

Verbal Reports in Consciousness Science

Literature Review

Frank P. DeVita

October 2026

Introduction

In this review, I examine some key pieces of literature on *verbal reports* as evidence in cognitive science and contextualize it with philosophical debates about memory, introspection, and consciousness. Verbal reports are a quick and easy means to collect data about consciousness. What is consciousness like? To find out, in the spirit of Williamson (2024), there is a sense in which we can *just ask* (188). However, just asking experimental subjects in neuropsychology experiments what their conscious experiences are like is a theoretical and practical mixed bag. Verbal reports are a reliable way to collect information about conscious experiences in experimental settings, however there are substantive worries about what kind of data they are, their integrity as data, and how they should be interpreted, all stemming from fundamental issues about their exact content, precision in capturing cognitive events, and overall reliability. Can subjects report on the phenomenology of ‘raw’ goings-on—occurrent, conscious mental phenomena and processes such as remembering—inside their heads? Can they only report *theories* of what those mental states and processes are or might have been based conjectures and interpretations arrived at by introspection? Neither? If subjects’ means of reporting their conscious experience is through introspection and subsequent verbalization, is introspective evidence glossed in

language reliable for doing science, or trustworthy at all? How much of the information in conscious experience is reportable to begin with, given material and mechanistic facts about cognition and the limitations of the nervous system?

This review aims to get clearer these questions, as well as practical ones about the collection and interpretation of verbal reports in consciousness research, and sharpen some methodological issues about techniques and norms for systematically constructing probes and eliciting verbal reports from experimental subjects, as well as interpreting and analyzing verbal report data in consciousness science. It focuses primarily on philosophical and psychological issues related to verbal reports. The reason for this lens is that discussion of verbal reports as empirical data about mental processes emerges in early experimental psychology literature, and is systematized there in a way that renders the worries of experimentalists concerned with the veracity, reliability, and role of such reports in scientific research connectable with the worries of philosophers of mind about the possibility of accurate introspection. The overall goal is to provide fair grounding in some of the relevant literature across both domains, and get clearer about what kinds of verbal reports should be elicited and when for the purposes of studying conscious experience, and its relation to other processes such as perception, and memory. Further, it develops a view on how to interpret and analyze verbal reports in systematic ways that can aid discovery and progress in the science of mind.

The nature of verbal reports as empirical data is addressed scientifically by theoretical work in psychology that offers a taxonomy of different types of verbal reports structured by their relation to, and influence on, cognitive processing (Ericsson and Simon 1980). This work sorts verbal reports by their level of *directness of access* to cognitive processes, and suggests that close-to-immediate verbalizations of ongoing cognitive processes provide the most reliable reporting of occurrently conscious mental phenomena (218-219). Philosophically, there is a conceptual basis for the relative *introspective authority*

this kind of view suggests that quells worries about the introspective reliability, and takes verbal reports as weighty evidence about perception, cognition, and conscious experiences generally speaking. However, there are also reasons to doubt that a view accepting introspective authority is at all tenable. Consider the systematic errors we may make in ordinary life—we mistake objects for one another, mishear utterances, and often explain the rationale for our actions in ways that did not occur to us in the moment. This review explores these views in tandem to establish a coherent way in which consciousness science can use verbal reports to reliably collect information about perceptual experiences and cognitive processes that reasonably controls for the possible complications brought about by introspection. In the perspective developed, it is in principle possible to be strategic about the types and timing of verbal report probes through intentional experimental design that raises the probability of collecting more precise introspectively grounded verbal reports about conscious experiences of cognitive phenomena including perception and memory in experimental settings.

The status of verbal reports as evidence in science traces to historical debates about introspection—the cognitive process by which one learns about one’s own ongoing or recent mental processes (Schwitzgebel 2024)—in psychological research. Critical discussions of introspection initially gained steam when psychophysicists began developing quantitative methods for studying the mind. Since they were looking for precise measures of processes such as perception and memory, they held that introspection was a specialized cognitive task, and therefore that significant training for experimental subjects (to the tune of 5,000+ trials) was required to introspect effectively and precisely discriminate phenomena such as perceptual changes in stimulus intensity (Boring 1953). More radically, some early psychologists held that it was impossible for an observant psychological being to observe itself (e.g., Comte 1830). Wundt (1883, in Schwitzgebel 2024) thought that the method of introspection was, at its core, “the method for holding

and attentively manipulating a memory image or reproduction of a past psychological process” and thus that introspection had real scientific value, but was disadvantaged due to the fact that (a) information in memory cannot discover anything new beyond what is occurrently remembered, and (b) that memories of processes, objects or other items might be easily confused with one another in reporting (Schwitzgebel 2024). For Wundt, cognitive processes were indeed accessible by introspection, but, if they were to be studied properly and objectively, had to be actively and skillfully monitored by subjects as they were occurring. In that way, Wundt thought, mental states and processes such as perception, memory, feeling, and volition could be reliably studied by analyzing the introspective reports of highly trained, expert subjects. For early psychophysicists, introspection was thus a viable path toward studying the mind, but only after subjects’ introspective abilities were finely tuned.

Looming always in early discussions of introspection in quantitative psychology were objections that uncontrollable subjective factors inherent in the mental act itself contributed to undesirable and unavoidable effects on the accuracy, believability, and confirmability of verbal reports from experimental subjects. With the rise of behaviorism in the early 20th century, disdain for introspective methods grew, especially with instance by behaviorists that psychology focus on behavior and behavior manipulation by intervention, which continued with the rise of functionalism and increased focus on the nature of cognitive processes in the latter part of the 20th century (Schwitzgebel 2024). The qualms about introspection from behaviorist and functionalist perspectives are valid and defensible. First, subjects might proffer verbal reports that aren’t in fact truthful to their occurrent cognition yet reasonably account for their behavior, and there is no systematic way to determine whether this is the case or not. Obviously, one cannot open up the heads of subjects to confirm that a proffered verbal report indeed tracks the phenomena it purports to describe. This lack of confirmability is a problem from general scientific and data

integrity and reliability standpoints. Further, subjects may proffer false reports due to factors as diverse as performance embarrassment, guessing or lack of confidence in decisions, misremembering what they actually experienced, or a slew of other possibilities, all of which would render verbal reports inaccurate, unreliable, or otherwise unbelievable—and thus unfit for robust scientific theorizing and analysis.

Nevertheless, frustrations with verbal reports as data are in important ways pressured by the needs of contemporary consciousness researchers, who require means to collect data about the phenomenology of conscious experiences in order to learn about the first person aspects mental phenomena such as perception and memory. Subjective data are essential for uncovering details about temporal and other dynamics of consciousness, e.g. when items become phenomenally conscious, how long traces of phenomenal experience remain in memory, or how lossy information available for recall from different memory systems is over time. With few other ways to generate data on what the character of conscious experience from the first-person perspective, verbal reports remain a necessary part of cognitive scientific research that aims to uncover specific facts about ‘what it’s like’ (Nagel 1974) to, for instance, perceive, remember, and think.

