

Semantic Integration as a method for investigating concepts

Derek Powell, Zachary Horne, and N. Angel Pinillos

I. Introduction

The last ten years in philosophy have been marked by widespread interest in the psychology of philosophy. Much of this work has been carried out by experimental philosophers, who aim to better understand the contours of philosophical concepts and intuitions by importing the methods of the empirical sciences. Their hope is that a better understanding of the psychology of philosophical concepts such as KNOWS, PERSONHOOD, FREE WILL, and many others, will allow them to better assess philosophical arguments which utilize such notions.

Experimental philosophers have amassed many interesting results, but compelling concerns have been raised about the survey-based experimental methods that they typically employ. Here we argue, on the basis of these concerns and our own, that the possibility of experimental artifacts is good reason to adopt a new experimental paradigm that we call Semantic Integration. This methodology uses a memory task as an implicit measure of the degree to which different situations instantiate concepts. This measure avoids the methodological challenges researchers must address if they are to continue to use surveys.

The plan of the paper is as follows: First, we consider some challenges associated with survey methodology (section II) and then describe how Semantic Integration tasks can be used to implicitly examine people's concepts (section III). Next we argue that, by investigating concepts implicitly, Semantic Integration offers important advantages over more explicit survey methods (section IV). Finally, we discuss caveats regarding Semantic Integration methods (section V), variations on these methods, and briefly consider how they might be used alongside survey-based research (section VI).

II. The Methods of Experimental Philosophy

Experimental philosophers investigate philosophical concepts by presenting participants with short passages and then asking them to make judgments about what they read. These passages, which are often derived from philosophical thought experiments, are designed to test whether certain features are parameters for instantiating a philosophical concept. Studies using this survey methodology have improved substantially when compared to early research that lacked proper control conditions, but the methodology is still limited in important ways. In this section we review some of the challenges faced by researchers that use surveys. We discuss issues raised by Simon Cullen (2010), as well as other limitations of survey methodology.

a. Pragmatic cues in experimental materials

In a recent critique, Cullen (2010) argues that researchers conducting surveys need to take into account both the semantic and the pragmatic features of their experimental materials.

Grice (1975) observed that when people attempt to comprehend some utterance of natural language, they do not attempt to comprehend the exact meaning of the words as spoken or as written on the page. Rather, they attempt to comprehend the speaker's meaning. Cullen argues that the participants in an experiment behave similarly, attempting to comprehend the experimenter's meaning, a consideration that many experimental philosophers have ignored.

According to Grice (1975), people make assumptions about the requirements for rational communication. These assumptions, often referred to as Gricean norms, allow listeners or readers to grasp what a speaker means to convey or what they think a speaker means to convey. People assume that speakers are "cooperative communicators"-- that their utterances are true, orderly, relevant, and non-redundant. Typically, speakers are themselves aware that their interlocutors make these assumptions and so they exploit these assumptions to help conversational participants understand them. For example, sometimes the best way of making sense of someone's communications, given that they are following Gricean norms, is by inferring that they mean something that goes beyond what they said or stated with their utterance. If Speak asks, "Has the number two bus come by yet?" a listener, Hear, can rightly infer that there is a number two bus, that its route passes by this location, and that Speak is hoping to catch the bus. Of course, none of these facts are explicitly stated in Speak's question. Hear can infer these things because she assumes that Speak is following Gricean norms. For example, Speak would not be following the norm of relevance if he was not planning to get in the bus. Moreover, Speak can count on Hear to infer these things about his utterance because Speak knows that Hear will assume he is following the Gricean norms. Roughly, the propositions that the speaker means to convey (but go beyond what is said) and are inferable in a conversation applying the Gricean norms, are called "conversational implicatures" (as opposed to propositions that are conventionally associated with the words used). It is widely accepted that the deployment and computing of conversational implicatures is pervasive in human communication. For this reason, Cullen (2010) argues that if experimental philosophers ignore conversational implicatures, then their instructions, stimuli, response options, and other experimental materials may not convey the meanings they intend.

For an illustration of how conversational implicatures can affect survey results, consider research on "base-rate neglect." Base-rate neglect is the tendency for people to ignore relevant statistical base-rates when judging the probabilities of events and to instead rely on simpler heuristics (for a review, see Nisbett & Ross, 1980). In one study, Kahneman and Tversky (1973) presented people with a description of a fictional college student and asked them to estimate the probability that the student majored in various fields. If the descriptions included traits that seem stereotypical of an engineering student (e.g., introverted, enjoys solving problems), then people estimated the probability that the student was an engineer was high. What's interesting is that people make similar probability estimates even when they were told that only a small percentage of students study engineering. Kahneman and Tversky concluded that people ignore base-rate information when making their probability estimates and instead employ a representativeness heuristic: since the student resembles an engineer, they judge that it is probable he is one, and

they ignore the base-rate information which would suggest that any individual student is most likely not an engineer.

However, more recent research suggests that base-rate neglect may be due, at least in part, to conversational processes rather than to decision processes. If participants assume that experimenters are cooperative communicators, then they assume that the information they've been given is the most relevant to the task at hand. This may lead them to place a greater weight on the descriptions given than they would have otherwise. Schwarz and colleagues (1991) examined this by manipulating the guarantee of relevance. Participants in one condition were told that the descriptive information presented to them had been compiled by psychologists (as in the original experiments of Kahneman & Tversky, 1973), and in another condition they were told that the same description had been compiled by a computer that randomly sampled from a database of information. Whereas communication from another person comes with an implied guarantee of relevance, computer-generated text does not. As predicted, researchers found that participants were significantly less influenced by computer-generated descriptions than human-generated descriptions (Schwarz et al., 1991).

