, but one of the stem facts they used was the incorrect information they had been provided in Encoding Phase 1. They therefore showed evidence of self-derivation, even though they arrived at the incorrect answer.

Responses were separated into the three conditions as they were presented in Encoding Phase 1 identified above: Consistent-True, Consistent-Untrue and Inconsistent. Though the information supplied in Encoding Phase 2 was all true, answers were still divided to reflect their condition in Encoding Phase 1 in order to determine the difference between changes in responses from Open-Ended and Forced Choice Phase 1 to Open-Ended and Forced Choice Phase 2.

Evaluations of Truth were also divided along these conditions, and the overall evaluations of each different condition for participants were summed and described. Evaluations of Truth between the Inconsistent condition were compared to Evaluations of Truth in the Consistent-Untrue condition and the Consistent-True conditions.

Within the Inconsistent condition, responses were further separated into two categories to assess whether the timing of the presentation of the inconsistent stem fact impacted successful self-derivation performance. On some of the trials, the first stem fact was correct and the second was inconsistent. On other trials, the first stem fact was inconsistent and the second was correct (see Figure 1).

Participants' answers in Open-Ended and Forced Choice Phase 2 were also compared between conditions, to determine whether there was a significant difference between the re-learning of untrue information that they had initially been presented inconsistent information on, as opposed to untrue information that they had not initially been presented inconsistent information on. Participants' likelihood of answering with the incorrect integration statement was also compared between phases and conditions.

Results

The results are presented in two main sections. In the first section I describe self-derivation performance by phase. Within this section, the review of the results follows the chronological order of each phase. Forced Choice results are only reported when they deviate from their Open-Ended counterparts. In the second section, I describe differences in Evaluations of Truth, these results are also reported in the chronological order of each phase.

Within each phase results are reported first by analyzing the differences in correct self-derivation performance across the three conditions: Consistent-True, Consistent-Untrue and Inconsistent. In the Consistent-True condition all information provided to the participant was completely correct and was consistent with itself. In the Consistent-Untrue condition, participants were provided information at both Prior Knowledge Building and Encoding Phase 1; one of the facts provided at Encoding Phase 1 was untrue but was not inconsistent with any of the information provided at Prior Knowledge Building. In the Inconsistent condition, participants were provided information at both Prior Knowledge Building and Encoding Phase 1; one of the pieces of information provided at Encoding Phase 1 was untrue and inconsistent with the knowledge presented at Prior Knowledge Building.

Next, further analyses between the Consistent-Untrue and Inconsistent conditions are presented. These analyses focused on potential between-condition differences on trials in which participants self-derived a novel fact but used the untrue stem fact to do so. Therefore, though the answer they ultimately produced was incorrect, it is still of interest because it is indicative of self-derivation. Finally, analyses within the Inconsistent condition are reported, focusing on differences in successful self-derivation performance and incorrect self-derivation performance based upon when the inconsistent information was presented.

Self-Derivation Performance

Phase 1

Consistent-True, Consistent-Untrue and Inconsistent Successful Integration

The focus of the analysis was to compare successful self-derivation performance in conditions in which participants had received no inconsistent or untrue information (Consistent-True) and inconsistent and/or untrue information (Consistent-Untrue, Inconsistent). Figure 2 depicts average performance by condition for each phase. A one-way ANOVA revealed a difference in performance across conditions in Open Ended Phase 1 $F(2, 43) = 39.34$, $p < 0.001$. A Tukey HSD revealed that participants were significantly more likely to successfully self-derive in the Consistent-True condition ($M = 29.03$, $SD = 17.77$) than they were in the

Inconsistent condition (*M* = 22.12, *SD* = 11.41), $p < 0.001$, which in turn was also higher than the Consistent-Untrue condition (*M* = 4.44, *SD* = 10.30), $p < 0.001$. In other words, successful self-derivation performance was negatively impacted by the administration of untrue information in general. Performance was especially negatively impacted when participants were presented untrue but consistent information.

In Forced Choice Phase 1, a one-way ANOVA revealed a difference in successful self-derivation performance across condition $F(2, 43) = 25.75$, $p < 0.001$. In departure from open-ended performance, a Tukey HSD revealed participants were not significantly more likely to successfully self-derive in the Consistent-True condition (*M* = 58.51, *SD* = 22.72) than in the Inconsistent condition (*M* = 50.14, *SD* = 14.64), $p = 0.13$.

Inconsistent and Consistent-Untrue Incorrect Self-Derivation Performance

The above analysis focused specifically on the generation of the factually correct integration fact. However, because individuals were provided untrue information in the context of the study, it is also important to investigate potential differences in the self-derivation of the incorrect integration fact. Figure 3 depicts average performance by condition for each phase. A paired *t*-test was conducted to determine whether incorrect self-derivation performance differed significantly across conditions. In Open-Ended Phase 1 participants generated the incorrect integration statement more frequently in the Consistent-Untrue condition (*M* = 32.70, *SD* = 22.58) than in the Inconsistent condition (*M* = 20.54, *SD* = 13.86), $t(44) = 4.97$, $p < 0.001$.

Inconsistent Information Timing: Successful Self-Derivation Performance

In the inconsistent condition on some of the trials, the first fact of a pair was true and the second was untrue, and in some trials the order was reversed. Figure 4 depicts average performance by condition for each phase. In Open-Ended Phase 1, participants were more likely to successfully self-derive when they encountered inconsistent information first (M = 27.04, SD = 16.62) as opposed to second (M = 14.78, SD = 14.18), $t(44) = -3.77$, p = 0.0015. The same pattern was observed in forced choice testing.

