GIUSEPPE LO DICO

Università Cattolica di Milano

giuseppe.lodico@unicatt.it

INTROSPECTION ILLUSION AND THE METHODOLOGICAL DENIAL OF THE FIRST-PERSON PERSPECTIVE

abstract

This paper will provide an evaluation of the Self/Other Parity Account, according to which introspection is an illusion and the data coming from it are unreliable for justifying theories. The paper will argue that the foundation of this account is based upon an a priori *denial of the first-person perspective, considered as an obstacle to a full naturalization of psychology, that affects both the choice of the methods of inquiry and the interpretation of the empirical data.*

keywords

Introspection, first-person, third-person

INTROSPECTION ILLUSION AND THE METHODOLOGICAL DENIAL OF THE FIRST-PERSON PERSPECTIVE

GIUSEPPE LO DICO *Università Cattolica di Milano*

1. It is Better not to Trust the Subject!

As a psychologist, I wonder why how the epistemological status of my discipline is still a matter of controversy. I think that the very question on the table is nothing but whether it is a naturalized science, a social science, or something hybrid between these two ones. I also think that researchers working in (experimental) psychology tend to consider themselves as natural scientists, no more and no less than physicists or biologists. In this sense, they tend to argue that their object of inquiry – the mind – is something inter-subjectively observable through inter-subjectively validated methods, no more and no less than what is argued by physicists or biologists. This approach can be called *objective* or *third-person* and has allowed psychology to gain results and credit. In this sense, it can be defined as a fruitful approach. But the open question here is whether this approach is completely true, that is, whether it tells the whole story about the mental. Personally, I think it does not and I want to face this problem by discussing a specific issue: the place of introspection in psychology.

I here define introspection as an empirical method of enquiry through which subjects are able to learn and then verbally report about their own currently going on, or very recently past, mental states (Schwitzegebel 2010). It is worth noting that, even though the term introspection rarely occurs in recent textbooks of psychology or research methodology (Hurlburt & Heavey 2001, p. 401), we can find traces of it under various aliases, such as *verbal report* or *self-report*, as stated more than sixty years ago by the historian of psychology E. Boring (1953, p. 163). I am sure that those familiar with psychological literature have no difficulties to say that many experiments or review papers refer to data coming from verbal reports. The point is how researchers consider these data for the justification of their theories, especially in comparison with other kinds of evidence.

My personal opinion is that verbal reports have no credibility among the majority of psychologists. For example, let us consider the following quotation from Philip Johnson-Laird, a prominent authority in the field.

> It is impossible to establish the veridicality of subjective reports. At worst, they may be fraudulent [...]; at best, they may be misleading, because none of us has access to the wellsprings of thought (Johnson-Laird 2006, p. 27).

As I said above, similar statements are quite common in psychological literature. These statements are formulated as showing a mere empirical fact generally accepted by the scientific community: introspection is a sort of illusion, since people have many mistaken notions about their introspective information and its value (Pronin 2009, p. 3). For example, in a documented review ranging from many research areas such as social psychology (*ivi*, pp. 15-26 and pp. 49-51), developmental psychology (*ivi*, pp. 45-46), and neuroscience (*ivi*, pp. 48-49), the psychologist Pronin reports a large body of empirical evidence that aims at demonstrating that the *introspection illusion* can be a source of danger, since it "causes problems. It can foster conflict, discrimination, lapses in ethics, and barriers to self-knowledge and social intimacy" (*ivi*, p. 2). On the basis of the results reported in the review, she individuates four components of the illusion (*ivi*, pp. 4-6):

1. Introspective weighting: When people have to assess themselves, they generally tend to be too confident of their introspections.
2. Self/other asymmetry: When people have to assess the others, they generally do not rely upon introspection.
3. Behavioral disregard: People generally tend to disregard observable behavior when they have to assess themselves, and to take it in full consideration when they have to assess others.
4. Differential evaluation: People generally tend to take into great account their own introspections and to underestimate those of the others.