Even relatively subtle pragmatic cues can have important effects on people's responses to survey questions. For instance, people seem to place greater weight on the last source of evidence they are shown: Krosnick and colleagues (1991) found that base-rates had a larger effect on participants' judgments when they were the last piece of information participants read before making their response. The guarantees of relevance and non-redundancy imply that if experimenters present an apparently sufficient source of evidence (e.g., base-rate information), and then present another source of evidence (e.g., an individual description), then this second source should be interpreted as non-redundant and highly relevant to the task at hand.

Cullen (2010) demonstrated that pragmatic cues can likewise affect people's responses to philosophical thought experiments. However, he argues that researchers can overcome these challenges if they are sensitive to the context in which participants interpret their experimental materials, and the norms that govern these interpretations. Following Schwarz (1994), he argues that experimenters and participants are engaged in a conversation governed by the norms of cooperative communication (Cullen, 2010; Grice, 1975; Schwarz, 1994). Since participants abide by these norms, and expect researchers to abide by them as well, experimental materials must be constructed with pragmatic cues in mind. We agree that addressing the pragmatic features of experimental materials would improve the conclusions that can be drawn from surveys. However, overcoming these challenges might prove difficult. In practice, researchers still need to determine exactly how materials and questions ought to be phrased, and what implicatures they ought to contain. To make matters more difficult, this would need to be determined for each concept that experimental philosophers intend to examine.

To illustrate the difficulty of designing appropriate questions and materials, consider the challenges faced by researchers studying causal learning: An important construct in research on causal learning is *causal strength*, defined as the probability that some cause produces an effect (Cheng, 1997). Although people often make judgments about causal strength, researchers can

ask participants to report such a judgment in any number of ways, and it is not obvious which way is optimal. In one experiment Buehner, Cheng, and Clifford (2003) asked their participants to make a causal strength rating on a scale from 0 (X does not cause Y at all) to 100 (X causes Y every time). They found that participants' judgments tended to cluster into two groups: one group of participants made judgments consistent with Cheng's (1997) probabilistic definition of causal strength, whereas other participants made judgments consistent with competing associative models. As causal learning is often taken to be a relatively basic cognitive mechanism (Cheng, 1997), it would be remarkable if some people learned causal relationships via wholly different cognitive mechanisms. However, Buehner and colleagues investigated whether ambiguities in the question they used to probe participants' judgments were responsible for the divergent pattern of responses. Indeed, they noted that the causal strength question they initially used can be interpreted as applying in one of two different contexts: 1) the experimental learning context where the effect is also produced by other background causes or 2) a counterfactual context where only the cause of interest is present. The clustering of participants' different responses was explained by these different interpretations: under the first interpretation, the best response is consistent with an associative model, whereas under the second interpretation the best response is consistent with the probabilistic definition of causal strength.

After further research, these researchers discovered that the best way to unambiguously probe participants' causal strength estimates was to phrase their questions counterfactually and in terms of frequencies (Buehner et al., 2003). For example, "Imagine 100 healthy people started smoking cigarettes, how many do you think would develop cancer?" This wording makes the context clear, and participants do not need to make any inferences beyond what is stated. The upshot of all this is that resolving ambiguities and constraining participants' interpretations of questions and materials is feasible, but can require systematic investigation for each concept at issue.

b. Demand Characteristics

The results of surveys can also be affected by *demand characteristics* (Orne, 1962). Crudely put, demand characteristics are artificial features of an experimental task that lead participants to perform some task other than what the researchers intended them to.¹ Demand characteristics can occur when participants are apprehensive about being evaluated (Weber & Cook, 1972). Apprehension can lead participants to respond in ways they perceive as either socially desirable, or "correct," irrespective of their actual attitudes or intuitions. Demand characteristics can also occur when participants assume the role of a *faithful participant*,

¹ Under this definition, there is some overlap with the concerns raised previously. For instance, pragmatic cues in base-rate experiments may have led participants to focus only on the descriptions they were given and to suppress information about base-rates. Since the experimenters were interested in how participants would use all the information they were given to produce the most accurate judgment they could, participants who interpret the instructions in this way clearly did not perform the intended task.

eschewing all pragmatic cues and following instructions exactly to the letter (Weber & Cook, 1972).

Survey materials in experimental philosophy studies are particularly likely to exhibit demand characteristics because experimental philosophers often present naive participants with bizarre thought experiments. Although the uniqueness of thought experiments is harmless in professional philosophy, there is evidence that survey participants are more likely to assume a faithful role, ignoring pragmatic and contextual cues, when experimental materials are particularly unrealistic (Weber & Cook, 1972). In other words, if experimental materials are convoluted or strange, then participants are more likely to ignore the contextual cues in experimental materials, or to interpret them under different assumptions. Additionally, if participants are apprehensive about being evaluated, then they are more likely to try to guess at desirable or “correct” response. When the passages they are asked to read are bizarre thought experiments, they may engage in a kind of amateur philosophizing, diverging from the aims of experimental philosophers. If demand characteristics cannot be ruled out, then it is unclear how to interpret the results of surveys.

III. Semantic Integration

In this section, we propose a new methodology for investigating concepts that we call Semantic Integration. First, we introduce research on memory and language processing that inspired the Semantic Integration methodology. Then, we describe the components of a Semantic Integration task, and two experiments in which we employ the method.