Verbal Reports as Psychological Data

The shift in psychology from aiming to understanding behavioral patterns and causes to aiming to understanding the mechanisms of the production and structure of those patterns via cognitive processes pushed researchers to look more closely at behaviors such as eye or bodily movements to explain cognitive performance. However, *verbal* behavior was discredited as too tainted by the woes of introspection, and thus deemed unreliable due to its lack of objectivity compared to behaviors such as motor movements.

Since there were informal and sometimes recalcitrant attitudes toward verbal reports as psychological data, strictness about their collection remained methodologically unformalized for much of psychology's early developmental period.

To treat verbal reports as data, there needed to be a measurement paradigm that tracks the processes that produce verbal responses. However, this is operationally difficult to do. Verbal reports are highly flexible—because of the flexibility of language there can be an infinity of probes and questions to subjects, as well as an infinity of responses from subjects to probes, making the task of systematic association seem insurmountable. Nevertheless, much progress has been made. Modern scientific work on verbalization as a neuropsychological data point flows from theoretical work by Ericsson and Simon (1980) that aimed to formalize verbal reporting for research purposes as *protocol analysis*. That work's major theoretical advancement was arguing that the *kinds of probes* used to elicit verbal reports, as well as their *timing* have a significant influence on how verbal reports should be both elicited and interpreted.

Ericsson and Simon put the recording of verbal information on a par with eye or bodily movement data at the level of “encoded patterns or information content” (217). Their view asserts that that verbal reports are data from the perspective of information processing, and presents a model for how subjects verbalize, on cue, information in short term memory that predicts what cognitive processes can be reliably reported. Crucially, their model implies that inaccurate reports are characterized by *inference*. The main idea is that accurate verbal reporting of cognitive processes inversely correlates with theory-ladenness on the part of the subject—essentially, the less theorizing subjects do in reports the better. Ericsson and Simon's account of verbal reports also implies that verbalizations made in times closer to perceptual events or neuropsychological stimuli will be more reflective of ‘raw’ cognitive processes, while reports at later times post-stimulus will be increasingly subject to additional information and cognition that affect what is

reported. They argue that using verbal reports as experimental data are no different than other reporting methods such as hand or eye motor movements in that they must be intentionally designed to target the mental phenomena of interest, and rule out alternative explanations.

Ericsson and Simon taxonomize *kinds* of verbalizations (1980, 218; Figure 1) based on whether they are *concurrent* with task performance or *retrospective* after task performance, since the interval between acquisition and recall is significant in the classification of both probes and reports:

Figure 1: Kinds of Verbalizations

Verbalization as the primary task	Verbalization as the secondary task
Concurrent with performance	Retrospective
Reporting specific events	Reporting generalizations

Concurrent verbalizations are elicited while a subject is actively attending, while retrospective verbalizations are elicited when perception or cognition occurred at an earlier point in time. According to Ericsson and Simon, *which kinds* of verbal reports are being dealt with as data thus bears on the way those reports should be understood. Depending on intervening information a subject might receive, or representations they may have in the time between task performances and verbal reports, the nature of report itself will be affected. This requires researchers to interpret reports differently based on the specific conditions in which the report was uttered.

Ericsson and Simon also taxonomize *levels* of verbalization, delineated by their directness and the intervention of cognitive processes between acquisition and recall (1980, 219-220; Figure 2):

Figure 2: Levels of Verbalization

Level 1 (Direct): Immediate Verbalization	Information produced in the form in which it was acquired during task performance
Level 2 (Encoded): Articulation from Non-Verbal Information	Information is encoded into descriptions that could be communicated to another person, using compact format and idiosyncratic referents.
Level 3 (Encoded): Inference	Scanning and filtering processes intervene to describe attended content, or to describe features of a scene that a subject would not have normally attended to

The difficulty, especially with higher levels of verbalization, is that experimenters may be interested in motives, reasons, or processes to which subjects may not have conscious access to at the time they are asked to make a report. For instance, subjects are often asked to retrospectively generalize over cognitive processes tens of minutes after they perform a task, and in some cases—especially in tasks involving probability (Tversky and Kahneman 1973)—experimental subjects are no better than observers at reporting cognitive processes relevant to their behavior (Nisbett and Wilson 1977)

Further, Ericsson and Simon delineate different kinds of *probes* (1980, 220-222; Figure 3) that themselves influence the content of verbal reports by way of allowing intervening representations or cognition, providing additional information, or other factors:

Figure 3: Types of Probes

Think or talk aloud	Instruction to verbalize during performance for a direct trace to recalled information
Probe for specific Information	Report of hypotheses or strategies used to accomplish an experimental task involving concept or discrimination learning
Retrospection: Interpretive probing	Subjects are probed at the end of an experimental session
Probing for hypothetical states	Asking subjects about counterfactuals, e.g. how they would have behaved had some information not been presented
Probing for general states	Questions asked and answered without reference to the experimental context
Specialized probing	Investigator asks for specific information with experimental context and constrains responses

Probes allow experimenters to systematically collect information about conscious states, strategies, and recall, however they are not themselves without complications that experimenters must be aware of. The ways probes are constructed, including language choice and instructions given, used to collect verbal reports will affect what cognitive processes and information the reports are ultimately about, or their *intentionality*, emphasizing the need for careful and strategic design based on the cognitive phenomena of interest. This is especially relevant in memory and consciousness research because the kinds of probes selected and used can be the difference maker for whether verbal reports reflect occurrent cognitive processes such as recall or perceptual discrimination, or reflect subjects' own theories and inferences about the mental underpinnings of their behavior.

Verbal reports are further complicated by the different ways subjects might arrive at descriptions of their own perceptual or cognitive processes. To enumerate a few possibilities, subjects may:

1. Be aware of general procedures or programs and can report them directly without reference to specific behavior
2. Remember specific parts or episodes and attempt to generalize these in reports
3. Remember specific tasks, then use this as information to infer procedures
4. Use knowledge of what one ought to do in a procedure to generate a report about their strategy

(Ericsson and Simon 1980, 221)

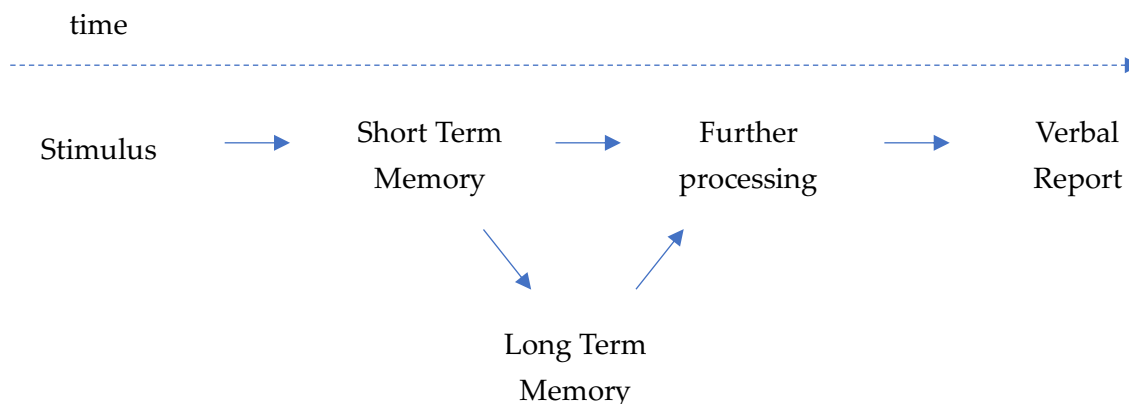
There is particular difficulty with reliance on verbal reports elicited with retrospective, interpretive probing that has subjects speculate or theorize about their cognitive processes rather than report 'what's going on in their mind'. These kinds of probes and reports represent a serious challenge to interpretation. As Ericsson and Simon put it, "there is no reason to suppose that the subjects themselves will or can be aware of the limitations of the data they are providing...moreover, the variety of inference and memory processes that might be involved in producing the reports make them extremely difficult to interpret or to use as behavioral data" (221). Even with probes of the specialized variety asking for specific information, there remain experimental risks that subjects will make reports about high-level procedures and strategies instead of occurrent cognitive processes, and further, that experimenters may ask about information that a subject cannot access. The provision of contextual information may help with recall, but it does introduce still

