# Data from Introspective Reports:

# Upgrading from Commonsense to Science

Gualtiero Piccinini[1]

Address for correspondence: Department of Philosophy, Washington University, Campus Box 1073, One Brookings Dr., St. Louis, MO 63130-4899. gpiccini@artsci.wustl.edu

**Summary**

Introspective reports are used as sources of information about other minds, in both everyday life and science. Many scientists and philosophers consider this practice unjustified, while others have made the untestable assumption that introspection is a truthful method of private observation. I argue that neither skepticism nor faith concerning introspective reports are warranted. As an alternative, I consider our everyday, commonsensical reliance on each other's introspective reports. When we hear people talk about their minds, we neither refuse to learn from nor blindly accept what they say. Sometimes we accept what we are told, other times we reject it, and still other times we take the report, revise it in light of what we believe, then accept the modified version. Whatever we do, we have (implicit) reasons for it. In developing a sound methodology for the scientific use of introspective reports, we can take our commonsense treatment of introspective reports and make it more explicit and rigorous. We can discover what to infer from introspective reports in a way similar to how we do it every day, but with extra knowledge, methodological care, and precision. Sorting out the use of introspective reports as

[1] I thank Jim Bogen, Matt Boyle, Carl Craver, Peter Machamer, Andreas Roepstorff, Andrea Scarantino, Becka Skloot, and two referees for helpful comments and discussions on this topic. Special thanks to Tony Jack for his specially insightful comments.

sources of data is going to be a painstaking, piecemeal task, but it promises to enhance our science of the mind and brain.

## 1. Introduction: To Introspect or Not to Introspect?

I *feel* tired, this page *looks* white to me, but I'm *thinking* I can't procrastinate any more—these are *introspective reports*. In everyday life, we rely on reports like these to learn about other minds. This paper discusses whether introspective reports are also legitimate sources of scientific evidence. Introspection is often associated with consciousness, and introspective reports may well be particularly useful in the study of consciousness. But the present topic is not consciousness or experience or qualia or any particular kind of mental state, as construed by a scientific theory or by folk psychology. I will not address the status of folk psychology or its relation with scientific psychology. My topic is the use of introspective reports to generate evidence. What that evidence is *about* is a separate question, on which I will remain as neutral as possible.

Science is supposed to be based on *public*, or intersubjective, methods. Public methods are such that (i) different investigators can apply them to answer the same questions, and (ii) when they do so—other things being equal—they obtain the same results (Piccinini, forthcoming). But introspective reports, when they are expressed sincerely,[2] are often construed as the output of a non-public method of observation: I introspect my mind not yours; you introspect your mind not mine. Under this construal, introspective reports violate (i) above: different investigators cannot answer the same questions about the same minds by introspecting.

Because they are private, introspective reports are sometimes said to be in principle

[2] Unless otherwise noted, from now on I will omit this qualification and talk about introspective

*unverifiable*, in the sense that aside from the introspecting subject, no one is in a position to determine whether a report is true or false.

The *introspection agnostic* thinks that since introspection is private and introspective reports are consequently unverifiable, scientists shouldn't take a stance on their truth value. According to the agnostic, scientists should not treat introspective reports as a special source of information about mental states (Lyons, 1986; Dennett, 1991). Instead, scientists should treat introspective reports as observable behaviors, to be explained on a par with all other behaviors (Dennett, 1991, 2001).

The *introspection believer*, by contrast, argues that introspective reports provide otherwise unavailable evidence, which we could not collect by any other means. The believer agrees with the agnostic that introspection is private and introspective reports are unverifiable. Because of this, the believer admits that introspection's truthfulness cannot be established by public methods. Nevertheless, the believer encourages scientists to assume that introspective reports are true at least most of the time, and hence they are legitimate sources of scientific evidence (Goldman, 1997, 2000; Chalmers, 1996). According to the believer, scientists should treat introspection as a genuine method of observation, yielding a special sort of first-person (private) data about mental states. These first-person data are special in that they cannot be obtained by third-person (or public) methods. The result is a new kind of first-person science—radically different from ordinary third-person science (Goldman, 1997, 2000; Chalmers, 1996, 1999; Price and Aydede, forthcoming).

Neither of the above attitudes is satisfactory. The problem with the agnostic position is that introspective reports do seem to be a precious source of information about minds, which

---

reports that are expressed *sincerely*.

many scientists are eager to exploit in their research (Jack and Roepstorff, 2002). Rather than rejecting introspective reports as sources of evidence, it would be preferable to look for a sound way to learn from them. Learning from them is what the believer wants to do, but all she offers to underwrite introspection is the untestable assumption that it's truthful. This is more akin to *faith* than sound scientific methodology. If we are going to use introspective reports as sources of scientific data, we'd better have good reasons.