Semantic Integration uses memory tasks as an implicit measure of how concepts are activated by different situations. As we discuss, this method has important advantages over survey-based research: it minimizes the influence of pragmatic cues and greatly reduces the possibility of demand characteristics. It also provides a more direct way of examining concepts. That is, Semantic Integration provides a measure of conceptual activation that is not influenced by downstream judgment or decision processes. In contrast, participants’ responses to survey questions in experiments typically constitute their judgments about whether a particular concept applies in a given situation. These judgments may be the products both of people’s concepts as well as downstream decision processes.

a. Memory and Language Processing Research

People tend to think of errors in memory as errors of omission-- they acknowledge that we sometimes forget things that have happened to us, but assume that we can only form memories for events that we have experienced. Yet, psychologists have amassed significant evidence that people sometimes remember events that never actually occurred (for a review, see Schacter, 1995), indicating that memory is not entirely dependent on external inputs. Bartlett (1932) is often credited with reporting the first experimental evidence for the formation of false memories. In his research, he had participants read a story and then recall it several times after

subsequent delays. His analyses were informal, but he reported that memories grew increasingly distorted after each recall. Since Bartlett, researchers have found evidence for the formation of false memories in list learning paradigms (Deese, 1959; Roediger & McDermott, 1995), as well as in retention of sentences (Bransford & Franks, 1971), longer prose passages (Sulin & Dooling, 1974), image sequences (Loftus, Miller, & Burns, 1978) and videos (Loftus & Palmer, 1974). These phenomena are more than just curiosities; researchers have leveraged false memory to investigate the nature of our mental representations as well as our language comprehension processes.

Psychological research indicates that people's memories are influenced by semantic processing, and that people's memory is better for semantic information than for specific episodes or verbatim utterances (Anderson, Bjork, & Bjork, 1994; Anderson & Ortony, 1975; Deese, 1959; Loess, 1967; Roediger & McDermott, 1995; Sachs, 1967). Even in simple experimental contexts (e.g., learning lists of words), experiences are processed and given semantic representations. In one study, Roediger and McDermott (1995; also see Deese, 1959) asked participants to memorize lists composed of different words that were semantically related to a single target word. When participants were asked later to recall the words they had been presented with, they were often just as likely to falsely recall the target word, which had never been presented, as any of the other words that actually appeared in the list. For example, when presented with a list made up of words like 'glass', 'pane', and 'shade', people are likely to recall the target word 'window', even if the word never appeared in the list. To introduce some terminology, the words in the list *semantically activate* the word 'window' -- which is to say that they cause people to form or retrieve stored mental representations associated with the word.

Researchers have leveraged the relationship between false memories and semantic activation to examine language processing (e.g., Bransford & Franks, 1971; Flagg, 1976; Gentner, 1981). In particular, prior research investigated how semantic information is combined to form meaningful structured representations, or *discourse meanings*. This process, sometimes called *semantic integration* (Bransford & Franks, 1972), enables people to comprehend complex ideas communicated through connected discourse. Early research by Sachs (1967) found that memory for the meanings of sentences is more robust than memory for their specific wordings. He asked participants to read passages and then tested their recognition for sentences either immediately, or after they had read different amounts of intervening material. Some of the tested sentences had actually appeared in the text, but others were altered semantically or syntactically. When the meanings of the sentences were changed, participants made few errors; even after substantial distraction, participants rarely reported memory for sentences that had not appeared in the passage. However, when the changes were syntactic (e.g., a shift from active to passive voice), participants often reported recognizing the new sentences. After distraction, their recognition performance was near chance. Sachs concluded that during language processing, the original form of presented material is stored temporarily, only long enough to be comprehended, whereas the material's meaning is encoded into long-term memory.

If semantic information is integrated during language processing and it is the meaning of a passage that is encoded into memory, then memory ought to exhibit *productivity*. That is, it should be possible for exposure to several basic, interrelated sentences to produce false memory for a sentence that expresses the integrated representation. A number of studies have confirmed this prediction, indicating that people integrate simple sentences to form representations for more complex sentences during language comprehension (Bransford & Franks, 1971; Cofer, 1973; Flagg, 1976). Additionally, people have been found to integrate information from text passages read during an experiment with their general background knowledge, leading to false recall for additional information that was not experimentally presented (Owens, Bower, & Black, 1979; Sulin & Dooling, 1974; Thorndyke, 1976).

To explain these findings, Gentner (1981) examined a model of language processing in which sentences are considered both individually and in the broader context in which they appear. Her model states that when a sentence is read within the context of a larger passage, the discourse meaning that a reader forms may incorporate information not contained in the original sentence. She focused her investigation on an examination of the integration of verb meanings in context. Following research in linguistics (e.g., Chafe, 1970), artificial intelligence (e.g., Schank, 1972; 1973), and psychology (e.g., Miller & Johnson-Laird, 1976; Stillings, 1975), Gentner hypothesized that complex verb meanings can be represented by networks of *subpredicates* that express semantic relationships. Crudely put, a verb's subpredicates are simpler verbs that function as components of the more complex verb's meaning. To illustrate, consider the relationship between the verb 'give' and the more specific verb 'pay'. On Gentner's analysis, 'giving' some item is to take some action that transfers ownership of that item to a recipient. 'Paying' is a more specific form of giving, in which the giver owes the recipient. Thus, a representation of 'gave' would include subpredicates like 'caused', 'changed', and 'possession', and a representation of 'paid' would add the subpredicate 'owed'. Gentner tested this hypothesis by asking her participants to read paragraph-long stories that each included a sentence with the verb of interest -- the *critical sentence*. For instance, one story contained the critical sentence, "Max finally gave Sam the money". In the experimental condition, additional context explained that Max owed Sam money, whereas the control condition lacked this context. After reading one version of the story, participants performed a recall task in which they were shown the critical sentence with the word 'gave' removed, and they were asked to fill in the word that had appeared in the story. In support of Gentner's predictions, participants who had been provided with the additional context were more likely to falsely recall the verb 'paid' than participants in the control condition.

b. Using Semantic Integration to investigate philosophical concepts

Whereas Gentner (1981) used a false recall paradigm to examine how verbs with known meanings are integrated during language processing, we propose that the same methods can be used to investigate the meanings of philosophically significant concepts.