further complications related to the influence of additional and supplemental information, long-term memory, and contextual reasoning processes on the content of verbal reports.

Considering the plethora of complications involved with verbal report data, there is a great need for systematization, and more basically, for clarification of the relevant relations between verbal reporting, behavior, cognition, and consciousness, and further, the role that intervening variables such as time, probe type, and supplemental information play in the determination of the content and validity of verbal reports. Ericsson and Simon's *Processing Model* (1980, 222-224; Figure 4) sorts some relations between verbal data and behavior, and requires two assumptions:

1. Human cognition is information processing.
2. Information is stored in several memories with different capacities and accessing characteristics: short term (STM) with limited capacity and immediate duration, and long term (LTM) with large capacity and permanent storage but with slow fixation and access times.

Figure 4: Processing Model of Verbal Reports



Ericsson and Simon assume that any verbalization of a cognitive process must be based on information in a subset of long term and short term memories, determined by time of verbalization relative to cognitive processing and the degree of intermediate processing. On the Processing Model, verbalizations *externalize* the content of short-term memory (STM). If the information in STM is not already verbally encoded, verbalization requires verbal encoding before reports can be uttered. The latter kind of processing takes time. Connecting the model back to probes, Ericsson and Simon argue that *thinking aloud* specifically cuts out any intermediating processing and will, on the model, be a reliable reporter of the contents of STM, while any longer time intervals between cognition and report will yield a theory-laden or generalized, *retrospective* report grounded in long term memory representations. As Ericsson and Simon put it, in retrospective reporting, “retrieval may not only fail to access previously heeded information but may on occasion access information that is confused with the events being queried, and hence is inaccurate” (226). Also on the model, probes that engage long term memory will not elicit reports about *occurrent* cognitive processes. The intervening time and cognition between stimulus/performance and report necessarily colors the verbal report with a subject’s theories and inferences—*their* interpretation—of the grounds of their behavior, and thus may not accurately bring out the details about occurrent cognition or conscious experience that cognitive scientists and consciousness researchers are after.

On the Processing Model, the important determining factors in the generation of verbal reports are the kinds of information that can be retained in short term memory, shuttled between long and short term memory, and the conditions for accessing and encoding that information verbally. The main conclusion that Ericsson and Simon draw is that instructions that require subjects to *recode* information in order to report it may affect the cognitive processes under study. It also may be the case that as subjects gain expertise in a task, less information about their cognitive processing may be available to them as

task performance is lifted to the algorithmic level, or subjects enter flow-like states. If the Processing Model is right, then the longer an interval between cognition and verbal report, the less likely the report is to be about an occurrent process or the content of short term memory, and the more likely it is about some information in long term memory or a generalized gloss of cognitive processing amounting to a strategy or higher level narrative the subject has constructed. One concern about the idea that subjects should be probed during task performance is that verbalization tasks would somehow impact cognition. However, a thorough review of empirical results, Ericsson and Simon claim, does not show this—and in fact rather shows that task performance by subjects is not negatively impacted by intervening verbalization requests (1980, 226-235). This is good news for verbal reporting experiments since it reduces the risk that an occurrent verbalization component in a study design would make experimental tasks significantly more difficult for experimental subjects.

The Processing Model makes some intuitive predictions. While instructions for subjects to verbalize aloud don't seem to affect cognition, the model predicts that instructions to provide *reasons* for actions would alter cognitive processing insofar as inferences have to be made and narratives constructed (227). For the verbalization of visual processing, the Processing Model entails an extra step that converts visually encoded information to verbally encoded information, which can affect cognitive processing (*ibid.*). Those effects, Ericsson and Simon hold, include slowing down main task performance, changing the structure and course of performance of the main task, and influence on what is remembered in retrospective recall (228-235). These effects will scale with the complexity and difficulty of the verbalization queries presented to subjects.

After Ericsson and Simon, the discussion of verbal reports as data continued, with psychological researchers beginning to see their scientific value and unique access to

phenomena of interest. For instance, it was argued (by Perone 1988; in Luque and O'Hara 2016, 160) that they were in fact useful for aims such as:

1. Observing behaviors with no other means of observation
2. Collecting data on behavior defined in terms of verbal responses, such as assessments of attitudes and beliefs
3. Evaluating the content of verbal reports and its behavioral role
4. Explaining causal roles of the content of verbal reports in behavior

Interest in verbal reports grew through the 1990s as a result of increasing general interest in verbal behavior, but their limitations were again quickly surfaced, this time at a practical level. Worries proliferated about both retrospective verbal reports elicited at the end of experiments and concurrent ones elicited during experimental tasks, albeit for different reasons than those pointed to by psychophysicists and dealt with by Ericsson and Simon's methodological formalizations.

For retrospective reports, it was thought that, since they are under the control of the experimenter's questions, they are *artificially selective*. The content of subjects' reports would not have occurred to them had the experimenter not asked, compromising the possibility that the report accurately tracks a genuine memory trace of a process that was cognitively registered by the subject when it occurred. In a phrase, such reports could be *post hoc* or genuine, with no surefire way to tell which. Retrospective reports could also not be systematically or causally linked to occurrent internal verbalizations or other cognitive processes, diminishing their explanatory usefulness. Further, retrospective probes might themselves be ambiguous both in their construction and to subjects insofar as they might refer to, and therefore elicit reports about, single timepoints in an experiment or

summation of a performance. Lastly, verbal behavior could not be systematically related to task performance, making it, for many researchers, irrelevant from a behavioral standpoint (Luque and O'Hara 2016, 160-161). Intra-experiment verbal reports were worried to have their own sets of limitations, including risk of disrupting task performance, introducing changes to task performance by the provision of hints or information about what the experiment is looking for, and the possible irrelevance of verbal reporting to performance itself.