As we can see, Pronin argues that introspection is a potential source of biases and errors (*ivi*, p. 15) and thus hopeless as a method of scientific inquiry. So, she derives the methodological claim of mistrusting introspection and verbal reports from a large body of empirical evidence. This methodological claim implies the preference for non-introspective methods of any sort in psychological research, such as behavior observation, non-conscious priming, and brain neuro-imaging "[...]in order to pursue the goal of understanding mental experience" (*ivi*, p. 55). Thus, only third-person and inter-subjectively validated methods are allowed in the understanding of the mind: because of the biases and errors that introspection can provoke, researchers cannot trust what experimental subjects tell about their own point of view.

2. The Self/Other Parity Account

It is important to stress that, in order to qualify a method as introspective, it must meet, among the others, the so-called first-person condition. On this condition, introspection aims at generating knowledge, judgments, or

beliefs about one's own mind and no one else's (Schwitzgebel 2010). In other words, this condition implies that, for making introspection, a person must adopt "[a perspective from which one thinks of oneself as an individual facing a world, as a subject distinct from everything else" (Baker 1998, p. 328). This does not mean simply that a person must possess a certain perspective towards the world and her thoughts (*ivi*, pp. 328-329). Rather, it means that she must possess the ability "to conceptualize the distinction between oneself and everything else there is" and "also to conceive of oneself as the bearer of those thought" (*ivi*, p. 330). That is, she must have a strong or robust (and not a weak or rudimentary) first-person perspective (Baker 1998, pp. 331-332; Baker 2013, pp. 147-150).

This implies that it is postulated an asymmetry between the way we can know our own mind and the way we can know others' minds: people cannot directly know others' minds through introspection, but they can indirectly know them only by making inferences from the observation of others' overt behavior. In other words, because of the first-person condition, introspection provides a sort of privileged access to the mind: the owner of mental states can have a better understanding of them than other people and thus, at the methodological level, psychologists can trust her when they ask about what happens in her mind. However, it is clear that this picture clashes with the conclusions reached by researchers emphasizing the presence of an introspection illusion. In fact, if we take a look at the four components of the illusion above mentioned, it appears evident that it is the first-person point of view that leads people to commit evaluation and judgment errors and thus to make the data coming from introspection scientifically unreliable. Roughly speaking, the four components aim at showing that the first-person perspective is so entrenched in subjectivity that it can lead to give unreliable interpretations of both inner mental states and outer behavioral phenomena. For this reason, various researchers appear to endorse a position that is counterintuitive from the standpoint of common sense: the so-called Self/Other Parity Account. This account points out that people can have a reliable and adequate comprehension of their own mind only on the basis of the same processes through which they acquire knowledge of the others' minds rather than through introspection. Thus, according to the simplest version of the Self/Other Parity Account, the first-person condition cannot be met (Schwitzgebel 2010): this means that reliable introspections cannot be met in any way and people can know both their own mental states and those of others only indirectly.

It is clear that these arguments echo the old criticisms moved by the psychologist Watson against the late 19th and early 20th Century

introspectionism in his 1913 behaviorist manifesto. However, I think it would be quite unfair to define the supporters of the Self/Other Parity Account as behaviorists, in spite of the similarities between them. In one of the most important contributions in favor of the Self/Other Parity Account, the developmental psychologist Gopnik strongly rebuts the charge of re-proposing a version of the old-fashioned behaviorism. In fact, differently from behaviorists, she stresses that internal psychological states do exist and that the discovery of their nature is the very aim of psychology: in this sense, the Self/Other Parity Account can be defined as a truly mentalist approach. She goes further by pointing out that there are also "[...] full, rich, first-person psychological experiences of the Joycean and Woolfian kind" (Gopnik 1993, p. 12). However, similarly to behaviorists, she points out that the first-person experiences cannot be considered as the genuine causes of people's thoughts and behaviors: this is so because first, people have internal psychological states, observe the behaviors and the experiences they lead to both in themselves and others; second, they build up theories about the causes of those behaviors and experiences that postulate the adoption of the first-person perspective; third, as a consequence, they experience the first-person perspective. This position is close to that proposed in one of the most-cited and controversial papers in psychology, that is, Nisbett and Wilson 1977, as explicitly claimed by Gopnik (1993, p. 9). The crucial argument of Nisbett and Wilson article can be briefly summarized in the two following points (Wilson 2002, p. 106):

> 1. Most emotions, judgments, thoughts, feelings, and behaviors are caused by an unconscious mind – in Gopnik terms, the *internal psychological states*.
> 2. Because people cannot have any conscious and first-person access to the unconscious mind, the conscious mind confabulates reasons – in Gopnik terms, *build up theories* – to explain emotions, judgments, etc.