In this paper, I argue that neither agnosticism nor faith concerning introspective reports is warranted. As an alternative, I consider our everyday, commonsensical reliance on each other's introspective reports. When we hear people talk about their minds, we neither blindly accept what they say nor refuse to learn from it. Instead, we (implicitly) weigh introspective reports against two relevant bodies of evidence: our beliefs about people and their circumstances, and our beliefs about the specific person and circumstance that generated the reports. Sometimes we accept what we are told, other times we reject it or suspend judgment, and still other times we take the report, revise it in light of what we believe, then accept the modified version. For example, if my daughter tells me she's still hungry after a regular meal, I may reasonably infer that she wants attention, not food. Whether we accept, reject, or revise introspective reports, we have (implicit) reasons for it.

And reasons can be made explicit. When we assess the accuracy of our neighbors' introspective reports, we do so by means that are (or can be made) public. If we—who have no scientific theories or methodology—can do this, so can scientists. A sound scientific use of introspective reports can be based on a more explicit and theoretically sophisticated version of the same methodology. The proper scientific use of introspective reports goes hand in hand with the requirement that scientific methods be public—as it should.

This leads me to reject an assumption that's common to both the believer and the agnostic—that learning from introspective reports means relying on a private method of observation. The idea of a first-person science misconstrues the only way we have—either in commonsense or in science—to learn from introspective reports, which is by observing public reports and subjecting them to public scrutiny. When learning from introspective reports is properly construed, introspective reports are no longer sources of unverifiable first-person data. They are public sources of public data, which are no less problematic than any other public data. Hence, using introspective reports does not fall outside of ordinary third-person science.

In conclusion, scientists need not be either blanket agnostics or blanket believers in introspective reports. They should rather be *connoisseurs*—experts who can finely judge, on public grounds, which kinds of introspective reports under which circumstances are informative about which mental states. A connoisseur is better trained and skilled, and therefore forms more accurate and reliable conclusions, than an amateur. But a connoisseur's knowledge is gained by the same public methods that inform the amateur. A connoisseur of introspective reports should do explicitly and rigorously what we attempt to do implicitly when we learn from each other's reports in our everyday life.

## 2. Mind reading

Introspective reports are informative. We ask each other how we feel, what we're thinking about, how things look to us, and we learn from our answers. Our ability to learn from introspective reports is part of our *mind reading*—the ability to discover the content of other minds. This section focuses briefly on some aspects of mind reading, to see how learning from our neighbors' introspective reports might be justified.

First, we can learn about the content of minds from nonlinguistic perceptual input. Almost from birth, we respond to smiling adults whose eyes are pointed in our direction by smiling back. After a few months, we understand—from people's behavior—what others are attending to and what their goals are. For instance, when mom turns her head suddenly to her right, her baby turns in the same direction and tries to locate whatever she must have seen. (But if a box turns to the right, a baby stares at the box and looks puzzled by its funny behavior.) In a few years, we develop a sophisticated understanding of people's perceptions, desires, beliefs, and other mental states. There is evidence that many of these skills are pre-linguistic: for instance, we share many of them with other animals. In short, we have a natural ability to respond to the content of other minds from nonlinguistic perceptual input.[3]

As members of a linguistic community, we have a mentalistic vocabulary in common. We know what it means to *feel*, *perceive*, and *think*; to *believe*, *hope*, and *fear*; to sense *pain*, *itch*, and *pleasure*; to be *focused*, *distracted*, and *bored*; to feel *gloomy*, *excited*, and *enamoured*; etc. Any competent speaker of our language knows how to apply mentalistic predicates, what mentalistic predicates name what conditions, and what inferences can be drawn from statements that contain mentalistic predicates. Any individual who shares those abilities with us, we say, *understands* our mentalistic language.[4] Those who understand our mentalistic language can share their information about minds, theirs and others', through first-person and third-person reports. Embodied within our linguistic competence, we inherit a lot of useful information that we could never acquire by dealing with nonlinguistic inputs: wisdom accumulated by our

[3] This ability is part of what psychologists call Theory of Mind. There is a vast psychological literature, which there is no room to summarize here, devoted to the study of Theory of Mind and its scientific explanation. A good review, focused on Theory of Mind in infancy, is Johnson, 2000.

[4] Marconi (1997) offers a general theory of semantic competence along these lines.

ancestors through millennia of dealing with minds. Furthermore, every mentalistic term, say *jealousy*, embodies within its meaning psychological assumptions, in this case about people's propensity to jealousy under appropriate circumstances. These assumptions are a precious part of the commonsense mentalistic beliefs of anyone who understands mentalistic terms.

With perception and language in place, and thanks to our propensity to form beliefs, we form personal beliefs about other minds, and we share psychological information with them. We observe our neighbors' behavior. We register what our parents, friends, and teachers have to say about minds. We learn from stories about other minds, and occasionally we even read psychology books. We accumulate a large body of psychological beliefs, ranging from beliefs about our friends' personalities to beliefs about how people react to advertisements (i.e., in many cases, by forming desires). To enter the business of mentalistic statements and generalizations, to find evidence for and against them, we have no need for a *scientific* psychological theory. All we need is our perceptual skills, our linguistic competence, and other cognitive capacities involved in forming beliefs.