On a traditional view, many philosophical concepts are complex mental entities constituted by simple concepts. The simple concepts jointly provide a “definition” of the complex concept. This means that the constituent concepts express properties which provide necessary and (jointly) sufficient conditions for the instantiation of the complex concept. In the terms of semantic integration research, the traditional view makes the prediction that a concept *C* will have subpredicates which are the constituents of *C*.² For example, a view which says that KNOWLEDGE is a complex concept constituted by JUSTIFIED TRUE BELIEF (e.g., Ayer, 1956) will make the prediction that these constituent concepts, expressing necessary conditions, will be subpredicates for KNOWLEDGE. Researchers can test whether including these subpredicates in a passage leads to false recall for words picking out KNOWLEDGE, offering evidence that these subpredicates were integrated to produce the concept. When this integration occurs, researchers can infer that the concepts JUSTIFIED TRUE BELIEF are constituents of KNOWLEDGE.

This application of Semantic Integration is straightforward because we are able to propose a jointly sufficient set of constituents for the concept. However, there may be other situations where this is not possible. For instance, some constituents of a complex concept might be unknown. Alternatively, some concepts may be simple, or may be sensitive to non-constituent parameters, or may be context sensitive. Fortunately, Gentner (1981) showed that concepts can play the role of subpredicates for a target concept even when they are not jointly sufficient or necessary for instantiating the target concept. For example, she shows that people falsely recall ‘painting’ when they integrate ‘working’ and “workers are carrying brushes, whitewash, and rollers.” Strictly speaking, these features are not jointly sufficient for painting. The workers might have carried the whitewash but ended up working on something unrelated to painting. Of course, we expect participants in Semantic Integration tasks to understand the story in a plausible way. As a result, there is no requirement that the items which are integrated to yield a concept actually form a jointly sufficient set for that concept in some strong metaphysical sense. Nor is there a requirement that the items playing the role of subpredicates in integration correspond to necessary conditions for the concept at issue. Recall Gentner’s example: ‘Carrying brushes, whitewash, and rollers’ is not a necessary condition for painting (think of spray painting). All that is required is that the context makes it more or less likely that the target concept is instantiated.

This is good news for four reasons: First, complex concepts can be examined even if some of their constituents are unknown. If we are interested in studying a complex concept, we can examine participants’ integration of a set of concepts that merely approximate its true constituent concepts. For example, suppose that a concept *C* has constituents X_1 , X_2 , and X_3 and we want to test whether X_1 is a constituent. We can test for false recall of the lexicalization of *C* in the presence of X_1 and X_2 without invoking X_3 or by approximating X_3 .

² Thus far we have treated subpredicates as linguistic items, but this is not necessary. One could also think of them as mental representations or concepts.

Second, some philosophers have argued that many philosophically interesting concepts are simple (e.g., Fodor, 2002; Williamson, 1998). Yet, these concepts may still have interesting necessary conditions. For example, Williamson (1998) holds that although KNOWLEDGE is simple, the concept still has philosophically important necessary conditions. Many philosophers hold that a necessary condition for S knowing P is that P be true. Semantic Integration can be used to examine whether the necessary condition is something that laypeople accept.

Third, some philosophers think that some concepts are sensitive to certain parameters and that this sensitivity is accessible to laypeople. For example, Joshua Knobe (2010) holds that competent folk mental state attributions are sensitive to the moral valence of the content attributed, and some epistemologists have claimed that competent folk knowledge ascriptions are sensitive to practical interests (Pinillos, 2012; Stanley & Sripada, 2012) and moral properties (Beebe & Buckwalter, 2010). In these cases, parameters like moral valence and practical interests do not necessarily constitute interesting necessary conditions for the concept. Yet, Semantic Integration is still apt for testing these parameters. We can do this by developing vignettes that include a critical sentence whose truth, together with a parameter, is thought to yield the target concept. If people consider the parameter to be relevant to the target concept, then the presence of the parameter ought to lead to greater false recall for words that lexicalize the concept (e.g., Henne & Pinillos in preparation; Waskan et al., submitted).

Fourth, Semantic Integration can be used to study the semantic contours of a word even if the word does not express a unique concept. Certain words may be context sensitive, expressing different concepts depending on the conversational setting. In the previous cases discussed, the concepts X_1, X_2, \dots, X_n are integrated into a concept C, yielding false recall of a lexicalization of C. If a word is context sensitive, then its use may express different concepts C_1, C_2, \dots, C_n depending on the discourse context. A set of concepts X_1, X_2, \dots, X_n might integrate to produce C_1 , but not C_4 . Whether the concepts X_1, X_2, \dots, X_n are integrated and produce recollection for a lexicalization of C depends on the discourse context.

Many philosophers accept that ‘knows’ is context sensitive, sometimes expressing different concepts KNOWS₁, KNOWS₂, ... , KNOWS_n corresponding to the different standards associated with ‘knowledge’ (DeRose, 1995; Cohen, 1986; Lewis, 1996). If a conversation takes place in a casual setting, the standards for knowledge might be lower than in a conversation taking place in a philosophy classroom. In the former, the use of ‘knows’ might express the concept KNOWS₁ while the use of ‘knows’ in the latter context may express the concept KNOWS₄. In examining the context sensitivity of a word, researchers can present information in the vignette to establish the discourse context and set the target concept. Being in a casual setting may communicate to the reader that the epistemic standards of the discourse context are low. Under low epistemic standards, the concepts LOW JUSTIFICATION and P is TRUE may yield KNOWS₁ which the word ‘knowledge’ expresses under low epistemic standards. Suppose the target sentence is “S believes that P.” In this case, participants would be likely to falsely recall ‘knows’ as having appeared in the sentence. Alternately, the discourse context associated with being in a philosophy classroom would establish a high epistemic standards, then the

concepts LOW JUSTIFICATION and P is TRUE might not yield KNOWS₄, and participants would be unlikely to falsely recall ‘knows’. If this account is on the right track, then researchers can also exploit Semantic Integration to examine context sensitivity.