As well, when we considered incorrect self-derivation performance, the same pattern was observed. That is, participants were significantly more likely to produce the incorrect self-derivation statement when they encountered inconsistent information first (M = 23.83, SD = 19.48) versus second (M = 17.04, SD = 15.69), $t(44) = -2.06$, p = 0.045. Figure 5 depicts average incorrect self-derivation performance by condition for each phase.

Incorrect self-derivation performance did not differ significantly in Forced Choice Phase 1. In Forced Choice Phase 1, participants were not significantly more likely to incorrectly self-derive when they encountered inconsistent information first (M = 60.07, SD = 29.36) versus second (M = 50.62, SD = 34.00), $t(42) = -1.60$, p = 0.12.

Phase 2

Consistent-True, Consistent-Untrue and Inconsistent Successful Integration

The purpose of Phase 2 was to determine whether participants would correct their knowledge base when presented with correct information. As outlined in the Method, participants re-encountered the Consistent-True facts that they had originally encountered in Encoding Phase 1. They were also presented with the corrected version of all facts in Encoding Phase 2, both in the Consistent-Untrue and the Inconsistent condition. To determine whether there were differences in self-derivation, a one-way ANOVA was conducted, which revealed a difference in performance across condition for both Open Ended Phase 2 $F(2, 43) = 3.675$, p = 0.028, and Forced Choice Phase 2 $F(2, 39) = 4.765$, p = 0.0102, consistent with results from Phase 1.

For open-ended testing, a Tukey HSD revealed that performance between the Consistent-True condition (M = 64.14, SD = 20.45) and the Inconsistent condition (M = 54.55, SD = 25.25), was not significantly different p = 0.92. Moreover, Consistent-True was not different from Consistent-Untrue, p = 0.089. This suggests that individuals corrected their knowledge base, as their performance in the Inconsistent and Consistent-Untrue conditions no longer differed significantly from performance in Consistent-True. However, performance was different from Inconsistent (M = 54.55, SD = 25.25), to Consistent-Untrue (M = 54.55, SD = 25.25), p = 0.034. Participants performed relatively better in the Inconsistent condition than the Consistent-Untrue condition.

Four participants were excluded from analysis in Forced Choice Phase 2 because, as outlined by the Method, they were not presented with any stimuli in the Consistent-True condition as they had answered these statements correctly in the previous phase. Like open-ended, analyses indicated that participants were not significantly more likely to successfully self-derive in the Consistent-True condition (M = 59.79, SD = 31.70) than the Inconsistent condition (M = 51.95, SD = 20.59), p = .46. However, participants were significantly more likely to successfully self-derive in the Consistent-True condition (M = 59.79, SD = 31.70) than the Consistent-Untrue condition (M = 39.67, SD = 34.98), p = 0.008. This suggests that participants did not correct themselves in the Consistent-Untrue condition.

Inconsistent and Consistent-Untrue Incorrect Self-Derivation Performance

In Open-Ended Phase 2, participants incorrectly self-derived more frequently in the Consistent-Untrue condition (M = 43.42, SD = 32.65) than in the Inconsistent condition (M = 26.82, SD = 17.58), $t(39) = 2.75$, p = 0.009. The same pattern was observed in Forced Choice Phase 2.

Inconsistent Information Timing: Successful Self-Derivation Performance

One participant was removed from analysis as they did not enter identify themselves in Encoding Phase 1, and therefore the order in which they encountered stimuli was unable to be determined. In Open-Ended Phase 2 participants were more likely to successfully self-derive when they encountered inconsistent information first (M = 57.39, SD = 26.24) as opposed to when they encountered inconsistent information second (M = 43.73, SD = 26.24), $t(43) = -2.15$, $p = 0.037$.

Performance between conditions did not differ significantly in Forced Choice Phase 2, and participants were no more likely to successfully self-derive when they first encountered inconsistent information (M = 51.41, SD = 19.23) as opposed to second (M = 43.73, SD = 35.54), $t(44) = -0.36$, $p = 0.72$. As well, incorrect self-derivation performance did not differ significantly across condition in either Open-Ended Phase 2 or Forced Choice Phase 2.

Evaluations of Truth

Phase 1

Consistent-True, Consistent-Untrue and Inconsistent Evaluations of Truth

Participant's Evaluations of Truth (EoTs), an assessment of a statement's truthfulness, were also collected. Four participants were removed from Evaluations of Truth analysis because they did not complete the task as specified. Figure 6 depicts average EoT by condition for each phase.

EoTs differed significantly only in Open Ended Phase 1 $F(2, 39) = 4.825$, $p = 0.0095$. A Tukey HSD revealed that participants evaluated Consistent-True (M = 68.30, SD = 14.39) and Consistent-Untrue (M = 67.43, SD = 15.40) statements as significantly more truthful than the inconsistent statements (M = 59.56, SD = 14.33), p = 0.015, 0.033. Evaluations between the

Consistent-True condition (*M* = 68.30, *SD* = 14.39) and Consistent-Untrue condition (*M* = 67.43, *SD* = 15.40) were not significantly different, *p* = 0.96. EoTs did not differ significantly in Forced Choice Phase 1 $F(2, 39) = 1.236$, $p = 0.294$.