Introspection contributes to our commonsense beliefs about minds. It helps a lot in knowing ourselves. It probably helps learn our mentalistic language and understand others in analogy with us. Perhaps it even helps develop our perceptual skills. This need not spoil the publicity of mind reading resources. Introspection informs us privately, but its behavioral output, to the extent that it contributes to our commonsense psychological knowledge, is as public as any other piece of psychological information. After we introspect, we can report our findings to the rest of our community, who trades them on the same market as the output of all of our cognitive faculties.

In the trade of introspective reports, our shared resources come especially handy. When

Rebecca tells us something about her mind, we not only understand what she says but also exploit what we already believe to evaluate it. Rebecca can lie, and sometimes we spot clues that she's lying. For instance, many times we asked her what she was thinking, and her answer was, "Nothing." Most of those times, we knew she was lying, and often we knew what she was thinking too. Of course we can be deceived but, in the absence of signs of lie, we take Rebecca to be sincere. This is good because, like the rest of us, she usually *is* sincere. But Rebecca can also engage in wishful thinking. Sometimes she asserts she is calm and relaxed even when she is visibly tense. In those cases, we don't take her introspective reports at face value, nor do we think she is lying to us: we correctly infer from both her report, the way she looks and acts, and other evidence we have, that she is tense but delusional. Jennifer, on the other hand, lately has been in a love frenzy. She goes from boy to boy, falling in love with all. Or better: this is what she says, but we know it isn't true. We've seen her really in love once, and it wasn't like *that*. As to these recent cases, she isn't in love but she desperately *wants* to be, though she'd never admit it.

Delusion is only one of many ways in which introspective reports are defeasible. Generating accurate introspective reports is likely to require the proper interplay of a large number of cognitive processes, ranging from perceptual to inferential to motor. Any of these processes may malfunction in any number of ways, and any malfunction may affect the accuracy of introspective reports. For instance, a subject's report that something looks green to him may be faulty because his introspective faculty is damaged, or because his ability to issue reports is defective, or because he doesn't know the meaning of "green" or of "looks," or because his color perception is impaired to the point that his color experiences of things are incommensurable to

those of others.[5] We can discover whether someone is cognitively impaired on public grounds, independently of any individual introspective report. So, when our neighbors give us a report, we don't accept what they say without question—we often form an opinion about their minds that's more accurate than theirs.

Most of the time, though, we take what our neighbors are saying—more precisely, what we understand of it—to be approximately true. We take their introspective reports to inform us accurately about their mental condition. This is not because we blindly trust what they tell us but because, most times, their reports sound kosher—not fishy—to our well-trained ears. We also relate what they say to their environment and to what we know about people in general and them in particular, which usually suggests that their reports are accurate. For example, our commonsense psychological assumptions include that people *perceive* objects, that perception is necessary both for certain manipulations of objects and for generating visual reports, and that objects are perceivable only when they are in somebody's visual field. So, if Andrea says that she *sees* the salad bowl, we can (implicitly) check the reliability of her report by looking at her and the salad bowl, listening to her accurate description of the bowl's shape, color, position, etc., and noting that she comes away from the shelves carrying the bowl we asked her to get. Under our commonsense psychological assumptions, both her report and her behavior hardly fit with the hypothesis that she isn't seeing the bowl. Although other reports can be subtler to evaluate than visual ones, the same applies, *mutatis mutandis*, to reports of non-visual perceptions, memories, feelings, etc. Andrea's sincere introspective reports can still be faulty in some

[5] I have a light acromatopsia myself. Before I was diagnosed, I used to occasionally argue with people as to the color of things: if I thought something looked green to me, I would argue that it was green. Now I defer to others, and I no longer trust *my own* introspective reports on colors: if I feel tempted to say that something looks green to me but people tell me it's blue, I conclude it doesn't really look green to me—my temptation to think so must be a side effect of my

respect, for example because she is color blind, but for the most part we have plenty of reasons to trust them. And when we don't, we distrust her reports on grounds that are public as well.

All we need for evaluating introspective reports—besides our commonsense beliefs about environments and circumstances—is the resources I briefly described above, all of which can be shared. The combination of the above, including our evaluation of introspective reports, constitutes mind reading. Like many skills, mastering mind reading takes a lifetime, and we can always improve at it. And, albeit we do it mostly automatically, it is painstaking: any new introspective report requires evaluation in its own right, given the evidence that's available in the context of its utterance.

This stands in opposition to the introspection agnostic, who thinks we should be neutral about the content of introspective reports. It also stands in opposition to the introspection believer, who thinks we must rely on the general assumption that introspection is truthful. If the believer were right, we could never establish that introspective reports are accurate or inaccurate; we should just resign ourselves to trust our neighbors' reports. But the believer's suggestion is the opposite of our experience: we start with some natural epistemic resources and, building upon them, we learn to evaluate single introspective reports—one by one. In the long run, at most we establish the accuracy of some introspective reports under some conditions and the inaccuracy of other reports under other conditions. Our mind reading skill is always improvable; our beliefs about whose reports are accurate under what conditions are always revisable. We never establish the truthfulness of most introspective reports under most conditions; we are never in a position to say: we have tested enough introspective reports; we safely agree with the believer that introspection is truthful; therefore, from now on we will trust what sincere people

---

acromatopsia.

say about themselves without question.