What these four points reveal is that the viability of the Semantic Integration method does not depend on any particular understanding of concepts. On the contrary, the method is applicable under a wide variety of assumptions about concepts. The versatility of the method is then especially useful for philosophers who themselves might disagree about the very nature of concepts.

c. Two Experiments Using Semantic Integration

In the remainder of this section, we discuss two experiments we conducted that demonstrate how Semantic Integration can be used to investigate philosophically significant concepts. In this research we focus on the concept KNOWLEDGE, but recently other researchers have adopted our methods in order to examine EXPLANATION (Waskan et al., submitted) and CAUSATION (Henne & Pinillos, in prep).

There are three main components in a semantic integration study. The first component is the passage containing the contextual information hypothesized to semantically activate the target concept. In order to construct passages that yield false recall of KNOWLEDGE, we altered contextual information in different versions of a main story, controlling for word count, sentence length, and overall structure. In a preliminary study, we constructed two versions of a story about a detective (Jack Dempsey) who forms the belief that a suspect (a teenager named Will) is guilty. In the experimental condition, the detective’s belief is justified by legitimate evidence and his belief is true (the suspect is in fact guilty). In the control condition, the detective cannot find any evidence and participants are not told whether the suspect is guilty, but the detective forms the belief anyway.

In each of these stories, we included a critical sentence containing a critical verb. Recall that when sufficient contextual information licenses using a more specific verb, people will falsely recall the more specific verb as having appeared in the passage. The critical verb must be consistent with the concept under investigation, but must not entail it. In our knowledge experiment, we chose ‘thought’ as our critical verb; thinking that P is consistent with knowing that P, but does not *entail* knowing that P. We predicted that when read in the right context, a sentence containing the word ‘thought’ would lead to false recall of the word ‘knew’. We predicted this will occur more frequently in the experimental condition where the appropriate context is supplied, than in the control condition.

Critical sentence: “Whatever the ultimate verdict would be, Dempsey thought Will was guilty.”

An additional consideration when choosing a critical and target words is the frequency with which that word occurs in English communications. Generally, it has been found that recall

performance is better for high-frequency than for low-frequency words, and that the opposite is true for recognition performance (Kintsch, 1970). That said, there is some evidence that low-frequency words might benefit at recall when they presented together with high-frequency words (Duncan, 1974; Gregg, 1976), as will likely be the case in Semantic Integration experiments. A good practice is to ensure that critical and target words are matched for frequency of occurrence as closely as possible. ‘Thought’ and ‘knew’ are reasonably well-matched as the 179th and 300th most common English words, respectively (Wolfram|Alpha, 2013).

The second component of a Semantic Integration study is a distractor task. In principle, this distractor task could consist of almost anything. The purpose of the distractor is simply to diminish the effect of episodic memory in the recall task. Importantly, however, distractors should not contain either the critical verb or the target word.

After reading the distractor, participants advance to the third part of the experiment, the recall task. There they are shown several sentences from the story, each with one word removed. Their task is to recall the word that appeared in the blank. In our experiment, we were interested in their recall performance for the critical sentence. During the recall task, participants were shown this sentence with the word “thought” replaced with a blank, as shown below:

Recall Task: “Whatever the ultimate verdict would be, Dempsey _____ Will was guilty.”

Participant typed in the word that they recalled as having appeared in the original story. Consistent with our predictions, participants were more likely to recall “knew” as having appeared in the sentence when the detective’s belief was justified and true (Powell et al., Draft).

Clearly, this finding does not demonstrate anything particularly interesting about KNOWLEDGE, but it does demonstrate that semantic integration can be used to examine philosophical concepts. Consequently, we investigated Gettier cases, a clearly more substantive issue in philosophy. We adapted our detective story and added another character named Beth. Beth is Will’s soon-to-be ex-girlfriend, who has it in for Will and interferes with Dempsey’s investigation. We created three versions of the story, a false belief version, a Gettier version, and a justified true belief version.

In the false belief condition, Will is innocent, but Beth framed him by committing the crime and planting evidence.

In the Gettiered condition, Will committed the crime and disposed of all the evidence, but Beth makes sure Will gets caught by planting evidence for Dempsey to find.

In the justified true belief condition, Will committed the crime and left behind evidence. Seeing his mistake, Beth does nothing and waits for Dempsey to arrest Will.

We found participants were more likely to falsely recall ‘knew’ as having appeared in the critical sentence in the justified true belief and Gettiered conditions than in the false belief condition. However, we also found *no* difference in recall between the Gettier and justified true belief conditions. That is, a case of Gettiered justified true belief activated participant’s knowledge concept to the same degree as non-Gettiered justified true belief. This suggests that our participants did not distinguish between Gettiered and non-Gettiered justified true belief (Powell et al., draft). Though our findings may be surprising to some philosophers, they are consistent with results reported by Starmans and Friedman (2012), who concluded that laypeople’s conception of knowledge greatly resembles the traditional definition: JUSTIFIED TRUE BELIEF. Still, more research is needed to explore the contours of laypeople’s concept of knowledge.

IV. Pragmatic Considerations and Demand Characteristics

Semantic Integration tasks offer two important advantages over more explicit survey methods. For one, Semantic Integration tasks avoid the concerns raised by Cullen (2010) over pragmatic cues. Researchers using survey methods need to account for pragmatic cues in the stimuli that they present to participants as well as in their instructions, questions, and response options. In a Semantic Integration experiment, participants are told they are performing a memory task and nothing in the instructions, response prompts, or options indicates otherwise. While these materials are not devoid of pragmatic cues, pragmatic factors in this context are considerably less problematic, and considerably better understood. Psychologists have studied memory since Ebbinghaus (1885/1964), and have developed reliable methods for testing people’s recollection of presented material. While it is clear that *stimuli* may still contain pragmatic cues and conversational implicatures, this fact is not in any way unique to Semantic Integration. For one, survey methods will also face these same concerns. Moreover, if one were skeptical about an experimental paradigm for this reason, one would also have to be skeptical about research on causal reasoning, decision making, psycholinguistics, or nearly any line of research that involves presenting text to participants. The pressing concern is that pragmatic cues in instructions will lead participants to approach the experimental task incorrectly, or to interpret their response options in a manner inconsistent with the researcher’s intentions. Semantic Integration tasks avoid these difficulties.