Inconsistent Information Timing: Evaluations of Truth

EoTs differed significantly between conditions in all phases based on timing of inconsistent information. Participants evaluated statements when they first received inconsistent information as significantly more truthful in Open Ended Phase 1 (*M* = 61.88, *SD* = 14.51) than when they received the inconsistent information second (*M* = 55.58, *SD* = 18.78), $t(44) = -2.83$, $p = .007$. This same pattern was observed in forced choice testing. Figure 7 depicts average EoT by condition for each phase.

Phase 2

Consistent-True, Consistent-Untrue and Inconsistent Evaluations of Truth

Like Forced Choice Phase 1, EoTs did not differ significantly in Open-Ended Phase 2 $F(2, 39) = 0.377$, $p = 0.686$ or Forced Choice Phase 2 $F(2, 35) = 0.569$, $p = 0.568$.

Inconsistent Information Timing: Evaluations of Truth

Similar to Phase 1, in Open Ended Phase 2 participants evaluated statements when they encountered inconsistent information first as significantly more truthful in Open Ended Phase 2 (*M* = 77.52, *SD* = 11.59) than when they encountered inconsistent information second (*M* = 67.00, *SD* = 17.16), $t(44) = -6.58$, $p < 0.001$. The same pattern was observed in Forced Choice Phase 2.

Discussion

In the present research, I aimed to answer three primary questions. The first was to investigate self-derivation when participants were given untrue information. The second was to assess whether the timing and consistency of untrue information impacted self-derivation performance. Last, I aimed to understand how a participant evaluated the truth of the statements they derived and if these evaluations differed based on the timing and consistency of untrue information that they encountered. These are all essential questions to understand as they shed light on the manner in which we encounter and work with information outside of the laboratory. The information that we come across in our daily lives is not always true or consistent with our knowledge base. Understanding how we integrate, assimilate, and extend new, potentially untrue and inconsistent, information is essential. This understanding would allow us to more deeply comprehend how individuals learn, as well as how untrue information is perpetuated and spread.

I first examined differences in successful self-derivation across the three conditions. In all conditions, participants were provided with prior knowledge regarding a given subject. They later learned pairs of facts related to that prior knowledge. Some pairs of facts contained information that was untrue. Within these untrue facts, some untrue information was consistent with the prior knowledge provided to the participant, while some was inconsistent.

Self-derivation performance was highest on trials in which participants encountered only true and consistent information on a given subject (Consistent-True). Performance in conditions in which participants encountered untrue information, whether it be consistent or inconsistent was lower. Performance was lowest in conditions in which participants encountered untrue and consistent information.

These results make clear that participants self-derive even when the information they learned was untrue, both in instances in which participants encountered consistent and inconsistent untrue information. I also observed differences between performance when the untrue information a participant encountered was consistent with their knowledge base as opposed to inconsistent. Participants performed better in the Inconsistent condition than they did in the Consistent-Untrue condition. For Phase 1 of the study, it is understandable that performance between these two conditions was different because in the Consistent-Untrue condition participants were not given the information needed to produce the correct integration fact in the context of the study. Instead, participants in this condition could demonstrate self-derivation using the two stem facts provided to them about a given subject, one of which was untrue. For Encoding Phase 1 when participants encountered untrue and consistent information the generation of the incorrect version of the integration statement is therefore comparable to correct self-derivation performance when participants only received true and consistent information. From these findings, I claim that untrue and true information are not necessarily treated differently by the participant at a fundamental level. That is, unless a participant encounters inconsistency at encoding, they are likely to self-derive using the information provided to them, whether that information be objectively true or untrue.

In addition to examining differences in how the consistency of untrue information presented affected self-derivation performance, I was also interested in investigating how the timing of inconsistent information affected performance. To investigate this, we varied the order in which participants were presented inconsistent untrue information. Contradictory to my initial hypothesis, I found that individuals were significantly more likely to successfully self-derive when they first encountered the inconsistent fact than when they encountered the inconsistent fact second.

Though different from our initial expectations on how individuals integrate untrue information, this finding could be explained by the fact that in the first condition, participants encountered the inconsistent information on each subject closer in time. This may have made it easier to recognize that the statements were, in fact, inconsistent and allowed the participant to more explicitly decide which of the two inconsistent stem facts they “trusted” to incorporate into their knowledge base. Because of the nature of the current study, I only asked participants to evaluate the truth of the integration statements that they produced and did not ask them to evaluate any of the stem facts that they encountered. For that reason, I cannot currently draw definitive conclusions about potential differences in this trust and evaluation of stem facts based on different timings of inconsistent information.

I also investigated how participants responded when they were presented with the corrected versions of the untrue information that they had initially encountered, and if this self-correction varied based upon condition. This was done in order to determine whether individuals corrected their knowledge, and then further used that corrected knowledge to self-derive. I found that individual’s performance in tasks of self-derivation when they received untrue and inconsistent information (Inconsistent) was no longer significantly different from instances in which they had never received untrue information (Consistent-True). However, this phenomenon was not consistently observed when participants encountered untrue and consistent information (Consistent-Untrue). This could suggest that individuals can correct the untrue information they receive when given the opportunity to do so, but that this correction differs based upon the initial consistency of untrue information that an individual encounters. Inconsistency at encoding appears to signal to participants in some way that the information they are encountering is untrue and needs to be corrected. In contrast, consistency may prevent individuals from identifying information as untrue, and therefore may prevent them from correcting their knowledge base.