We can, however, collect the information that we accumulate and ask questions or make tentative generalizations about the minds of individuals or groups: Andrea gets easily excited; does Lance tend to depression? Moya enjoys thinking about her Catholic background; Italians perceive public authority as an enemy; who's more emotional, women or men? These questions and generalizations are made, and answered or tested, by the same inferential processes that we use in other domains of our commonsense beliefs, without need for any assumption about introspection's truthfulness. These commonsense generalizations, in turn, become part of the beliefs that help us evaluate future introspective reports. In our philosophical moments, we might feel inclined to make a generalization to the effect that, *ceteris paribus*, people's introspective reports are accurate. We might even call the process that leads us from individual introspective reports to this generalization a global validation of "introspection's truthfulness." But this is an empty claim unless backed up by our ability to (fallibly) evaluate the accuracy of individual reports in specific circumstances.

After all I've said, we still can't introspect other minds or otherwise directly observe them, but—contrary to what both agnostics and believers suggests—we do have means to evaluate the accuracy of introspective reports. In our normal interactions, usually without being aware of it, we constantly check the validity of each other's reports by these means. This practice warrants our reliance on introspective reports to gauge other minds. For some limited purposes, such as treating patients, this may be all the warrant needed by scientists who use introspective reports as evidence. But if scientists want to use introspective reports in testing hypotheses about minds, they need a more explicit, rigorous, and sophisticated version of our everyday, implicit method. In the rest of this paper, I consider how we may upgrade our use of

introspective reports from commonsense to science.

## 3. The Epistemic Role of Introspective Reports in Science

Using introspective reports in science doesn't require treating introspective reports as a source of first-person, unverifiable data, gathered by observers practicing a private method of observation. A proper scientific treatment of introspective reports should proceed in the third-person, just like the proper scientific treatment of any other sources of data.

A method is a series of operations performed on some objects and instruments—what scientists call *materials*. Any material has properties, and empirical observation methods are designed to exhibit some fact or phenomenon by exploiting some properties of some materials. Scientific observation and experimentation is largely a matter of know-how: in manipulating their materials, researchers need to know how to do what they are doing, but they need not know how their materials exhibit their properties, much less have a theory of them. The microscopist needn't know how her microscope works, nor does the anesthetist need to know everything about how the patient is made insensitive to pain. In fact, the point of doing research is to uncover something *unknown* through one's materials. As long as materials have the properties needed for a method to work, ignorance of how those properties are generated by the materials, or even ignorance of what properties are at work, does not undermine the validity of the method.

Psychologists and neuroscientists use a somewhat special material—people. Human subjects of experiments are not in the lab to apply scientific methods; they are part of the materials. Researchers instruct their subjects to perform tasks, and then they record their subjects' performance. The recording can be more or less sophisticated. These days, recording rarely consists of simple perception and memory; it involves tape recording or videotaping the

subjects' behavioral responses, perhaps in combination with recording the subjects' neural activity by EEG or neural imaging methods. Subjects are often instructed to give verbal reports, which may include reports about their mental states. Such reports are recorded together with any other relevant piece of the subjects' responses, and then transcribed. Records of the subjects' responses, including transcriptions of their verbal reports, are sometimes called *raw data*; they are only the first stage in the production of data properly so called—those that will appear in scientific journals (Bogen and Woodward, 1988).

Any (relevant) behavioral or neural outputs of a subject can be used to collect information about that subject's mental states. The recordings of the subjects' responses, or the transcription of their verbal reports, are analyzed according to some standard procedure aimed at extracting some relevant information. The extraction of usable data from raw data requires careful methodology, especially when many assumptions are needed. A case in point is the use of neural imaging technologies in neuroscience and psychology, which face many methodological pitfalls (Bogen, 2002). In the case of verbal reports, procedures for data extraction may be as informal as ordinary mind reading, or as formal as the automatic analysis of reports by a computer program (Ericsson and Simon, 1993). The records of these analyses are further processed, and their contents summarized in a perspicuous form, until researchers come up with pieces of information that they call *data* in their scientific reports.

There is no reason why introspective reports should be used to collect information about people's mental states any less than any other relevant human output. Just as other outputs—whether neural or behavioral, verbal or nonverbal—can be exploited as sources of information about people's mental states by judicious methodology, so can introspective reports.

In fact, it's not clear whether there is a principled distinction between introspective

reports and other sources of data. First, notice that being verbal is neither necessary nor sufficient for being introspective. This is because on the one hand, verbal reports need not be about mental states; on the other hand, subjects can be instructed to give any behavioral response (e.g., pressing a button) to report on their mental states. Second, notice that being instructed to report on one's mental states is neither necessary nor sufficient for doing so. This is because on the one hand, people may fail to follow instructions to introspect for a variety of reasons (e.g., misunderstanding, malice, or confabulation); on the other hand, under many conditions people volunteer information on their mental states without being asked for it.[6] "Introspective" is a mentalistic term, which—like belief, desire, and other mentalistic terms—is likely to be definable only in mentalistic terms. A scientific definition of "introspective" will thus have to wait for a scientific theory of mind, which would include a theory of introspection. As of today, such a theory is at best a work in progress. So we don't seem to have a principled distinction between introspective reports and other behaviors. And without a principled distinction, it's unclear what a ban of introspective reports would amount to.