Optimists about verbal reports, however, see these limitations not as evidence that verbal reports have no scientific use, but rather that there are deeper questions on how verbal reporting should be used in research. This optimism led Simon and Ericsson to, from their initial ideas about verbal reporting, to develop a mature version of protocol analysis (in Ericsson and Simon 1993), the bones of which was described in detail above. The refined protocol analysis is meant to address some of the worries and methodological limitations of verbal reports also already mentioned, especially issues related to experimenters as an audience. The key to the refined protocol analytic approach is that the subject is asked to verbalize while performing a task as if she is alone and narrating her cognitive activity. Those responses are recorded during tasks, then transcribed, broken down, and analyzed further to build cognitive models or for other purposes.¹ Though these models get to a cleaner approach to verbal reporting for the analysis of behavior, they fall short for the purposes of cognitive processing research because protocol analysis still leaves no way to correlate verbalizations with internal processing.

¹ It is worth observing that protocol analysis looks a lot like a scientific adaptation of the logical positivist framework for analyzing experience that involved 'protocol sentences' originally constructed by Otto Neurath in his *Protocol Sentences* (1932) in Ayer (1959, 199-208). The logical positivism program ultimately failed in constructing a systematic first-order language that captured experienced relations in a systematic way. I'm unclear whether one ought to take this historical fact as a bad omen for protocol analytic methods.

These shortcomings led to the development of the “Silent Dog” method for verbal report collection favored by cognitive psychologists, which, practically, requires that behavioral tasks be run in both silent and think aloud conditions, then compared to determine whether inferences about mental states and processes can be made from a subject’s verbal reports.² Theoretically, the Silent Dog method essentially adds some methodological assumptions to the basic methods protocol analysis: (a) that performances in experimental tasks are rule-governed, and (b) that public verbal reports are functionally equivalent to private verbal behavior. These assumptions lead to the following ‘controls’, essentially further assumptions based on (a) and (b):

1. Subjects are saying aloud what they are saying privately, so saying aloud will not affect task performance.
2. Disrupting verbal behavior must affect performance, so a lack of differences in performances with and without disruption is attributable to verbal behavior
3. Probes (protocols) should be presented to other subjects and affect behavior in a systematic way, so that their content is task-relevant.

(Luque and O’Hora 2016, 167)

The critical move in the Silent Dog method is to ensure *functional similarity* between public and private verbal reports and self-rules by running tasks in talk aloud and silent conditions, and comparing the results. The key concept is that *if* behavioral responses in think aloud conditions *match* behavioral responses in silent conditions, then this licenses an

² On the reason for the name: “These controls were named the silent dog method, after the Sherlock Holmes’ novel “The Silver Blaze” (Doyle, 1892), in which the famous detective knew the identity of a murderer because the dog in the house had not barked” (Luque and O’Hora 2016, 167).

experimenter to make valid inferences about the mental states and processes of the subjects proffering the reports from the data of the reports themselves.

With the controls and constraints of mature protocol analysis and the Silent Dog method in hand, verbal reports might be seen in a more favorable light than their less than illustrious past might suggest. However, do additional theoretical and methodological constraints avoid the worries of the early psychophysicists about introspection? Also, what ought we think of the idea of *functional equivalence* between private and public language? To the first point, it seems that experimental controls don't quite answer more foundational worries about the reliability of introspection. To the second, it is unclear whether the functional equivalence assumption itself is warranted based on behavioral evidence alone. A subject could behave equivalently in two situations with different thoughts in her head, yet appear to pass the smell test for functional equivalence. These preliminary doubts ought to at least shake one's confidence that verbal reports, even constrained in various ways, can be completely trusted—an issue to which we now turn.

Trusting Verbal Reports

Studying subjective experience is difficult no matter the approach one takes. This fact brings out the reason why verbal reports have a unique and important role in in consciousness research, or any other research aiming to understand what goes on inside the head from the outside. Verbal reports are one of the only types of collectable data that can capture information about what it's like to perceive, remember, learn, and so on. This what its likeness, or *phenomenality*, is impossible to measure directly. Not even the most advanced experimentalist can detect, for instance, what it's like to see red, hear middle C, or remember what happened five minutes ago 'from the data' as it were. Any attempt

to do so seems to require prompting and querying of the experiencer, and her proffering of experiential reports from her perspective.

Subjective reporting may include verbal responses to open-ended questions about experiences during a controlled experiment, e.g., “What was it like to remember the red square you were shown?”, or behavioral responses to specific queries or closed-ended questions during experimental task performance, e.g. moving a cursor or joystick to the location of a perceived stimulus, or rating a stimulus on a scale. Psychophysical tasks can be set up to elicit behaviors that experimenters can then reasonably *infer* are grounded in specific conscious experiences, however, experimenters must nevertheless make inductive moves in accounts of how behaviors systematically involve particular types or tokens of conscious experiences. Based on our own ordinary experience, these inductions are indeed *licensed*—trivially, we use evidence in consciousness to act—but nevertheless one might question how to *confirm* the ways in which experimental subjects are conscious of stimuli and use that data in their consciousness to act. It’s possible that subjects may be performing tasks like matching cues to colors or locations by methods and processes postulated by experimenters, yet they equally well might be guessing or using unorthodox cognitive strategies to complete such tasks that aren’t obvious from outward behavior. One of the only ways to know which is in fact the case is to ask subjects to report their experiences, and to *trust* the content of those reports.

How trustworthy are verbal reports? The answer depends on where one stands on deeper questions about what makes verbal reporting possible or not. Philosophers and scientists differ greatly on the concept of a *reliable narrator*, that is whether verbal reports come from an authoritative point of view on what happens in the head. The *heterophenomenological view* advanced by Dennett (1982) holds that the subject of experience is a fully reliable, genuine author capable of detailed verbal reporting of conscious experiences, and further that the reports themselves are interpretable *as a text*. The

heterophenomenological view entails taking up the *intentional stance* toward subjects as having beliefs and desires for particular reasons, and putting them, by explicit direction and experimental design, in a position to desire to report whatever they believe about their own experiences truthfully and accurately. Dennett outlines the basic approach for studying consciousness empirically given these theoretical assumptions:

“...we tell our subjects that what we want to hear is whatever they believe, and we take care not to let them know what it is we hope they believe. We do what we can, then, to put them in a situation where, given the desires we have inculcated in them, they will have no better option than to try to say what they in fact believe.” (Dennett 1982, 162)

This controls for cases in which subjects might say what they think experimenters want to hear, what will make them look like good performers, or anything else other than direct evidential reports about consciousness. That is, heterophenomenology *incentivizes* a subject to veridically communicate the facts about their conscious experience by reducing noise that might come from, say, competing desires to be perceived as competent, or not feel embarrassed. Any kind of covert or conflicting attitude related to the rules or expectations of an experiment could in principle confound a verbal report, so removing those kinds of influences at the outset of data is crucial. Dennett does so by inculcating desires to communicate about consciousness veridically.

One might object that although an experimenter may try hard to create conditions under which subjects feel compelled to veridically communicate their conscious experiences, it nevertheless remains possible that subjective factors may systematically affect or distort verbal reports despite those efforts. Despite these possibilities, it is precisely the ethos of heterophenomenology to trust verbal reports entirely, then consider how their contents square with theories and other data to make discoveries about the nature and

dynamics of consciousness. On this view, verbal reports elicited in the right experimental conditions are *authoritative* about the contents and dynamics of conscious experience, and it is up to investigators to parse and interpret verbal reports as uncorrupted data, refitting theories and frameworks to accommodate new findings from verbal data.