Second, Semantic Integration tasks largely preclude demand characteristics. Even if participants are apprehensive about being evaluated, their apprehension is unlikely to lead researchers to any erroneous conclusions. Evaluation-apprehension should motivate participants to perform the task well, and since there is no reliable way for participants to produce “desirable” answers except by probing their own memory, there is little risk of evaluation-apprehension leading to spurious findings. In addition, because the memory task is both intelligible and experimentally realistic, participants are less likely to take on the role of the faithful participant (Weber & Cook, 1972). Even if some participants do ignore experimenters’ conversational

implicatures, this is unlikely to affect their performance, as the instructions of a memory task can be made comprehensible without many contextual cues.

V. Caveats

The interpretation of findings from semantic integration tasks depends on resolving three questions:

- (a) How are concepts structured?
- (b) What mental process leads to integration of semantic information?
- (c) Does “impure” semantic integration complicate matters?

a. The structure of concepts

If Semantic Integration directly measures the semantic activation of people’s concepts, then one might wonder about the nature and structure of these concepts. As discussed, Gentner (1981) hypothesized that verb concepts are represented as structured collections of subpredicates. On the basis of this view, she made and confirmed very specific predictions about how representations would be combined during the processing of connected discourse, lending support for this theory. Still, psychologists have attempted to describe concepts using a number of representational formats (e.g., Posner & Keele, 1968; Medin & Schaffer, 1978). This may prompt some to doubt that Gentner’s model of concepts is accurate, or to worry that, even if it accurately describes the representations of certain concepts, different types of concepts may be represented in other ways (e.g., natural kind terms, prototype or exemplar models, distributed representations, etc.). Although these possibilities may complicate the interpretations of Semantic Integration experiments, researchers who use Semantic Integration can remain agnostic to the “true” psychological theory of concepts. In fact, the method rests on two basic assumptions: (1) semantic concepts are mentally represented in some fashion and (2) memory for the *meaning* of a passage is more robust than memory for its exact wording. The first claim is a fundamental assumption of modern psychology and one which we will not defend. The second is supported by a large body of research on memory, some of which we discussed in section III (e.g., Bransford & Franks, 1971; Brewer, 1977; Barclay, 1973; Cofer, 1973; Flagg, 1976; Sachs, 1967).

b. Mental Processes and Semantic Integration

Thus far we have reasoned as if integration occurs during comprehension and encoding, but another possibility is that integration actually occurs at recall. That is, during encoding people store the meanings of individual propositions separately. Then, at recall, they integrate these meanings by a process of inference to form a reconstruction of the memory for an individual sentence or proposition. Supposing this is true, it is worth noting that Semantic Integration still overcomes concerns about demand characteristics and pragmatic cues. However,

it can no longer be said to provide a direct a measure of semantic activation. Rather, in this case the responses that participants give to recall prompts are just as dependent on inferential processes as their responses to surveys. Fortunately, Gentner (1981) tested this possibility by inserting contextual information both before and after the critical sentence in a passage. She found that false recall for critical items was greater when the inserted material came before the critical sentence, supporting the interpretation that meanings are integrated on-line during discourse comprehension rather than after the fact during recall. This supports the claim that Semantic Integration isolates conceptual activation from downstream inferential processing.

c. Impure Semantic Integration

“Pure cases of semantic integration” (Gentner, 1981, p. 371) occur when the subpredicate structures of n propositions are directly combined to produce some unified structure. For example, Gentner describes ‘gave’ and ‘owed’ resulting in recall for the verb ‘paid’. However, she also provides evidence that semantic integration can occur when the context does not directly specify any of the subpredicates in the new semantic structure. As discussed earlier, she found that people recall ‘painting’ in place of ‘working’ when they are told that the workers are ‘carrying brushes, whitewash, and rollers’. For integration to occur, people need to infer that the workers are using these materials, and thus, that they are painting. Ideally, when researchers are examining complex concepts they can make inferences about the subpredicate structure of a concept based on people’s integration performance. Yet, it would clearly be an error to infer on the basis of Gentner’s findings that ‘carrying brushes, whitewash, and rollers’ is really a component of the subpredicate structure of ‘painting’. As we noted earlier, spray painting does not require any of these materials. Earlier we identified the possibility of “impure” semantic integration as advantageous, allowing researchers to apply Semantic Integration methods in many different situations. However, this also means that researchers should exercise caution when making inferences about the subpredicate structures of putatively complex concepts on the basis of integration performance.

VI. Alternate Experimental Designs and Surveys

a. Similar experimental paradigms

In this paper we described an experimental method modeled after Gentner’s (1981) work on the semantic integration of verb meanings, and described its use for examining people’s concept of KNOWLEDGE. It bears noting that there are a number of other related experimental paradigms that have been used to examine semantic integration in discourse comprehension (e.g., Bransford & Franks, 1971; Brewer, 1977; Barclay, 1973; Cofer, 1973; Flagg, 1976; Sulin & Dooling, 1974; Thorndyke, 1976; Owens et al., 1979), and that some of these paradigms might also be employed by experimental philosophers. However, Gentner’s (1981) paradigm has

several qualities that are desirable for experimental philosophers, even relative to other semantic integration tasks.

First, the use of a free recall task makes its results more compelling than tasks that rely on recognition judgments. Participants' responses to recognition tasks can be influenced by both true recollection as well as mere feelings of familiarity (Tulving, 1985). In contrast, explicit recall of the word 'knew' provides unambiguous evidence for the semantic activation of the concept KNOWLEDGE.