These findings, which implicate the importance of inconsistency in self-correction, may shed additional light on the origins and perpetuation of the Moses illusion (Bottoms & Marsh, 2010, Erikson & Mattson, 1981). Throughout the work of Bottoms and Marsh, the incorrect information that participants overlook has high "surface similarity" to the correct information. In the question "How many of each animal did Moses bring on the ark?," Moses has high surface similarity to the correct answer (Noah), as both are Biblical characters present in the Old Testament. My findings on inconsistency may be another way of understanding "surface similarity." That is, because Moses is consistent with one's expectations of the correct name, participants overlook this error more frequently than they would if Noah's name was replaced with one that was inconsistent with their expectations.

In general, the findings regarding differences in both successful and incorrect self-derivation performance based on the consistency of untrue information presented to the participant are particularly interesting because, with the exception of Open-Ended Phase 1, individuals did not evaluate the truth of the statements they generated differently. These findings are, again, similar to that of Bottoms and Marsh (2010) on the subject of the Moses Illusion, which suggests that individuals often overlook untrue information, even after having indicated possessing the prior knowledge needed to identify those statements as untrue. However, because there were significant differences between successful and incorrect self-derivation by condition, the reported findings on Evaluations of Truth are puzzling. Further research to investigate why these differences between conditions exist are necessary.

In summary, I examined differences in the manner in which individuals identified and incorporated information that was consistent with their prior knowledge and true, consistent with their prior knowledge and untrue and inconsistent with their prior knowledge and untrue into their knowledge base. I found evidence that there were substantial differences between these conditions when participants were first asked to self-derive. However, I found that, after being given the opportunity to correct the untrue information they had encountered, participants were capable of doing so. This correction was dependent on condition, as participants were significantly more likely to correct their knowledge base when they encountered information that was untrue and inconsistent, but not information that was untrue and consistent. Thus, I have generated initial evidence that identifying untrue information that one encounters is based on comparison to one's established knowledge base.

I also found evidence that the order in which an individual encountered untrue and inconsistent information had a strong impact on the individual's ability to successfully self-derive. These findings are potentially indicative of deeper differences in the timing of how and when individuals process and compare true and untrue information. However, to get a clearer picture of these differences, Evaluations of Truth should be collected at time of encoding.

Limitations and Directions for Future Research

The present research is not without limitations. Because of the nature of the study design, the diversity of the participant pool was limited, which is essential to consider as we think about extending our findings beyond the current sample. As outlined in the Method, all participants were current undergraduate students at Emory University. This significantly limits the diversity of the sample, and further research that includes different age groups and levels of education is necessary.

Outside of demographic considerations, the study was also conducted online via Zoom per COVID-19 restrictions outlined by the university. In these conditions, it proved more difficult to monitor participant's level of engagement with the task, as well as limit their access to sources of information outside of the study (i.e., looking up answers online). Because of these significantly different conditions from previous studies which have assessed participant's ability to successfully self-derive, performance may in some way have been affected. Reproducing these results in an in-person setting in which the environment can be better controlled would be a necessary next step in extending and strengthening our findings.

The study was also limited in time, and therefore scope of analysis. Additional data were collected that I was unable to analyze due to time constraints.

Further, because previous research indicated that early assessments of truth could “alert” the participants of the untrue nature of some of the information being presented in the context of the study (Brashier et al., 2020), I made the decision to only ask participants to make Evaluations of Truth at integration. However, given my findings that the order in which participants encounter inconsistent information is significant, additional research which includes assessments of truth at Encoding Phase 1 and 2 could be enlightening.

Conclusion

In conclusion, the present research provides insight into how individuals assess, encode and incorporate untrue information into their knowledge base. Most significantly, the present research reflects the notion that our ability to identify untrue information relies heavily on comparison of new information to our established knowledge base. This is the first study that extends research on the identification and reproduction of untrue information to self-derivation and integration and represents a meaningful step in understanding how and why untrue information is spread.

These findings may help us explain why untrue information is perpetuated amongst different groups of people and within individuals, even if they have the true information ultimately available to them. There is a wealth of information available to us from a multitude of different sources, some of which is true and some that is not. However, these pieces of information frequently are consistent with themselves, and only found to be inconsistent when compared with one another. A good example of this is news outlets. Though different news outlets may report events inconsistently from one another, information presented within a single news article or source is frequently consistent with itself. If an individual consumes information from a single source and does not encounter inconsistent information until a later period or date, they may be less likely to recognize the statements as inconsistent, and therefore may accept the information that they initially encountered as true. They therefore will continue to use the untrue information throughout their daily lives to make evaluations and draw conclusions.