Even with explicit direction to report on the features of conscious experience, there is a deeper question as to whether subjects are in principle capable of accurately reporting on their experiences at all, since even an intentionally motivated subject will report how things *seem* to her, which itself might be a gloss, interpretation, or otherwise indirect accounting or recounting of what happens in the head occurrently, bringing us right back around to the longstanding issue of introspective reliability. However, as emphasized, the thrust of the heterophenomenological approach is to bracket both skepticism about introspection, and any presupposed distinction between how things seem to a conscious subject and how they really are phenomenologically. From the heterophenomenological perspective, how things really are phenomenologically just are the ways that they appear to conscious subjects. Doubting the accuracy of verbal reports simpliciter is distinctively disallowed by the view. However, one might think that there are independent grounds on which to doubt such a radically trustworthy position.

Sperling (1960) illustrated that subjects' ability to report on their conscious experiences is constrained by the amount of information presented to consciousness and the capacities of memory systems. In his landmark experiments on iconic memory—the fast-decaying store of visual information or *sensory register* (Di Lollo 1977)—subjects were shown a 9x9 array of letters and were asked to recall what they had seen. Sperling found that subjects shown the array reported seeing all the letters in the array, but could only accurately report a subset of the specific letters when prompted. This result was relatively unsurprising to psychologists since all memory is *lossy*, however Sperling also paired each 3-letter row with an audible tone, then cued recall for the letters composing the rows.

In this condition, subjects were, surprisingly, able to report the specific letters in each row with high accuracy on cue. This result suggested that information entering perceptual consciousness can be recalled with precision, albeit to a limited extent constrained by capacity limitations of memory systems. Iconic memory has a capacity of a handful of items, accounting for both subjects' increased accuracy in recall of smaller lines of information on cue, and their inability to recall entire large arrays of letters. Given Sperling's data, the accuracy of verbal reporting about conscious experience is at least a function of the *amount* of information initially presented to a subject. So while not *completely* authoritative, subjects do have access to at least a partial subset of veridical information about their experiences for verbal reporting.

Block (1995, 2007, 2011) interprets these results as exemplifying the distinction between phenomenal, or *p-consciousness*, and access, or *a-consciousness*, and advanced the idea that these results show that *p-consciousness*—what we are perceptually conscious of—contains more information than what is *a-conscious*, or accessible for actions like verbal or behavioral reporting. To rehearse Block's iconic assertion, "*perceptual consciousness overflows cognitive access*" (Block 2011, italics added). On Block's interpretation of Sperling's results, *p-consciousness* is rich and information-dense, while *a-consciousness* is comparatively sparse since it can only cite a subset of information that was perceptually conscious to a subject. This means that only a subset of details about conscious perceptual experiences is reportable from iconic and short-term memory, which itself rapidly decays. This in turn raises legitimate concerns about the content and accuracy of verbal reports as data about consciousness. If perceptual consciousness is rich and access consciousness is partial, and rapidly-degrading iconic memory retains the most information about immediate perceptual experience, then verbal reports are necessarily highly lossy compared to occurrent perceptual experience and thus largely inaccurate. Contrary to what was suggested at the outset of this review, we may, given the issues discussed, be

unable to *just ask* about consciousness to collect data about it, since it may be the case that what is verbally reportable about conscious experience, perception, or memory necessarily falls short of capturing what it's like to occurrently experience, perceive, and remember.

Which account of the relation between consciousness and reports is right has broad practical ramifications for consciousness science, particularly in approaches for collecting and analyzing verbal report data from which we hope to glean something about the nature of perceptual experience, memory, and other cognitive processes. If Dennett is right, then interpreting verbal reports consists in focused analysis of the reports collected taken at face value, and we ought to interpret verbal reports as we would any other spoken or written sentence that we would have no reason to otherwise suspect is false or misleading. On the heterophenomenological view, verbal reports are unendangered by the concerns about memory degradation, and the sparsity of access consciousness—Dennett would have us take it that verbal reports are *fully authoritative* about the happenings of conscious experience. If we treat verbalizations about consciousness as an authoritative text, we ought to analyze them on the assumption that thinkers are fully conscious of, and have full cognitive access to, the goings on in their heads. On this view, conscious experience, and reports about it, are both rich, and there are no overflow worries to speak of. If Block is right, then interpretation of verbal reports requires hard thinking and theoretical machinery. If perceptual consciousness overflows cognitive access, verbal reports are necessarily *impoverished*, so there ought to be attention paid to how and when verbal reports are elicited, and interpretive heuristics developed for analyzing them to uncover facts about consciousness in experimental research.

Which view about the relation between consciousness and reporting one takes on board matters for interpreting the content of verbal reports, and their status as empirical data. Can a verbal report accurately communicate when one became conscious of some

stimulus? What kind of questioning could reliably bring out such details? Would an experimental subject know the difference? Whether perceivers are introspectively authoritative, and whether phenomenal consciousness overflows cognitive access matter for answering these questions, and further, for constructing a coherent picture of relations between perceptual consciousness, memory, and reportability. Empirically, these issues matter for understanding what verbal reports are about, how to elicit them in ways that aid subjects in targeting occurrent mental phenomena of interest, and how to interpret verbal reports effectively. Theoretically, these issues matter for taking a stance on whether verbal reports can supply researchers with reliable data at all, which we'll consider next.

Introspective Inference and Skepticism

Beyond issues regarding what subjects are reporting when they verbalize, i.e. subjectively glossed seemings-to-them vs. “real” features of the phenomenology of types of conscious experiences, a deeper concern about the dynamics of introspection raises concerns about whether verbal reporting can *ever* be accurate. Introspection is “a means of learning about one’s own currently ongoing, or perhaps very recently past, mental states or experiences” (Schwitzgebel 2024). We introspect when we “observe” our own thoughts, making it a complex metacognitive activity that is both deeply revealing as mode of access to internal states, but is also subject to distortions and interference as discussed above. On theoretical grounds, it’s not at all clear that the introspections that ground and make verbal reports possible can be trusted as data since there are no surefire ways to differentiate between introspective reports as dry, unadorned communications of raw mental processes, or as post-hoc narratives or personal theories about those mental processes prompted by querying.