A famous example in support of Nisbett and Wilson's view comes from the results of a selection task in which the participants were required to choose between four consumer products that were actually identical and to verbally justify their choice (Nisbett & Wilson 1977, pp. 243-244; see also Newell & Shanks 2014, p. 5). As a result, it was found that the participants tended to select the right-most of the four alternatives without mentioning the position as a justification of their choice. Rather, for this justification, they *built up a theor*y based upon certain attributes of the chosen product.

Thus, according to Nisbett and Wilson, the subjects' missed report about position effects on choice is evidence in favor of the dissociation between (unconscious) third-person psychological states and (conscious) first-person psychological experiences.
I think that the quotation of Johnson-Laird proposed above follows the same line of reasoning of Gopnik and Nisbett and Wilson: most of our mental life is unconscious and, since verbal reports are not the product of any genuine introspection, but rather *post-hoc* theories of what is supposed to happen in the mind, they cannot be considered as reliable tools in psychological research. Thus, the Self/Other Parity Account appears to deny the use of introspective reports in the justification of psychological theories because they are irremediably biased by the subject's first-person point of view. In this sense, such a point of view seems to preclude the possibility to have reliable data at disposal.

3. Is the Empirical Evidence in Favor of the Introspection Illusion Really Grounded?

What amazes me of the literature in favor of the Self/Other Parity Account is the amount of empirical evidence reported for its justification (see Nisbett & Wilson 1977; Gopnik 1993; Wilson 2002; Pronin 2009). In this sense, as I pointed out above, the prevalence of an unconscious mind over the conscious appears to be an empirically well-grounded scientific theory. In fact, the claim that our unconscious mental states play a significant role in the determination of thoughts and behaviors seems to be empirically confirmed and generally accepted by the scientific community. However, all the scientific theories must be continually revised and put into question and psychological ones are not exceptions. Recently, the psychologists Newell and Shanks have proposed a critical review of the role of the unconscious mind on decision-making and have reached conclusions different from those of the supporters of the Self/Other Parity Account. The focus of their work is on the methods used to test whether experimental subjects are conscious or not of the mental processes involved in the determination of behavior during decision-making tasks (Newell & Shanks 2014, pp. 1-2). They point out that, in decision-making tasks, it is made a comparison between a behavioral performance and a conscious assessment based on subjects' verbal reports (*ivi*, p. 3): researchers infer that a mental state occurs unconsciously if the subjects' behavioral performance is clearly guided by this mental state but their verbal reports do not reflect it in any way. According to their proposal, in order to be reliable in assessing the presence or the absence of consciousness, a test must meet four criteria (*ivi*, pp. 3-4, Table 1):

a. Reliability: the assessment test must be unaffected by those factors that do not influence the behavioral performance.
b. Relevance/Information: the assessment test must consider only the amount of information relevant to the behavioral performance or the decision in question.
c.Immediacy: the assessment test must occur concurrently or as soon as possible after the behavioral performance to avoid possible lapses or distortions.
d. Sensitivity: the assessment test should occur under optimal retrieval conditions.