References

Bauer, P. J., King, J. E., Larkina, M., Varga, N. L., & White, E. A. (2012). Characters and clues: Factors affecting children's extension of knowledge through integration of separate episodes. *Journal of Experimental Child Psychology*, *111*(4), 681–694. https://doi.org/10.1016/j.jecp.2011.10.005

Bauer, P. J., & Souci, P. S. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of Experimental Child Psychology*, *107*(4), 452–465. https://doi.org/10.1016/j.jecp.2010.05.012

Bauer, P. J., & Larkina, M. (2016). Predicting remembering and forgetting of autobiographical memories in children and adults: A 4-year prospective study. *Memory (Hove, England)*, *24*(10), 1345–1368. PubMed. https://doi.org/10.1080/09658211.2015.1110595

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. https://doi.org/10.1016/j.jarmac.2017.01.006

Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory*, *18*, 670–678. doi: 10.1080/09658211.2010.501558

Cantor, A. D., & Marsh, E. J. (2017). Expertise effects in the Moses illusion: Detecting contradictions with stored knowledge. *Memory*, *25*(2), 220–230. https://doi.org/10.1080/09658211.2016.1152377

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning & Verbal Behavior*, *20*, 540–551. doi: 10.1016/S0022-5371(81)90165-1

Varga, N. L., & Bauer, P. J. (2017). Young adults self-derive and retain new factual knowledge through memory integration. *Memory & Cognition*, *45*(6), 1014-1027.

Figure 1

Inconsistent Fact Placement

First	Second
Prior Knowledge Given: *Hematopoiesis is the cellular formation of blood*	***Prior Knowledge Given:*** *Modern day Belgium is the location of Waterloo*
Stem Fact 1 (Inconsistent): *Hematopoiesis is the cellular formation of neurons*	***Stem Fact 1:*** *Napoleon's final defeat was at Waterloo*
Stem Fact 2: *Hematopoiesis is a process that occurs in the skeleton*	***Stem Fact 2 (Inconsistent):*** *Modern day Ireland is the location of Waterloo*
Integration Fact: *The skeleton is the site of the production of blood*	***Integration Fact:*** *Napoleon's final defeat was in Belgium*