It is difficult, on this line of reasoning, to know if introspection a window into the workings of the mind or an *inference* about them. On another line of reasoning, there is solid ground on which to assume the existence of introspective authority. This line, expounded by Davidson (1984) holds that though it is possible to be mistaken about the happenings and content of our own minds, such cases are non-standard, since we can be confident that linguistic and social norms will lead individuals to talk about their own internal states in ways that are both coherently interpretable and true. This introspective and interpretive confidence, Davidson argues, does not come from being error-free or introspectively infallible like it does on the heterophenomenological view, but rather from introspection being grounded in a perspective that is only accessible first-person subjectively by a conscious subject, and thus *unchallengeable*:

“Because we usually know what we believe (and desire and doubt and intend) without needing or using evidence (even when it is available), our sincere avowals concerning our present states of mind are not subject to the failings of conclusions based on evidence. Thus sincere first person present-tense claims about thoughts, while neither infallible nor incorrigible, have an authority no second or third person claim, or first person other-tense claim, can have.” (16)

For Davidson, we know our own mind because no one else can, and for that reason must be considered an authority on our own conscious experiences. If Davidson is right, then conscious subjects, though constrained by informational and cognitive limitations discussed above, are the *bona fide* best source of phenomenological evidence an experimenter can get. By Davidson’s lights, there is no outside individual that could argue against an experimental subject’s own claims about her experience, whatever their grounds may be.

Though one may find some comfort in views like Davidson's and Dennett's, it is hard to resist the force of the evidence from Sperling's experiments and Block's arguments that though we are fully and richly conscious of the world around us, our ability to report about it impoverished by comparison, and may be systematically mistaken or erroneous in certain respects. A supporting philosophical for this kind of view and its contrastingly radical conclusions comes from Schwitzgebel (2011, 2008, 2007) in his skeptical arguments against introspection, and by extension, introspective authority. In Schwitzgebel and Hurlburt (2007), Schwitzgebel highlight issues in the translation of one's inner experiences into language, and doubts that verbal reports can ever be interpreted at face value, thinking that no reconstruction of conscious experience via language will be theory or inference-free, and will be confounded by unsystematic and problematic uses of metaphor to describe inner experiences, for example, as being film-like or photographic.³ In general, Schwitzgebel argues that it is nearly impossible for a conscious subject to reliably distill the chaos and non-linearity of occurrent thought and conscious experiences into linear verbal descriptions that can be interpreted as precise or accurate. This position derives from his (2008) broader philosophical skepticism about self-knowledge, on which he completely diverges from the likes of Dennett and Davidson, holding instead that conscious subjects are broadly inept at introspection:

"We are both ignorant and prone to error. There are major lacunae in our self-knowledge that are not easily filled in, and we make gross, enduring mistakes about even the most

³ Schwitzgebel (2022) provides evidence from the history of science supporting that people do indeed systematically use metaphor in introspection and verbal reporting. Before the 1950s, experts assumed that people dreamed in color. With the rise of black and white film however, experts shifted to believing that dreams were black and white based on rising verbally reported evidence that subjects were dreaming in black and white. By the 1960s, when color film was on the rise, reports of dreams in black and white fell, and reports of dreaming in color increased. For Schwitzgebel, this illustrates that our introspections are wildly unauthoritative, and in this case, can be shaped by vogue technologies.

basic features of our currently ongoing conscious experience (or “phenomenology”), even in favorable circumstances of careful reflection, with distressing regularity.” (247)

For Schwitzgebel, introspective errors are systematic and pervasive, extending across perceptual and cognitive domains. In the case of emotion, he holds that we barely have a grip on answers to questions about, for example, whether we are happy or not or angry or not because our judgements about these states there are significant ‘distortive factors’ at play, such as self-conception and emotional involvement that undermine those very judgements (252). Schwitzgebel extends his skepticism to basic visual experiences by offering that our systematic introspective inaccuracy is demonstrated by the utter lack of clarity we have about details of features in the peripheral visual field even a few degrees from the foveal center, a phenomenological fact that is often surprising to the average introspector.⁴ How could such a pervasive and systematic visual character be missed by ordinary perceivers? Schwitzgebel sees it as a manifestation of our impoverished ability to reliably introspect even the qualities of our own visual field. Going further still, he extends his skepticism again to our ability to introspect what it’s like to think, challenging us to point to some specific phenomenological feature of occurrent thought that disjoins from things like mental *imagery* or discursive representation in chains of *verbal* thought that transpires without either. This radical failure of introspective ability across domains leads Schwitzgebel to generalize his arguments to the skeptical conclusion that we are, “prone to gross error, even in favorable circumstances of extended reflection, about our ongoing emotional, visual, and cognitive phenomenology” (259).

⁴ To see this, place an object outside the center of your field of vision into the periphery and observe how difficult it becomes to reliably discern its color or shape. Introspective reliability about basic features of objects rapidly degrades with even small magnitudes of de-centering. Schwitzgebel uses the example of a playing card, whose color and suit can barely be identified outside the foveal center to illustrate the point.

Accompanying his introspective skepticism, Schwitzgebel (2011) takes the line that what we intuitively think of as introspection and inner observation of occurrent thought is in fact *retrospection* insofar as it is an examination and analysis of an immediate memory of a conscious event. On this view, any introspective report, verbal or otherwise, is a function of memory, and as such, is not a form of direct access to the features or details of conscious experiences but rather a reconstruction influenced by a slew of intervening factors and information ranging from cultural embeddedness, semantic habits, self-conception, environmental situations, schemas of different kinds, and logic. This puts subjective reports on perhaps unescapably shaky grounds because:

“...stimulus error, memory error, and the potential subtlety of the target experiences make even carefully collected retrospective reports difficult to interpret...Any concurrent self-report method—*any method that asks the participant to report on his experience after it occurs*—will be polluted by the introspective act itself. Any retrospective self-report method will invite concerns about short-term memory at least...no self-report method can, I think, effectively avoid [these] sorts of concerns.” (113, italics added)

If Schwitzgebel is right, then any attempt to collect data about consciousness by introspective report is futile. Ought we share his skepticism? Even if it is right that “it is challenging to evaluate the accuracy of introspective reports not only when the relationship between the stimulus and experience is likely to be complex and variable...but even sometimes when [it] appears to be comparatively simple and dependable” (*ibid.*, 90), there remains, in my view, a middle ground we might seriously consider in consciousness science for using verbal reports as a way to make discoveries about conscious experience under different conditions.

Controlling for Noise: Time and Content

Is there a middling view between the naïve authoritativeness of the heterophenomenological and epistemic priority views of Dennett and Davidson on the one hand, and the pervasive introspective skepticism of Schwitzgebel on the other that squares with Sperling's experimental evidence for iconic memory, and Block's argument that phenomenal consciousness overflows cognitive access? I think there is, and it involves seriously considering the issues about introspection that provide reasons to doubt the *unconditional* acceptance of verbal reports as evidence. It also, however, critically involves recognizing that consciousness science cannot operate without dealing with the problem of introspective unreliability on theoretical and practical levels. Without querying conscious subjects about what consciousness is like, how else is one to collect data about the phenomenon? There are some efforts to find and enrich properties that may be operationalized as 'markers of consciousness' (see Bayne et al. 2024, and Koch et al. 2016 for scientific issues, and Irvine 2013 for philosophical issues), but this will only get us so far. To collect data about the phenomenology of perception and memory, and how it relates to behavior requires, at least at present, prompting subjects.