Fortunately, just as other scientists need not have a definitive scientific understanding of the processes on which they rely in collecting their data, we don't need to have an exact understanding of how introspective reports are generated in order to use them as sources of data. Like other scientists relying on other materials, we need not know the details of the introspective process for our use of introspective reports to be legitimate. Reliance on introspective reports to generate data should work in the same way as reliance on any other aspect of a subject's behavior. The introspecting subjects, as part of the materials of the experiment, execute their task, which—in this case—includes giving introspective reports. The subjects' responses,

[6] In fact, according to some researchers, even certain non-linguistic creatures, such as some

reports included, are recorded. Then, records of the reports are analyzed to extract information about the mental states of subjects, just as records of other behavioral and neural responses are analyzed for the same purpose.

The method followed by scientists to generate data from introspective reports includes instructing the subjects, recording their behavior, and analyzing their reports according to appropriate assumptions and standard procedures. In the context of scientific methodology, introspection need not be a first-person "method" for generating first-person data, but only a cognitive process whose outcome is observable behavior, which in turn is a legitimate source of data about the subject's mental states.

A final point should be considered. Human subjects elicit introspective reports in response to instructions from human investigators, and the interaction between investigators and subjects is relevant to the form and content of introspective reports. Because of this, Jack and Roepstorff have argued that scientists eliciting and using introspective reports should adopt a "second-person" perspective in which they treat subjects as "responsible conscious agents capable of understanding and acting out the role intended" (Jack and Roepstorff 2002, Box 2). I agree so far. But according to Jack and Roepstorff, this "second-person" perspective is methodologically distinct from the third-person perspective scientists adopt towards other sources of data. I disagree here. Human subjects are different from other experimental materials, and as such they require specialized knowledge and attitude on the part of the experimenters. Dubbing this specialized knowledge and attitude "second-person" may be a good way to acknowledge that our fellow humans are capable of responding to us in extraordinarily subtle ways. None of this requires that experimenters step outside third-person science. On the

---

primates, can be made to exhibit behaviors that deserve to be construed as introspective reports.

one hand, Jack and Roepstorff's considerations apply not only to introspective reports but also to many other uses of human subjects in empirical research. On the other hand, extracting data from *any* materials—not only from humans—must be based on specialized knowledge of the materials. So taking a "second-person" perspective towards experimental subjects is not in contrast with taking a third-person perspective towards the subjects' reports and the data extracted from them. The remaining question is how to validate the data.

**4. How to Validate Data from Introspective Reports**

When scientists collect data, they face at least three distinct issues of validity.[7] Collecting data from introspective reports results in the same issues, which can be dealt with by the same strategies that scientists rely on in collecting any data.

First, raw data must correlate with a genuine phenomenon and not be an experimental artifact. This issue may be called *process reliability*. If the instruments are imprecise, if some variable is not controlled for, if some unpredicted condition defeats the intended outcome, then the raw data—no matter how carefully handled—will mislead the investigators. For example, if a cell line is contaminated by the wrong kind of cells, it will yield useless data. Scientific chronicles are replete with anecdotes of bogus effects generated by unexpected confounding factors. A lot of scientific ingenuity goes into minimizing these risks—it's a never-ending struggle. So, when a process or tool, like a microscope or telescope, becomes standard in a discipline, usually scientists study its properties systematically to determine under what conditions it can be effectively exploited in research. For example, if a researcher wants to

[7] Validity should not be confused with verifiability. Statements, such as introspective reports, may be verifiable or unverifiable; data, such as data extracted from introspective reports, may be valid or invalid.

anesthetize animals, she needs to know, for any given drug, animal, and experiment, what dosage to administer, whether the experiment will interfere with anesthesia, and whether anesthesia will interfere with the experiment. That information is specific to anesthesia, and useless when anesthesia is not part of an experimental method. Creating a correlation between a measurement outcome and a genuine phenomenon requires a lot of specialized knowledge and skill—there is no general recipe.

Solving the problem of process reliability is mainly about preparing the materials in a careful way, learning about confounding factors (often by trial and error), and taking precautions against them. In the case of introspection, this involves formulating effective and unambiguous instructions for the experimental subjects, recognizing specific confounding factors—such as lack of cooperation, delusion, confabulation, etc.—and learning to avoid those confounding factors. Dealing with the specific confounding factors that affect introspective reports requires specialized knowledge and a systematic study of introspection, and it may be quite difficult to make all the relevant knowledge explicit. But there is no principled difficulty.