Figure 2
Successful Self-Derivation Performance by Condition

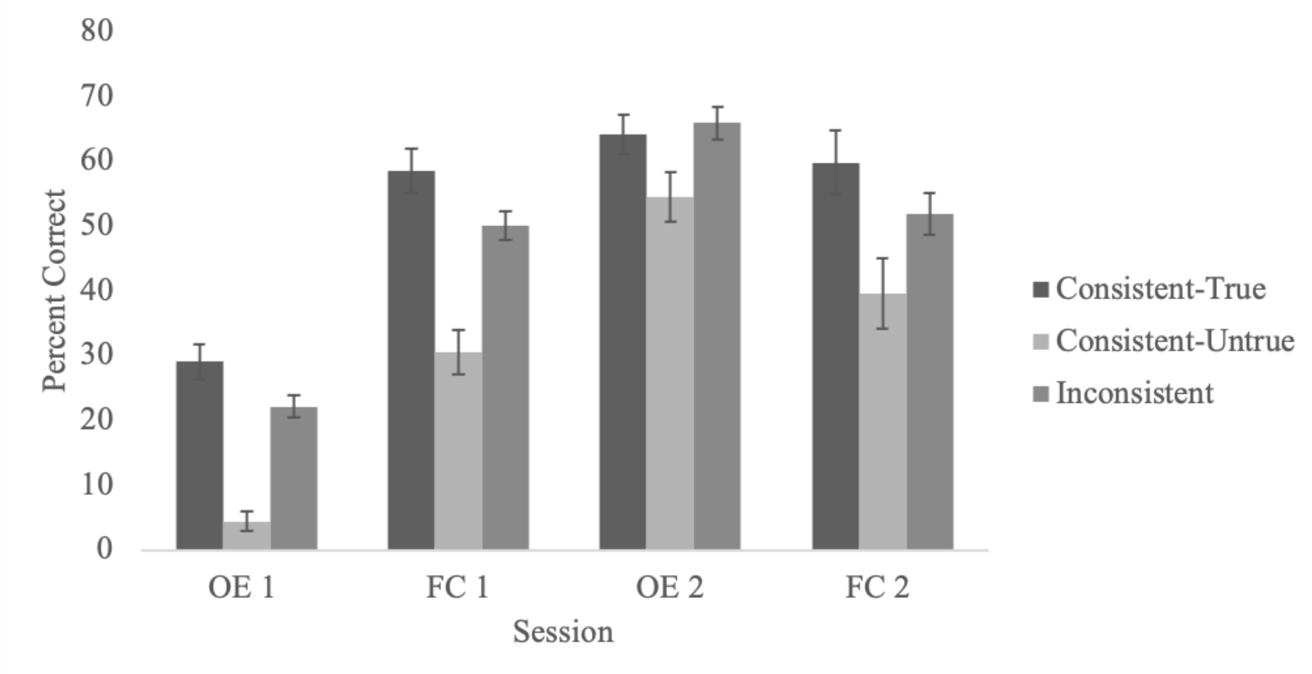


Figure 3

Incorrect Self-Derivation Performance by Condition

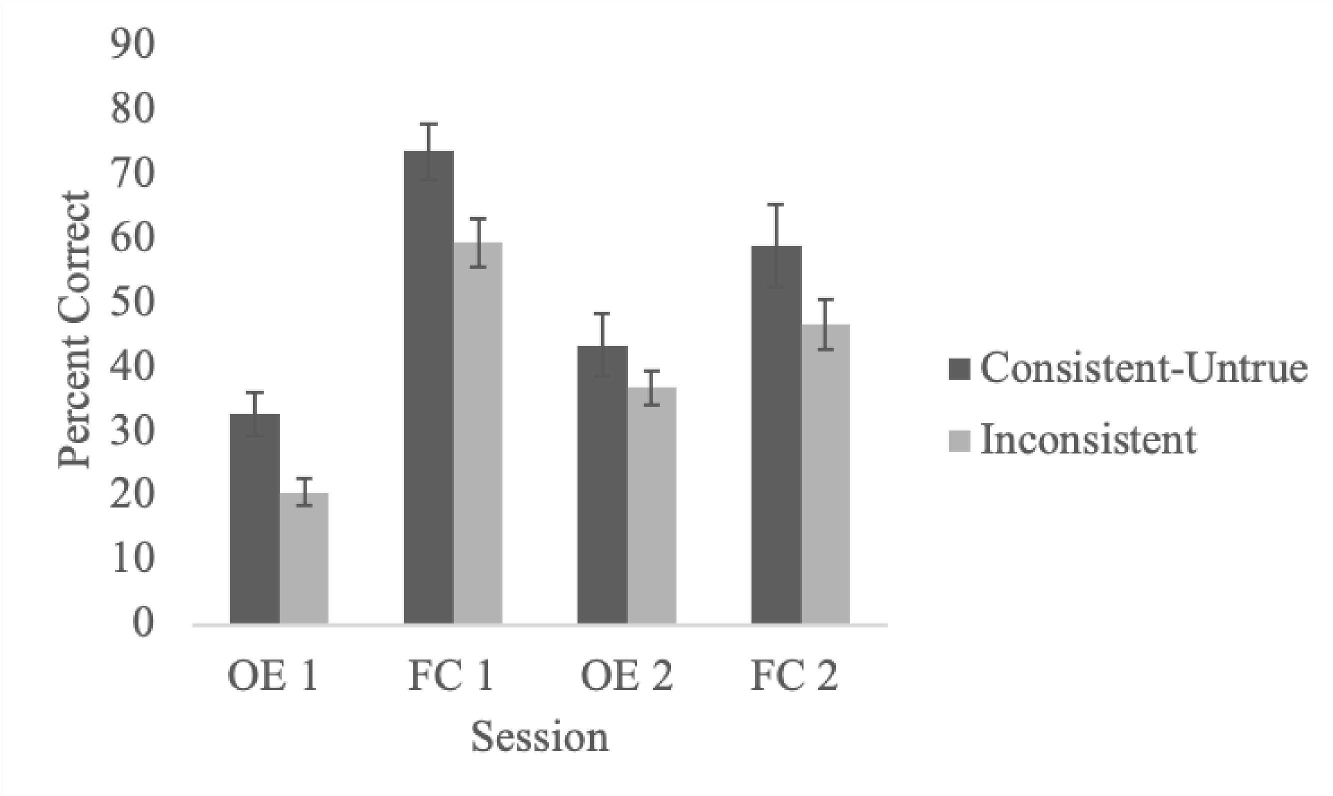


Figure 4

Inconsistent Information Timing: Successful Self-Derivation Performance

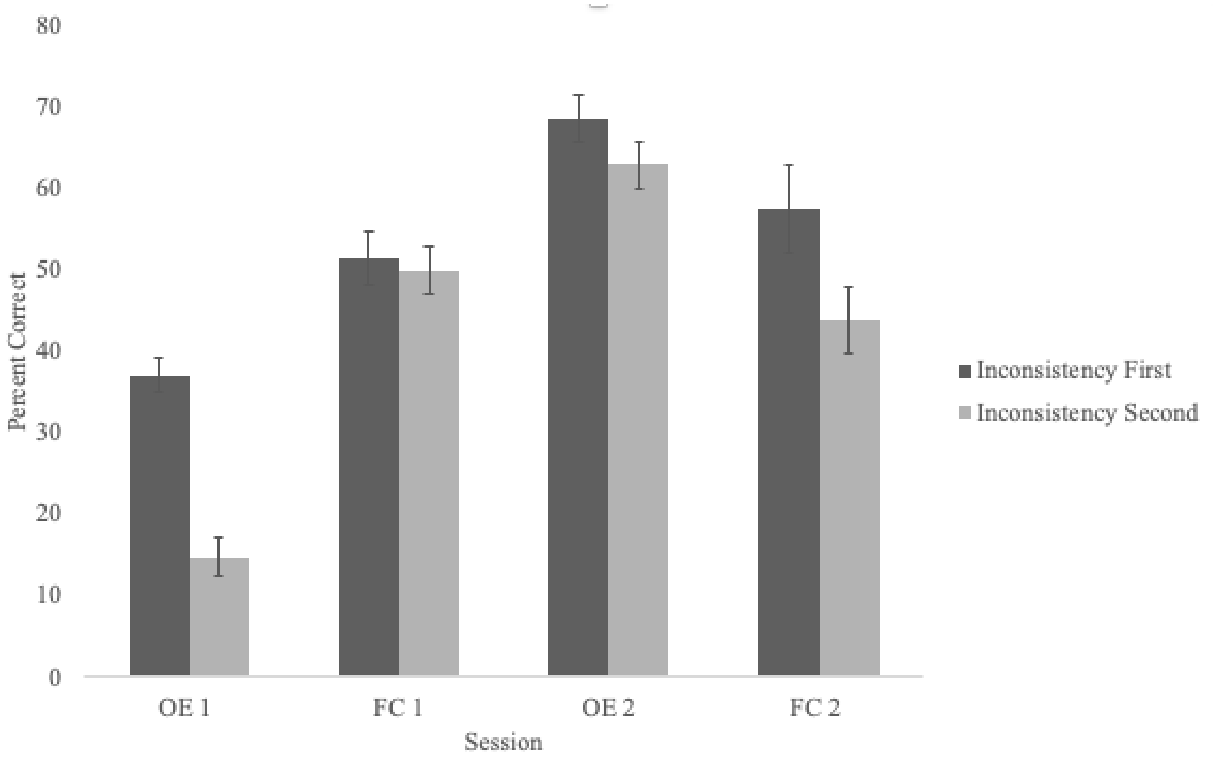


Figure 5

Inconsistent Information Timing: Incorrect Self-Derivation Performance

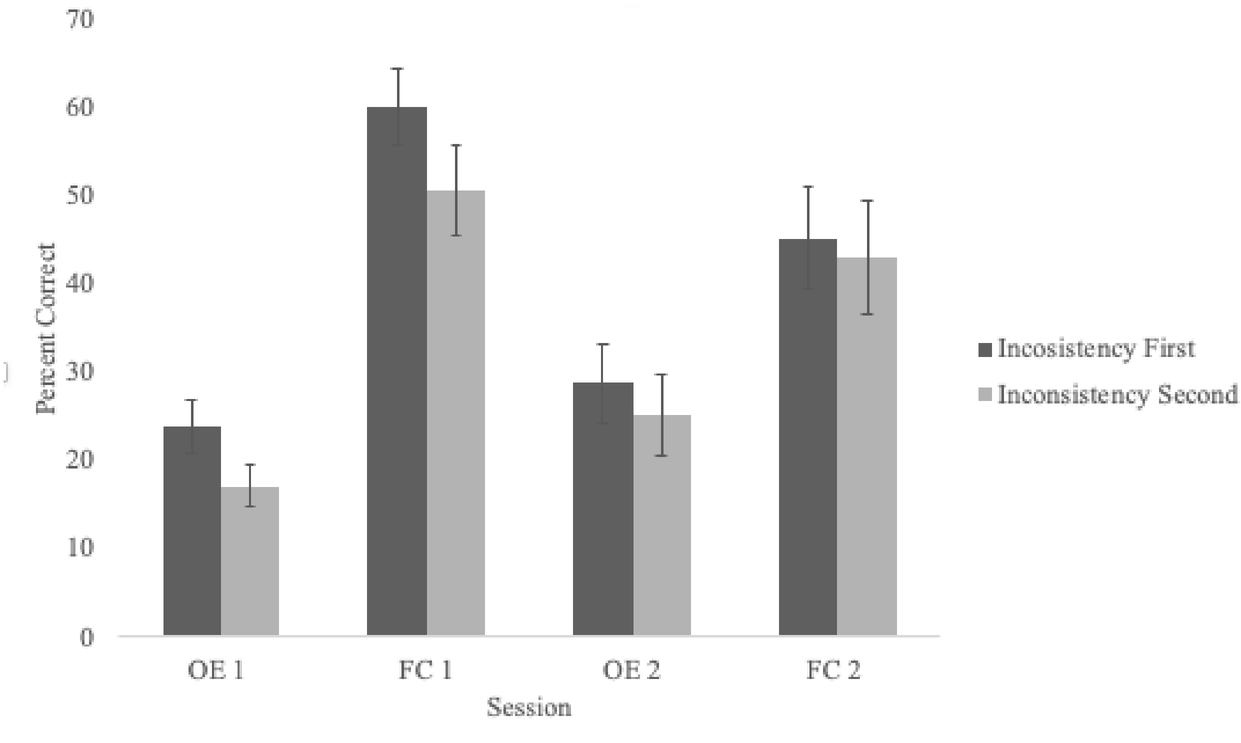


Figure 6
Average Evaluations of Truth by Condition

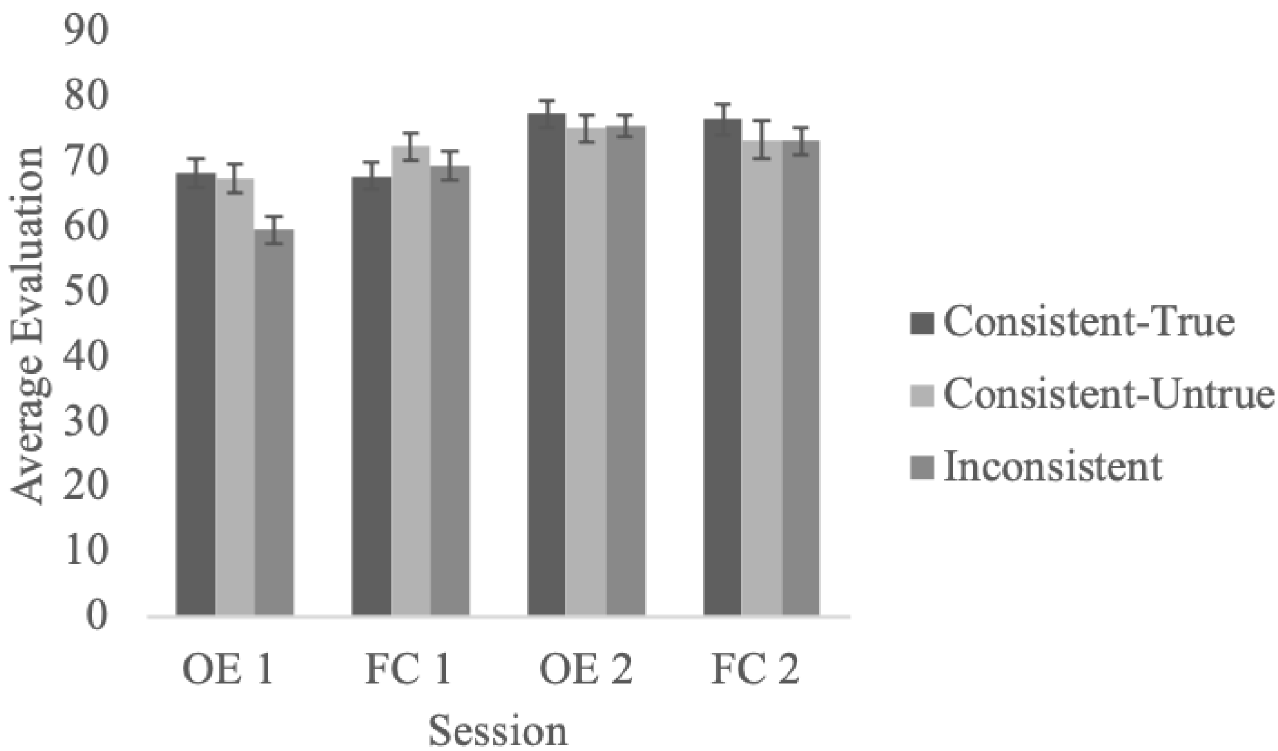