Second, this paradigm focuses responses onto a single specific word of interest, whereas other semantic integration paradigms often ask participants to evaluate larger semantic units, such as phrases or sentences (e.g., Bransford & Franks, 1971; Sulin & Dooling, 1974). Specifying a target verb can reduce ambiguity in investigations of individual concepts. Thus, where possible, the Semantic Integration tasks we describe here are a superior method for examining the parameters involved with instantiating people's concepts.

Of course, not all concepts of interest will necessarily have a verb form ('knew'), with nearby synonyms ('thought', 'believe'). Where this is not the case, other Semantic Integration tasks may be more appropriate. The disadvantages associated with Semantic Integration tasks measuring recognition for sentences or phrases (e.g., Bransford & Franks, 1971; Owens et al., 1979) are not insurmountable. In particular, memory researchers have developed procedures, like the remember-know procedure (Tulving, 1985), that can help distinguish between genuine recollection and familiarity. With sufficient care, phrases or sentences can be crafted to unambiguously express whatever concept may be of interest to researchers (e.g., Waskan et al., submitted).

b. Surveys and Semantic Integration

The methodological advantages of Semantic Integration owe to the implicit nature of the task. However, this also marks Semantic Integration tasks as importantly different from the explicit measures collected during survey tasks. Different research questions might warrant the use of either surveys or Semantic Integration.

Many experimental philosophers hope to assess philosophical arguments by examining the psychology of concepts they employ. We have argued that, in general, Semantic Integration tasks are well suited for accomplishing this goal. Semantic Integration tasks provide an implicit measure of conceptual activation, making them ideal for capturing these sorts of intuitive reactions. However, some philosophical concepts may also be applied to situations by more effortful cognitive processes. In these cases, explicit survey questions that elicit conscious consideration may be better suited, if these questions can be adequately constructed. Additionally, surveys may be more appropriate where experimental philosophers are interested in people's judgments. For instance, some researchers may not be interested in KNOWLEDGE per se, but in knowledge ascription behavior. In this instance, Semantic Integration tasks are inappropriate and surveys would be preferable.

VII. Conclusion

In this paper, we discussed the ways in which pragmatic cues and demand characteristics can affect the results of surveys. In light of these problems, we argued that experimental philosophers should adopt a new experimental paradigm that we call Semantic Integration. Our experimental investigations of KNOWLEDGE demonstrate how this method can be used to examine philosophical concepts. Semantic Integration can be applied to investigate complex concepts in a manner consistent with the aims of traditional conceptual analysis, and used to examine other parameters relevant to the instantiation of concepts. This method avoids concerns about pragmatic cues and demand characteristics because participants' conceptual activation is measured implicitly through a memory task. For these reasons, Semantic Integration represents an important methodological advance in experimental philosophy.

References

- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of experimental psychology. Learning, memory, and cognition*, 20(5), 1063–87.
- Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles — A problem of polysemy. *Cognitive Psychology*, 7(2), 167–180.
- Ayer, A. J. (1956). *The Problem of Knowledge*. London: Macmillan.
- Barclay, J. (1973). The Role of Comprehension in Remembering Sentences. *Cognitive Psychology*, 254, 229–254.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.
- Beebe, J. & Buckwalter, W. (2010). The Epistemic Side-Effect Effect. *Mind and Language* 25(4), 474–498.
- Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331–350.
- Bransford, J. D., & Franks, J. J. (1972). The abstraction of linguistic ideas: A review. *Cognition*, 1(2-3), 211–249.
- Brewer W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition*, 5(6), 673–678.
- Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of experimental psychology. Learning, memory, and cognition*, 29(6), 1119–40.
- Chafe, W. L. (1970). *Meaning and the structure of Language*. Chicago: University of Chicago Press.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2), 367–405.

- Chisholm, Roderick M. (1956). *Perceiving: A Philosophical Study*. Ithaca, New York: Cornell University Press.
- Cofer, C. N. (1973). Constructive Processes in Memory: The semantic and integrative aspects of memory may reach far beyond the usual notions of the retention of dates, names, and events. *American Scientist*, 61(5), 537–543.
- Cohen, S. (1986) Knowledge and Context. *The Journal of Philosophy*, 83: 574–583.
- Cullen, S. (2010). Survey-Driven Romanticism. *Review of Philosophy and Psychology*, 1(2), 275–296.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology*, 58(1), 17–22.
- DeRose, K. (1995) Solving the Skeptical Problem, *The Philosophical Review*, 104(1), 1–52.
- Duncan, C. P. (1974). Retrieval of low-frequency words from mixed lists. *Bulletin of the Psychonomic Society*, 4, 137-138.
- Ebbinghaus, H.E. (1964). *Memory: A contribution to experimental psychology* (H.A. Ruger & C.E. Bussenius, Trans.). New York: Dover. (Original work published 1885)
- Flagg, P. (1976). Semantic integration in sentence memory? *Journal of Verbal Learning and Verbal Behavior*, 15, 491–504.
- Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Clarendon Press, Oxford. New York, NY.
- Gentner, D. (1981). Integrating verb meanings into context. *Discourse processes*, 4, 349–375.
- Gregg, V. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall and recognition*. (pp. 183-216). New York: Wiley.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics: Vol. 3: Speech acts*. (pp. 41-58). New York: Academic Press.
- Henne, P. & Pinillos, N. (in preparation). Cause by Omission and Norms.
- Kahneman, D., & Tversky, A. (1973). On the Psychology of Prediction. *Psychological review*, 80(4), 237–251.
- Kintsch, W. (1970). *Learning, memory, and conceptual processes*. New York: Wiley.
- Knobe, J. (2010). The Person as Moral Account and Its Alternatives. *Behavioral And Brain Sciences*, 33(4). pp 353-365.
- Krosnick, J. A., Li, F., & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgment. *Journal of Personality and Social Psychology*, 59, 1140-1152.
- Lewis, D. (1996). Elusive Knowledge. *Australasian Journal of Philosophy*, 74(4): 549–567.
- Loess, H. (1967). Short-term memory, word class, and sequence of items. *Journal of Experimental Psychology*, 74(4), 556–561.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic Integration of Verbal Information into a Visual Memory. *Journal of Experimental Psychology*, 4(1), 19-31.

- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, 13(5), 585-589.
- Medin, D. L., & Schaffer, M. M. (1978). Context Theory of Classification Learning. *Psychological Review*, 85(3), 207-238.
- Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. New York: Prentice Hall.
- Orne, M. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 776-783.
- Owens, J., Bower, G., & Black, J. (1979). The "soap opera" effect in story recall. *Memory & Cognition*, 7(3), 185-191.
- Pinillos, N. (2012). Knowledge, Experiments and Practical Interests. in *New Essays on Knowledge Ascriptions*. M. Gerken and J. Brown (Eds). Oxford.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of experimental psychology*, 77(3p1), 353.
- Powell, D., Horne, Z., Pinillos, A., & Holyoak, K. J. (Submitted). Justified true belief triggers false recall of "knowing."
- 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.
- Thorndyke, P. (1976). The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 15, 437-446.
- Sachs, J. (1967). Recognition Memory for Syntactic and Semantic Aspects of Connected Discourse. *Perception & Psychophysics*, 2(9), 437-442.
- Schacter, D. L. (1995). Memory distortion: History and current status. In J. T. Coyle (Ed.) *Memory Distortion: How minds, brains, and societies reconstruct the past*, (pp. 1-43). Cambridge, MA: Harvard University Press.
- Schank, R. C. (1972). Conceptual dependency: a theory of natural language understanding. *Cognitive Psychology*, 3, 552-631.
- Schank, R. C. (1973). Identification of conceptualizations underlying natural language. In R. C. Schank & K. M. Colby (Eds.) *Computer models of thought and language*. San Francisco, CA: W. H. Freeman & Co.
- Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. *Advances in Experimental Social Psychology*, 26, 123-162.
- Schwarz, N., Strack, F., Hilton, D. J., & Naderer, G. (1991). Judgmental biases and the logic of conversation: The contextual relevance of irrelevant information. *Social Cognition*, 9, 67-84.

- Stanley, J., & Sripada, C. (2012). Empirical Tests of Interest Relative Invariantism. *Episteme* 9 (1):3-26.
- Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, 124, 272–283.
- Stillings, N. A. (1975). Meaning rules and systems of inference for verbs of transfer and possession. *Journal of Verbal Learning and Verbal Behavior*, 14, 453-470.
- Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, 103, 255–262.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, 26, 1-12.
- Waskan, J., Harmon, I., Spino, J., Horne, Z., & Clevenger, J. (Submitted) Explanatory Anti Psychologism Overturned by Lay and Scientific Case Classifications.
- Weber, S., & Cook, T. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, 77(4), 273–295.
- Williamson, T. (2002). *Knowledge and Its Limits*. Oxford.
- Wolfram Alpha LLC (2013). Wolfram|Alpha. [http://www.wolframalpha.com/input/?i="knew"&a=*C.knew-_*Word-](http://www.wolframalpha.com/input/?i=) (access March 3, 2013).
- Wolfram Alpha LLC (2013). Wolfram|Alpha. http://www.wolframalpha.com/input/?i=thought&a=*C.thought-_*Word- (access March 3, 2013).

Appendix I

Example vignette: Justified true belief condition

Gary Hawkins was a counselor who treated troubled youths with long histories of abuse. He was having an especially hard time getting through to two of his clients, a pair of fourteen year-olds named Will and Beth, who both seemed to dislike him. Most of Gary's clients grew up poor and were at-risk youths.

One morning, Gary was out for a jog in Millennium Park on the east side of Chicago. Gary's path ran under Columbus Drive, and when he entered the unlit tunnel his eyes were unadjusted to the dark. Suddenly, Gary felt a terrible pain at the back of his head and he fell to the ground. He hadn't seen the attacker waiting in the tunnel with a weapon in their hand. The attacker continued to hit Gary with the weapon, bruising his ribs and arms. Then the attacker ran off, and Gary laid in the tunnel, dazed.

Another jogger discovered Gary about a half an hour later and called the police. Detective Jack Dempsey was assigned to the case. Dempsey was a veteran detective who loved police-work, so he hurried to the hospital to interview Gary as soon as his doctors would allow it. Unfortunately, Gary was useless as a witness. He hadn't seen the attack coming, and the blow to the head had left his memory hazy. Next, Dempsey started to question Gary's clients, and Will really rubbed him the wrong way. Dempsey was immediately suspicious of him.

Dempsey wasn't the only one who disliked Will. Beth and Will were dating, and she suspected he was going to leave her. She wanted a way to get even with Will, and Will had told her a couple weeks before that he was planning to attack Gary in Millennium park.

Dempsey started his investigation and found several pieces of evidence that pointed to Will. First, another officer found Will's baseball bat near the scene of the crime. Then, Dempsey got a warrant and searched Will's phone, where he found texts bragging about beating Gary up.

Actually, Beth wanted to get payback for Will leaving her. She hoped Will would be caught for his crime. It sure looked like he was going to be. Will wasn't careful to cover his tracks after he attacked Gary. He left his baseball bat at the crime scene, and then he sent texts from his phone bragging about the attack.

After finishing his investigation, Dempsey wrote up his report for the district attorney based upon the evidence he had collected, including Beth's testimony. He worked on his other cases until Will's case went to trial. Whatever the ultimate verdict would be, Dempsey thought Will was guilty.

Dempsey tried not to worry about work and just look forward to the weekend. His daughter was visiting colleges, and they were flying to New York together to visit NYU. Dempsey had never visited New York before, and he really needed a vacation. It would be a good chance for a break, although he kept warning his daughter that Chicago's pizza was vastly superior.