A further upgrade towards a more scientific use of introspective reports would be the identification of mechanisms and processes responsible for generating introspective reports and of ways in which these mechanisms can malfunction. Neuroimaging studies, as well as lesion studies and other traditional techniques, are likely to help in this respect.[8] In principle, knowing the neural correlates of introspection can help determine whether a behavior is or isn't a genuine introspective report. If and when we have robust results about what mechanisms and processes are responsible for introspection, we could legitimately conclude that when the wrong mechanisms or processes are involved, the ensuing reports are not introspective.

[8] There is a growing literature that appears to be relevant here. For example, see Frith *et al.*,

A second validity issue pertains to mining the raw data for information. This issue may be called *extraction publicity*. Even if the raw data do correlate with a genuine phenomenon, processing them might add unwarranted or question-begging assumptions that skew the results. A classical example is the case of René Blondot and other physicists of his time, whose apparatus was designed to reveal the presence of a new type of radiation called N-rays. Since N-rays do not exist, the raw data generated by the apparatus correctly correlated with the absence of N-rays. Some physicists who observed raw data from similar apparata registered no effects of N-rays. But Blondot and some colleagues, who were convinced N-rays existed, claimed they were seeing the effects of N-rays—and proceeded to publish dozens of papers about them. Their illusion was publicly exposed by physicist R. W. Wood, who visited Blondot's lab in 1904. While Blondot was putatively measuring the effects of N-rays, Wood surreptitiously removed a crucial part of Blondot's apparatus, without Blondot noticing any diminution in the intensity of the N-rays' effects (Klotz, 1980; Nye, 1980). To reduce the risk of spoiling data during their processing, data must be analyzed and processed according to standard procedures that others can reproduce.

In psychology, to insure extraction publicity in the face of the qualitative nature of many raw data, the data analysis is often done by two different observers studying experimental records independently of each other, without knowing what hypothesis is being tested by the principal investigators. If the two observers obtain the same results from the same raw data, the information they extract is considered trustworthy.

To make the extraction of data from introspective reports as public and reproducible as possible, similar precautions should be taken. In their methodological study of introspective

---

1999 and Vogeley and Fink, 2003.

reports, Ericsson and Simon mention some of the important ones (Ericsson and Simon, 1993, pp. 4-5, 276-8, 286-9): (1) researchers should record the reports rather than take selective notes; (2) recordings should be transcribed verbatim; (3) theoretical assumptions used while analyzing the reports should be as few and as weak as possible; (4) all assumptions should be explicit; (5) categories that are used to analyze reports should be chosen *a priori*, before the analysis takes place, on the basis of a formal analysis of the task; (6) each report should be analyzed using only information contained in that report, independently of the surrounding reports, (7) the method of analysis should be constant from one report to another; (8) individuals analyzing the reports should be blind to the hypotheses being tested; (9) different individuals should analyze the same reports independently, and their outcome should be the same.

All these procedures are analogous to those used for other kinds of behavioral data, like eye movements and fixations. It may not always be possible or even appropriate to follow all of these precautions. In searching for theories in new domains, for example, the data analysis may be less formal than this methodology suggests, but whenever possible—especially when testing theories—the more rigorous method should be followed (cf. Ericsson and Simon, 1993, p. 6).

A third validity issue pertains to what the data correlate *with*. The phenomenon being measured and investigated must be properly understood and described. This issue, which may be called *framework validity*, is part and parcel of the science that is being created during the investigation. Different researchers—working under different theoretical assumptions—may genuinely disagree about what a method is measuring. For instance, when particle physicists announce the results of their data analyses, there may be a period of genuine controversy over what they have measured; for example, as to whether a new particle has been discovered, what particle it is, and what its properties are. These controversies are normally resolved by a

combination of further experimentation and theorizing. The former yields more data about the phenomenon, while the latter aims at either accommodating the new findings within the existing conceptual frameworks or extending existing frameworks so that they can accommodate the new findings. In the end, hopefully an uncontroversial interpretation of the data emerges.

To determine what introspective reports correlate with is not straightforward. Answers that have been offered in the literature include the contents of short-term memory during a task execution, the contents of long-term memory traces of past mental states of the subject, the subject's representation of herself, the subject's qualitative state of consciousness, and what the subject believes the experimenter wants to hear. (These answers are not always mutually exclusive.) Different kinds of introspective reports may correlate with different phenomena under different conditions. For example, reports issued during the execution of a task may correlate with the contents of short-term memory, whereas reports issued after a task is completed may correlate with the contents of long-term memory. Different experimental paradigms may be able to exploit those correlations for some purpose or another. The answer to these questions is for the empirical science to find. It will require both a theoretical language in which to couch the description of what is being measured and a lot of empirical investigation.

A model for how to proceed in the matter of framework validity is the classic study by Ericsson and Simon (1993; see also Ericsson, forthcoming). Ericsson and Simon developed a rigorous methodology for extracting data from two kinds of introspective reports pertaining to the execution of cognitive tasks. Reports of the first kind—*concurrent* introspective reports—are issued during the execution of a task. Reports of the second kind—*retrospective* introspective reports—are issued after a task has been executed.