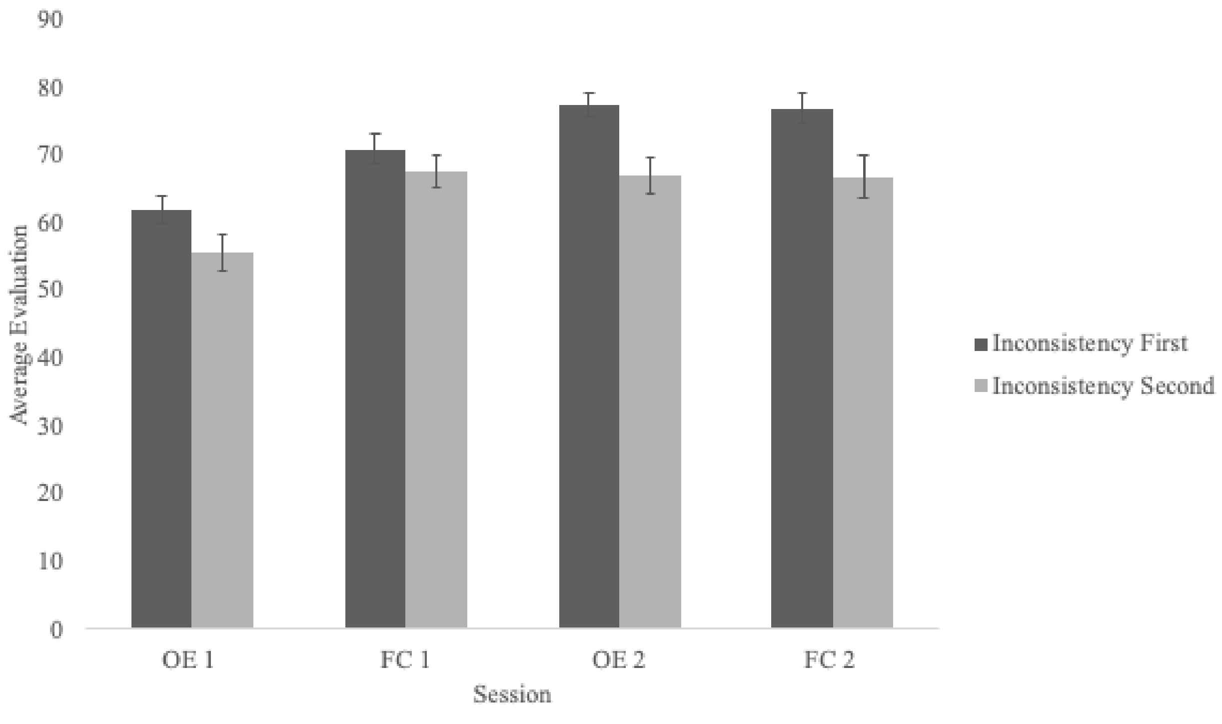


Figure 7

Inconsistent Information Timing: Average Evaluations of Truth



Appendix A

1. I am an Emory University undergraduate student (Agree/Disagree)
2. My gender identity is (Male, Female, Nonbinary, Other, Prefer not to say)
3. My race/ethnicity is (White/Caucasian, Black/African American, Asian, Latinx, Other, Prefer not to say)
4. I learned English before the age of 2 (Agree/Disagree)
5. I expect to graduate in (Fall 2020, Spring 2021, Fall 2021, Spring 2022, Fall 2022, Spring 2023, Fall 2023, Spring 2024, Fall 2024)
6. My prospective primary major is best classified as (STEM, Social Science, Business/Finance, Liberal Arts/Fine Arts)

Appendix B

1. The first vitamin was discovered by _______
   a. Marie Curie –*untrue but related*
   b. Don’t Know - *unknown*
   c. Casmir Funk – *correct self-derivation*
   d. Stephen Hawking – *incorrect self-derivation*
2. Softer body hair is common in people with ______
   a. Diabetes – *incorrect self-derivation*
   b. Malnutrition – *correct self-derivation*
   c. Cancer – *untrue but related*
   d. Don’t know – *unknown*