It is important to note that the idea behind their criteria for assessing consciousness dates back to two papers written by Shanks himself and the psychologist St. John in 1994 and in 1997. In these papers, Shanks and St. John provide a criticism to what they call the *Thesis of Implicit Knowledge and Learning*, according to which the most of people's knowledge is the primary cause of their behavior but it cannot be represented into consciousness. Further, also the learning of this knowledge takes place unconsciously both at the time of learning and at the time of retrieval (St. John & Shanks 1997, p. 164). According to their criticism, most studies in favor of the Thesis of Implicit Knowledge and Learning use invalid tests of consciousness, that is, tests clearly violating criteria (b) and (d) listed above (Shanks & St. John 1994, pp. 73-75 and p. 377; St. John & Shanks 1997, p. 167). For this reason, they conclude that the empirical evidence in favor of the Thesis of Implicit Knowledge and Learning is not as grounded as it might appear at a first sight (Shanks & St. John 1994, p. 367 and p. 394; St. John & Shanks 1997, pp. 162-163). In this sense, the paper by Newell and Shanks 2014 can be seen as an application of Shanks and St. John's (1994 and 1997) work to the area of decision-making – where the Thesis of Implicit Knowledge and Learning seems to be prevalent. However, the 2014 paper goes beyond the conclusions reached in the 1994 and 1997 papers. In fact, in the older articles, Shanks and St. John aim only to show that the reviewed studies using tests of consciousness violate the criteria (b) and (d). Instead, in the newer article, Newell and Shanks also show that the studies using tests of consciousness that respect the four criteria above described "either demonstrate directly that behaviour is under conscious control or can be plausibly explained without recourse to unconscious influences" (Newell & Shanks 2014, p. 19). Thus, the points moved by Shanks and colleagues seem to overturn the picture sketched by the supporters of the Self/Other Parity Account. In fact, at the empirical level, they argue that the data coming from verbal reports, if adequately treated, cannot be defined as illusory or confabulatory in any way and can be legitimately used for justifying psychological theories.

Instead, at the theoretical level, the mind seems to be much more conscious and introspectively accessible to a first-person perspective than many researchers can think, and the appeal to the unconscious in psychological theories often appears not to be justified.

4. Concluding Remarks

Now it is time to go back to the starting questions: should introspection and the data coming from it be eliminated from psychology? Should the methods of inquiry of psychology be limited only to the third-person and objective/inter-subjective ones? Is psychology a naturalized science? We have seen that Shanks and colleagues' work casts doubts on the supposed unreliability of the data coming from introspection and on the claim that most of the mind is unconscious (Newell & Shanks 2014, pp. 18-19). I think that their arguments are compelling and that the four criteria they propose should be met in the construction of every test for assessing consciousness. Personally, I think that these criteria are reasonable and not so difficult to be met and can provide a useful guide for evaluating the validity of the results of psychological research.

I believe that the most relevant conclusion of Shanks and colleagues' work can be summarized in this way: researchers should start to take subjects' introspective or verbal reports much more seriously than they actually do. However, to do this, they should also assume that the subject is able to adopt a first-person perspective allowing her to access her own mental states. This assumption seems to clash with the possibility of using only objective/inter-subjective and third-person methods for assessing psychological facts. That is, if we take a look all along the Newell and Shanks 2014 review, we can find that sometimes the authors must focus upon the single data obtained from a single participant for assessing their degree of consciousness and not only upon the results of the overall sample of subjects. This is clear, for example, in one of the studies that they review – and one of the few ones respecting the four criteria above discussed – that is, the Maia and McClelland 2004 paper on the re-examination of the Damasio's Somatic Marker Hypothesis, specifically when they discuss the results of two single participants, respectively number 36 and number 41 (Maia & McClelland 2004, pp. 4-5). As we can see, the adoption of the first-person perspective seems to imply the adoption of methods typical of an idiographic approach. This means that the primary goal of verbal or introspective reports is to provide an accurate description of a particular person's experiences, no matter whether they can be similar to or different from some or most other people's experiences (Hurlburt & Akhter 2006, p. 274). Of course, this does not appear to fit with an idea of psychology as a naturalized science, aiming

at being nomothetic, objective/inter-subjective, and based on the average responses of a large sample of individuals to the introduction of some experimental manipulation in comparison with the response to certain control conditions (Hurlburt & Akhter 2006, p. 297; Barlow & Nock 2009, p. 19). To put it in another way, psychology cannot be limited to the study of the (universal) unconscious and sub-personal mechanisms necessary for a first-person perspective because the knowledge of these mechanisms cannot "supplant or replace knowledge of phenomena that the mechanisms make possible" (Baker 2007, p. 206). Personally, I do not want to argue that psychology should rebut the nomothetic approach in favor of the idiographic one: I believe that these two approaches should be viewed as the *methodological legs* of psychology, in spite of their irreconcilable differences, aims, and historical and philosophical traditions they come from (see von Wright 1971, Chapter 1). Thus, in my opinion, the denial of one of these two methodological tenets would lay psychology on the line to be incomplete. However, it is important to stress that these two approaches appear to be difficult to conciliate, since the explanations used in the idiographic approach cannot leave aside the adoption of the first-person perspective (see Baker 1998, pp. 336-337) and those used in the nomothetic one seem to work exclusively in a third-person perspective.