Schwitzgebel may be right about the indeterminacy of translation (cf. Quine 1960) between phenomenology and language, and we have good reasons to doubt that a precise translation between conscious experience and verbal reports is possible. However, the existence of the indeterminacy doesn't preclude the act of translation entirely. It can be *lossy*. As two languages may be inter-translated, so can the idiolects of inner experience and thought and public language, *if* we are willing to accept a reasonable degree of information loss between the two formats. We can take on board a healthy dose of skepticism about introspection by saying that introspective reports are not always reliable, and often

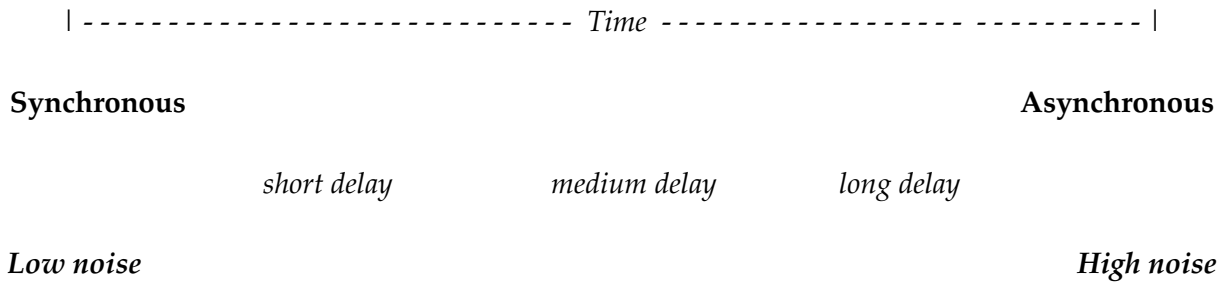
distorted by intervening information, co-occurrent mental states and processes, environmental conditions, and a slew of other variables we can group under the heading of *noise*. While noise is a fact of the situation, it need not torpedo the entire enterprise of getting at consciousness through verbal reports. Despite the inherent noise that may preclude the *absolute* authority of verbal reporting that some theorists want, or even demand, one can imagine ways to minimize and control it.

First, the positioning of prompting about the features, dynamics, or any other aspects of conscious experience in experimental study design must be strategic. If Ericsson and Simon are right, the longer an experimenter waits to elicit a verbal report, the more likely that report is to be noisy — that is, affected by intervening information and processes from different subsystems. Thus, there is a continuum along which the *placement* of a verbal report in time varies with its *content*. However, the existing terminology of “concurrent” and “retrospective” verbal reports stands in the way of clarity.

Let us understand that “concurrent” verbalizations, which indeed may be unavoidably retrospective as Schwitzgebel argues, are those verbal reports elicited while a subject is actively attending to a stimulus. Instead, call these reports *synchronous*. One might define a millisecond threshold after which reports are no longer considered *synchronous* depending on their experimental context, but the important part is drawing a line for motivated theoretical or empirical reasons. Synchronous verbal reports have lowest degree of *retrospective distortion* one might hope for, since they are, by Ericsson and Simon’s analysis, the least ‘contaminated’ by noise. On the other hand, “retrospective” verbalizations are elicited sometime after a subject was attending to a stimulus in intervals that may range from seconds to minutes. Instead, call these kinds of reports *asynchronous*. One might taxonomize asynchronous reports more finely, say by short, medium, and long *delay* designations proportional to the delay time between initial stimulus and the eliciting of a verbal report about it. The continuum just described (Figure 5)

allows for a coherent sorting of verbal reports based on their proximity to a stimulus, and, combined with idea that more proximally elicited verbal reports will be less noisy. Thus we have the idea that verbal reports can vary along a *temporal* dimension associated with their delay from the initial stimulus.

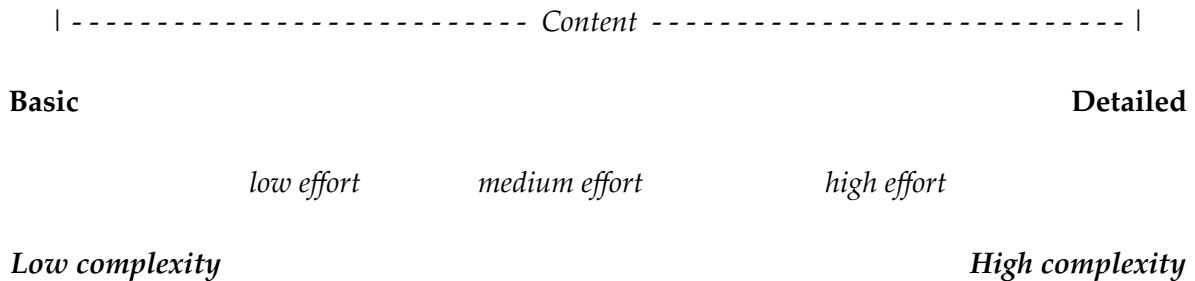
Figure 5: Temporal dimension of verbal reports



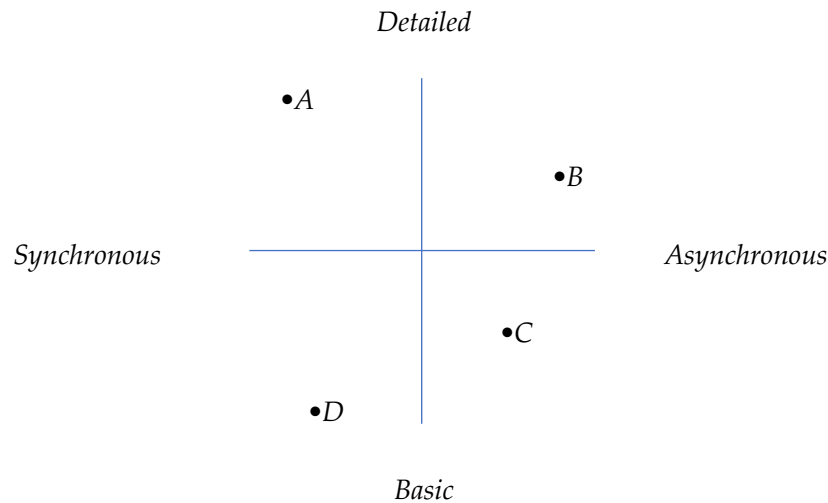
A second dimension we might consider is one for *content*. Though it is a bit of a trickier notion, we might think about content along a dimension of *complexity*. Call content about mental states *basic* when it is about the ordinary perception of objects and properties. Things like seeing a color, recognizing an object, or processing a piece of language syntactically are basic in this sense. Accordingly, call reports on these kinds of features *basic*. Let's also make a theoretical assumption that basic verbal reports can be made using a relatively modest number of words. If basic reports consist in 'raw' identifications, they would by definition not mention higher order cognitive states, processes, or items such as memory/ies, inference/s, reasoning, or cognitive strategies. Call reports that involve

this kind of more complex information *detailed*. Let's also stipulate that these kinds of reports take, on average, more words to articulate clearly. We can further taxonomize detailed reports, say by low, moderate, and high *effort* designations proportional to the volume of words a subject generates in response to a prompt, probe, or query, drawing on Ericsson and Simon's (1993) more mature articulation of their approach. The continuum just described (figure 6) thus allows for a coherent sorting of verbal reports based on their (verbal) content, and, combined with the idea that more basic verbal reports will be less complex:

Figure 6: Content dimension of verbal reports



On this picture, verbal reports can be taxonomized two-dimensionally, in terms of their temporality and their content. By combining the spectra, we can plot both verbal reports and probes on the following accordingly two-dimensional time-content axis to categorize them for interpretation and analysis:

Figure 7: Time-Content Axis for Probes and Verbal Reports

Point *A* indicates a short synchronous high detail probe or report. This might be a probe such as “What strategy did you use to remember the stimulus after it disappeared?” asked shortly after responding to a behavioral cue, or a categorization for a verbal report such as “I held the image in memory by creating a mnemonic out of the first letters of the objects I saw.” Point *B* is a long asynchronous medium detail probe such as “How clear were the details of the stimulus you saw in memory?” asked several minutes after viewing, or a report such as “I could remember the shape but not the color.” A probe or verbal report at Point *C* would be similarly medium detail to Point *B*, except with a medium delay between stimulus and prompting. A verbal report or probe at Point *D* falls under a medium asynchronous basic classification, which might include a probe such as “What color and shape were most of the items you saw?” asked with a moderate delay, or a report such as “I saw mostly blue circles”. One can imagine possibilities for other possible combinations along the time and content dimensions for other probes and reports.

The time-complexity axis constates a systematic heuristic for both constructing probes systematically, as well as parsing and interpreting verbal reports that can be used in consciousness science. Of course, like any heuristic it makes certain theoretical assumptions, namely those just laid out, importantly including that *recall noise*, or the amount of interference by peripheral cognitive processes on retrospective reports increases with time, and that increasing *verbal volume*, or the number of words in a report, is a reflection of complex content. One might disagree with these assumptions, and they might not satisfy the skeptic about introspection (especially one as steadfast as Schwitzgebel), however, they do enable a much needed beginning of a formal approach to probe design and verbal report interpretation. The heuristic itself is also open to augmentation and revision—one may want to add additional axes in the z-plane for instance, or define different axes for different purposes. I'll, however, leave the simple model as it is here, since such moves would require independent justification beyond the scope of this review, and deserve detailed and careful treatment in future work. For now, the 2D time-content axis can be put to empirical work on both sides of subjective reporting.

In study and questionnaire design, researchers can target specific kinds of content by choosing a complexity level or range for probes, and defining a synchronicity based on the kind of information they are after. Basic synchronous probes, for example, will elicit information about what subjects believe their conscious experience to be like in 'raw' form, and ideal for capturing verbal data about phenomenality. On the other hand, detailed asynchronous reports will elicit reports about how subjects are thinking, and ideal for capturing data about reasoning or choice strategies. On the interpretive side of things, researchers can use the time-complexity axis to begin analyzing the reports they receive, for example to determining if the systematically and intentionally designed probes worked as intended by mapping verbal reports onto it. Further analysis can be enriched by examining the features of reports that fall in different parts of the space,

including word count/volume, usage, and additional analyses for content and linguistic features aided by computational methods. With these methods in hand, we can at least hope to unearth something interesting about the ways that subjects verbalize about their subjectivity, and be optimistic about patterns and regularities that such studies might reveal about what it's like to be a conscious subject.

References

- Bayne, Tim, Anil K. Seth, & Marcello Massimini et al. 2024. "Tests for Consciousness in Humans and Beyond." *Trends in Cognitive Sciences* 28 (5): 454–466.
- Block, Ned. 1995. "On a Confusion About a Function of Consciousness." *Behavioral and Brain Sciences* 18 (2): 227–247.
- Block, Ned. 2007. "Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience." *Behavioral and Brain Sciences* 30 (5–6): 481–499.
- Block, Ned. 2011. "Perceptual Consciousness Overflows Cognitive Access." *Trends in Cognitive Sciences* 15 (12): 567–575.
- Boring, Edwin G. 1953. "A History of Introspection." *Psychological Bulletin* 50 (3): 169–189.
- Comte, Auguste. 1830. *Cours de philosophie positive*. Vol. 1. Paris: Rouen Frères.
- Davidson, Donald. 1984. "First Person Authority." *Dialectica* 38 (2–3): 101–111.

- Davidson, Donald. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60 (3): 441–458.
- Dennett, Daniel C. 1982. "How to Study Human Consciousness Empirically or Nothing Comes to Mind." *Synthese* 53, 159–180.
- Di Lollo, Vincent. 1977. "Temporal Characteristics of Iconic Memory." *Nature* 267 (5608): 241–243.
- Ericsson, K. Anders, and Herbert A. Simon. 1980. "Verbal Reports as Data." *Psychological Review* 87 (3): 215–251.
- Ericsson, K. Anders, and Herbert A. Simon. 1993. *Protocol Analysis: Verbal Reports as Data*. Revised edition. Cambridge, MA: MIT Press.
- Irvine, Elizabeth. 2016. "Measures of Consciousness." *Philosophy Compass* 8 (3): 285–297.
- Koch, Christof, Marcello Massimini, Melanie Boly & Giulio Tononi. 2016. "Neural Correlates of Consciousness: Progress and Problems." *Nature Reviews Neuroscience* 17, 307–321.
- Luque, Francisco C. & Denis O'Hara. 2016. "Verbal Reports in the Experimental Analysis of Behavior." *International Journal of Psychology and Psychological Therapy* 2, 157–177.
- Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *The Philosophical Review* 83 (4): 435–450.
- Nisbett, Richard E., & Wilson, Timothy D. 1977. "Telling more than we can know: Verbal reports on mental processes." *Psychological Review* 84, 231–259.
- Neurath, Otto. 1932. "Protocol Sentences." In *Logical Positivism*, edited by A. J. Ayer, 199–208. New York: The Free Press, 1959.

- Perone, Michael. 1988. "Laboratory Lore and Research Practices in the Experimental Analysis of Human Behavior: Use and Abuse of Subjects' Verbal Reports." *The Behavior Analyst* 11, 71–15.
- Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge, MA: MIT Press.
- Schwitzgebel, Eric. 2024. "Introspection." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. Spring 2024 ed. <https://plato.stanford.edu/archives/spr2024/entries/introspection/>.
- Schwitzgebel, Eric. 2011. *Perplexities of Consciousness*. Cambridge, MA: MIT Press.
- Schwitzgebel, Eric. 2008. "The Unreliability of Naive Introspection." *The Philosophical Review* 117 (2): 245–273.
- Schwitzgebel, Eric, and Russell T. Hurlburt. 2007. *Describing Inner Experience? Proponent Meets Skeptic*. Cambridge, MA: MIT Press.
- Schwitzgebel, Eric. 2002. "Why Did We Think We Dreamed in Black and White?" *Studies in History and Philosophy of Science* 33, 649–660.
- Sperling, George. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs: General and Applied* 74 (11): 1–29.
- Tversky, Amos & Kahneman, Daniel. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5, 207–232.
- Williamson, Timothy. 2024. *Overfitting and Heuristics in Philosophy*. Oxford: Oxford University Press.
- Wundt, Willhelm. 1883. "Selbstbeobachtung und innere Wahrnehmung." *Philosophische Studien*, 4: 292–309.