In determining what introspective reports correlate with, Ericsson and Simon adopted the

following framework, which they saw as supported by decades of cognitive psychology. Human cognition is information processing. Information structures, often coming from perceptual input, are acquired by a central processor and temporarily stored in short term memory. While in short term memory, information structures are available for further processing, including the process of verbalization. From short-term memory, information can be transferred into long-term memory, from which it can be retrieved and verbalized.

Ericsson and Simon used their framework to determine what introspective reports correlate with. According to them, verbalization encodes information structures into concurrent reports—reports containing information being processed in short-term memory at the time the report is made—or retrospective reports—reports made after the time when the information they contain was processed by the sybject (cf. Ericsson and Simon, 1993, p. 11).

Ericsson and Simon's theory may well, in practice, fail to account for some experimental procedures involving introspective reports. For that matter, it may well be false. Those are empirical issues of no concern here. What matters for the current argument is that their theory is supported by evidence collected by public methods, which can be applied by their colleagues. Other researchers may disagree, but not accuse Ericsson and Simon's theory of being based on untestable assumptions, e.g. that introspection is truthful.

Ericsson and Simon do not stop at formulating a rigorous methodology for extracting data from introspective reports. They actually offer evidence that the kinds of reports they study offer reliable information about the mental states of subjects. They do so in two steps. First, they exploit their framework to predict what introspective reports should be like under certain circumstances; second, they review three sources of evidence showing that their predictions are correct. The first source of evidence is independent of introspection itself, while the other two

compare the information contained in different kinds of introspective reports (ib., p. 137).

First source. Many cognitive tasks require the use of perceptually available information. For example, if a subject performing arithmetic calculations with paper and pencil reports that she is summing up two numbers while her gaze is shifting between two digits on paper that represent those numbers, the eye fixations of the subject—combined with the result of her calculation—are evidence of the accuracy of the subject's report. This evidence can be gathered because the same information that is perceived by the subject is also available to the experimenters, and can be recorded by a camera together with the subject's eye fixations. This is far from trivial: in the case of Anton's syndrome—unawareness of one's own blindness—subjects describe what they claim to be seeing while their eyes point in a certain direction, but there's no correspondence between the content of their reports and the stimuli in their visual field. In normal cases, though, there is an experimentally established correlation between eye fixations and the information contained in introspective reports. So, via the theoretical assumption that eye fixations tell what information is processed in short term memory, eye fixations give researchers public means—independent of introspection itself—to establish the accuracy of introspective reports (ib., 173).

Second source. Ericsson and Simon's theory predicts that, if identical cognitive processes are reproduced, introspective reports made during those processes will contain identical information. Subjects can perform a task and give reports on it more than once. If each time their reports are the same in relevant respects, this identity of reports supports Ericsson and Simon's theory, and in turn the conclusion that the reports are accurate. It may be difficult to reproduce exactly the same cognitive process in the same subject, because learning and memory intervene between each two executions of a task, making slight changes to the cognitive process.

But some tasks, like mental arithmetic, do reproduce the same processes each time they are executed, and the results corroborate the reliability of introspective reports (ib., 356-7).

Third source. Introspective reports can be either concurrent or retrospective relative to task execution. Concurrent reports draw on short-term memory, whereas retrospective ones draw on the trace of a process that's stored in long-term memory. By asking subjects to give both concurrent and retrospective reports on the same task, researchers can compare the content of two sets of reports pertaining to the same cognitive process. Ericsson and Simon's theory predicts that retrospective and concurrent reports will be mutually consistent, and that the former will contain less information than the latter. The authors review experimental evidence showing that their predictions are fulfilled (ib., 357ff).

Ericsson and Simon's rigorous theory of introspection, and their consequent validation of concurrent and retrospective reports, is a valuable contribution to the methodology of introspective reports. Other scientists using introspective reports to gather data need not subscribe to Ericsson and Simon's theory, nor do they need to have an alternative theory of introspection. As long as they have no reasons to believe their raw data are experimental artifacts, and as long as they are following public procedures of data extraction, they should feel free to (fallibly) search for the best interpretation of their data.

## 5. Conclusion

Our ordinary mind reading skill allows us to understand, evaluate, and interpret introspective reports so as to extract useful information about our neighbors' minds. This commonsense treatment of introspective reports is wiser than either agnosticism or faith. It shows that we can learn from introspective reports without needing to construe them as the output of a private

method of *observation*. Introspection can be construed as a cognitive process yielding a kind of behavior, which can yield information like any other process and behavior. When this is done, the supposed "unverifiability" of introspective reports becomes a non-issue. If we establish what information can be extracted from which sorts of introspective reports under what circumstances, there is no reason why we shouldn't use them as sources of scientific data.

For some scientific purposes, our ordinary mind reading, perhaps supplemented by the results of relevant scientific research, might suffice to underwrite the extraction of information from introspective reports. For instance, neuropsychologists routinely resort to this practice in interpreting their patients' reports. But relying on our ordinary mind reading is not enough for many scientific purposes, especially the rigorous testing of hypotheses in psychology and neuroscience. A proper scientific methodology of introspective reports should go beyond ordinary mind reading. Scientists can take our commonsense treatment of introspective reports and make it more explicit and rigorous.