In conclusion, Shanks and colleagues' papers above considered suggest that there are no empirical reasons to reject the idea of a central role of the conscious mind in psychology. This is because the empirical results in favor of the unconscious mind appears to be theoretically affected by naturalistic presuppositions *a priori* dismissing the first-person perspective. Now, the question at play is no more empirical but philosophical/logical: can a naturalistic framework be a proper account for psychology? More precisely, if the acceptance of the first-person perspective appears to be undeniable for psychology, does the naturalism have the resources for coherently dealing with it? If not, the consequence should be to renounce to an idea of psychology as a fully naturalized science (see Baker 1998, pp. 336-337 and pp. 342-343 and Baker 2007) and to radically revise and reinterpret many psychological concepts and constructs.

REFERENCES

Baker, L.R. (1998), "The First-Person Perspective: A Test for Naturalism", *American Philosophical Quarterly,* 35, pp. 327–348;
Baker, L.R. (2007), "Naturalism and the First-Person Perspective", in G. Gasser (ed.), *How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society*, Ontos-Verlag, Frankfurt, pp. 203-226;
Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, Oxford;
Barlow, D.H. & Nock, M.K. (2009), "Why Can't We be More Idiographic in Our Research?", *Perspectives on Psychological Science,* 4(1), pp. 19-21;
Boring, E.G. (1953), "A History of Introspection", *Psychological Bulletin,* 50(3), pp. 169-189;
Gopnik, A. (1993), "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality", *Behavioural and Brain Sciences,* 16, pp. 1-14;
Hurlburt, R.T. & Heavey, C.L. (2001), "Telling What We Know: Describing Inner Experience", *Trends in Cognitive Sciences*, 5, pp. 400-403;
Hurlburt, R.T. & Akhter, S.A. (2006), "The Descriptive Experience Sampling Method", *Phenomenology and the Cognitive Sciences*, 5, pp. 271-301;
Johnson-Laird, P.N. (2006), *How We Reason*, Oxford University Press, Oxford;
Maia, T.V. & McClelland, J.L. (2004), "A Re-Examination of the Evidence for the Somatic Marker Hypothesis: What Participants Know in the Iowa Gambling Task", *Proceedings of the National Academy of Sciences*, 101, pp. 16075-16080;
Newell, B.R. & Shanks, D.R. (2014), "Unconscious Influences on Decision-Making: A Critical Review", *Behavioral and Brain Sciences*, 37,pp. 1-61;
Nisbett, R.E. & Wilson, T.D. (1977), "Telling More than We Can Know: Verbal Reports on Mental Processes", *Psychological Review*, 84, pp. 231-259;
Pronin, E. (2009), "The Introspection Illusion", in M.P. Zanna (ed.), *Advances in Experimental Social Psychology*, 41, Academic Press, Burlington, United States, pp. 1-66;
Schwitzgebel, E. (2010), "Introspection", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, retrievable at http://plato.stanford.edu/entries/introspection/;
Shanks, D.R. & St. John, M.F. (1994), "Characteristics of Dissociable Human Learning Systems", *Behavioral and Brain Sciences*, 17, pp. 367-447;
St. John, M.F & Shanks, D.R. (1997), "Implicit Learning from an Information Processing Standpoint", in D.C. Berry (ed.), *How Implicit is Implicit Learning?*, Oxford University Press, Oxford, pp. 162-194;
Von Wright, G.H. (1971), *Explanation and Understanding*, Cornell University Press, Ithaca, United States;

Watson, J.B. (1913), "Psychology as the Behaviorist Views it", *Psychological Review*, 20, pp. 158-177;
Wilson, T.D. (2002), *Stranger to Ourselves. Discovering the Adaptive Unconscious*, The Belknap Press of Harvard University Press, Harvard, United States.