Scientists using introspective reports face the same problems that other scientists face, and they can solve them by specialized versions of the strategies that other scientists employ. They need to make sure their raw data are not experimental artifacts, by learning about factors that confound their experimental results and avoiding them. They should steer clear of question-begging assumptions while processing their data, by following public procedures that others can replicate. And they should discover, by a mixture of empirical research and theoretical construction, what mental states correlate with what introspective reports under what conditions.

None of the above constitutes a special "first-person" science. The proper scientific use of introspective reports falls well within the boundaries of ordinary third-person science. This is what makes the use of introspective reports scientifically legitimate, and why scientists should

have no blind faith in introspective reports.

Introspection's truthfulness is not something to be established by armchair philosophizing; it takes hard empirical research to find out what can be learned by what reports under what circumstances. We can discover what to infer from introspective reports in a similar way to how we do it every day, but with the extra knowledge, methodological care, and precision that comes with scientific study. In doing so, we should rely on our best-established scientific theories of mind and brain and our specific knowledge of the circumstances in which the reports are generated (the environment, the task being performed, and the instructions given to the subject). We should also develop empirical tests of the reliability of introspective reports. A serious start in this direction has already been made, at least for certain kinds of reports (Ericsson and Simon 1993; Ericsson, forthcoming). Sorting out the use of introspective reports as sources of data is going to be a painstaking, piecemeal task, but it promises to enhance our science of the mind and brain.

The generic term "introspective reports" blurs important distinctions between different kinds of reports. There are concurrent reports, which are generated while performing a task, and retrospective reports, which are generated from memory traces. There are reports that can be issued automatically and reports that require attention. There are reports about tasks whose execution can be independently observed and reports about qualitative states, like pain. Not all introspective reports are informative in the same way, and not all of them present the same methodological difficulties. A proper methodology of introspective reports will have to take all these differences into account.

## References

Bogen, J. (2001), 'Functional Imaging: Some Epistemic Hot Spots', in Theory and Method in the Neurosciences, P. Machamer, R. Grush and P. McLaughlin, eds. (Pittsburgh: University of Pittsburgh Press), pp. 173-199.

Bogen, J. and J. Woodward (1988), 'Saving the Phenomena', Philosophical Review, XCVII (3), pp. 303-352.

Chalmers, D. (1999), 'First-person Methods in the Science of Consciousness', Consciousness Bulletin, Fall 1999.

Chalmers, D. J. (1996), The Conscious Mind: In Search of a Fundamental Theory (Oxford: Oxford University Press).

Dennett, D. C. (1991), Consciousness Explained, (Boston: Little, Brown & Co.)

Dennett, D. C. (2001), 'Are We Explaining Consciousness Yet?' Cognition, 79, pp. 221-237.

Ericsson, K. A. (forthcoming), 'How to Elicit Verbal Reports that Provide Valid Unobtrusive Externalization of Concurrent Thinking?' Journal of Consciousness Studies.

Ericsson, K. A. and H. A. Simon (1993), Protocol Analysis (Cambridge, MA: MIT Press)

Frith, C., R. Perry, et al. (1999), 'The Neural Correlates of Conscious Experience: An Experimental Framework', Trends in Cognitive Sciences, 3 (3), pp. 105-114.

Goldman, A. I. (1997), 'Science, Publicity, and Consciousness', Philosophy of Science, 64, pp. 525-545.

Goldman, A. I. (2000), 'Can Science Know When You're Conscious?' Journal of Consciousness Studies, 7 (5).

Jack, A. I. and A. Roepstorff (2002), 'Introspection and Cognitive Brain Mapping: From Stimulus-Response to Script-Report', Trends in Cognitive Sciences, 6 (8), pp. 333-339.

Johnson, S. (2000), 'The Recognition of Mentalistic Agents in Infancy', Trends in Cognitive

Sciences, 4, pp. 22-28.

Klotz, I. M. (1980), 'The N-Ray Affair', Scientific American, 242 (5), pp. 168-175.

Marconi, D. (1997), Lexical Competence, (Cambridge, MA: MIT Press).

Nye, M. J. (1980), 'N-rays: An Episode in the History and Psychology of Science', Historical Studies in the Physical and Biological Sciences, 11 (1), pp. 125-156.

Piccinini, G. (forthcoming), 'Epistemic Divergence and the Publicity of Scientific Methods', Studies in the History and Philosophy of Science.

Price, D. and M. Aydede (forthcoming), 'The Experimental Use of Introspection in the Scientific Study of Pain and its Integration with Third-Person Methodologies: The Experiential-Phenomenological Approach', Journal of Consciousness Studies.

Vogeley, K. and G. R. Fink (2003), 'Neural Correlates of the First-person-perspective', Trends in Cognitive Sciences, 7 (1), pp. 38-42.