Jorge Gonçalves – Working Papers
Introspection and scientific study of consciousness

The problem I'm going to talk about in this lecture is whether we can ascribe scientific validity to introspective data in consciousness studies.

These data have been used indirectly in several areas: neuroscience, cognitive science, subjective well being, pain, etc.

However, when it comes to consciousness these are not merely indirect data: they are essential data, especially when it comes to phenomenal aspects/qualia/what is it like?

Without introspective data we can't really study consciousness. The question is to know whether using this data is in accordance with the demands of the scientific method, or if its subjectivity prevents any objective use of them.

The introspective data I'm referring to are the data we collect when we direct attention to the occurring states of our consciousness. These data can be grammatically described as first-person data, in contrast with third-person data, and collected by external observation of the behaviour and of the nervous system.

These introspective data are normally verbal reports but they could also be behaviours which are subjective responses to a situation programmed by the experimenter: for example, to press a button when a certain stimulus appears.

Examples of verbal reports are:

"I feel a throbbing headache", "I'm falling in love with Mary", "I'm seeing a bright red", "I'm seeing little yellow balls after rubbing my eyes", "I can feel the roughness of this cloth", "I'm getting excited by this book".

We know by our own experience that these data exist and so it makes no sense to say that introspection isn't possible. What makes sense is to question its scientific validity.

Introspection was an important method in the early times of Scientific Psychology. At that time, its proponents insisted that it was a systematic method among others, that implied training of the subjects and of the experimenter.

They considered that this method was in perfect accordance with the demands of scientific rigour. Their definition of empirical, therefore, includes data from introspection.

However, a movement initiated in Russia with Pavlov, which later spread to the US, would become dominant in Psychology, especially in the Anglo-American world.

This movement, known as behaviourism, was an attempt to refound the scientificity of Psychology.

To achieve this goal, they excluded all that they considered not objectively valid. So they refused the introspective method and deemed consciousness irrelevant.

They centred exclusively in the externally observable behaviour and, whenever possible, in neurology.

Not only consciousness, but meaning and value as well were eliminated from the domain of scientific psychology and relegated to the domain of old pre-scientific beliefs.

Only the pure movement produced by a biological system – human or animal – was eligible. Therefore, according to this rationale, behaviour is explained by the same principles whether we are dealing with writing books or the flight from a predator.

In order to differentiate themselves and to downplay the first scientific psychologists at the same time, the Behaviourists coined the term "introspeccionism", which refers to a kind of Psychology where introspection was the only method.

Nevertheless, an attentive reading of the sources (**see Costall, A. (2006). Introspectionism and the mythical origins of modern scientific psychology. Consciousness and Cognition, 15, 634-654)** reveals that the Behaviourists were unfair to the so called introspeccionists, since the introspective method was only one among others like neurology or psychology. Wundt, the creator of the method, was actually more interested in data originating from history and anthropology than data originating from introspection.

The concept of scientificity, as defined by the behaviourists, continued to exert strong influence over psychology and cognitive sciences even when behaviourism began to be criticized. The scientific study of subjectivity and consciousness continued to be resisted. However, it is to be noted that the reluctance to accept introspective data was many times only rhetorical, because in practice these data were used by psychologists of behaviouristic orientation.

Even so, cognitive sciences and neurosciences have been progressively introducing introspective methodologies in order to complement their objective analyses. These sciences normally avoid the use of verbal introspective reports, asking instead for the performance of certain tasks by the subject. However, the experimenters understood that objectivity demands greater control over data. Actually, it could happen that the subjects didn't understand the instructions correctly, or understood them in different ways, depending on the experimenters.

Another difficulty is that of the subjects sometimes using different strategies to formulate the same reply. For example, if the experimenter asks a subject to give an answer to a logical problem in order to observe his brain states, it could be important to know what strategy was employed in order to arrive at the answer.

All of this can be cleared up by asking the subject to describe his/her mental states while he/she is doing the tasks.

Brain imaging can't do without introspective reports because in the phase of establishing correlations it can be necessary for subjects to describe the contents of their consciousness. To know that a certain brain state correlates with a certain mental state, we need to know which mental state it is and, as I have explained, the objective description of tasks is not enough. Therefore, brain imaging is insufficient to describe mental states: introspective data are necessary.

Introspective data are sometimes used indirectly to study mental processes. In this case, introspective data are not fundamental for investigation, but only complementary.

However, there are experiments where introspection has a fundamental role.

One such case is the research of Thomas Ramsoy and Morten Overgaard (xxxxxxxxxxx) on subliminal perception.

They presented a stimulus to a test subject on a computer screen and varied the time of presentation.

When subjects described what they saw, the answers were dichotomised: they either saw the stimulus or they didn't. But when forced to choose (the experimenters asked them to guess) they discovered a graded situation that they put in categories:

No experience – Brief Glimpse – Almost clear experience – Clear experience.

The brief glimpse was the so called sensorial fringe of consciousness.

("Fringes" was a term coined by William James. For example, when we see the Cube of Necker, the part that is not within the focus of consciousness is in its fringe.)

The "clear experience" was consciousness.

The "no experience" was no consciousness.

They didn't use categories at first, but after these categories were discovered they could be used in subsequent experiences.

It was later discovered, through brain imagiology, that correlations existed between these categories and brain states. Different neural states correspond to consciousness, no consciousness and fringes of consciousness.

The point I want to stress here is that introspection was necessary to establish the categories; without introspection it wouldn't have been possible to discover the categories between *no experience* and *clear experience*, and therefore, it wouldn't have been possible to study the fringes of consciousness and its neural correlates.

Neuroscientists like Anthony Jack and Andreas Roepstorff hold that we must assume "a methodological triangulation in which objective behavioural measurement, recordings of brain activity and introspective evidence can be related to each other." (**Jack, A. I. and A. Roepstorff (2002). "Introspection and cognitive brain mapping: from stimulus-response to script-report." Trends In Cognitive Sciences 6(8): 333-339.)**

Why are these authors disposed to accept the objectivity of introspective data?
Why must we accept this objectivity?

One solution consists in denying that there really are subjective and ineffable data. Actually, what would have been would only be third person data since the so called first-person data would become a kind of third-person data. Experimenters would not deal with inner mental states, but only with the observable data produced (verbal reports and records of behaviour measurements) by the subjects in the course of the experimentations.

Piccinnini **First-person Data** compares it with data from other sciences: astronomic data are not distant physical objects, but facts collected by researchers through their instruments. Likewise, in Paleontology, the data aren't creatures long since gone; they are not even fossils, but fossils duly measured, observed and recorded by the investigators.

In the case of introspective data, the real data are verbal reports and not the states that these verbal reports are about. These verbal reports are measurable and manipulable.

For Dennett (**Sweet Dreams: Philosophical Obstacles to a Science of Consciousness)** it's not even necessary to believe that the verbal descriptions are true: they only *seem* to be true. The experimenter must assume a neutral attitude, like the anthropologist who records the myths of a give culture without believing in them.

These positions don't seem to me to be correct.

Firstly, against Dennett, I believe that there are real subjective states. I am not going to discuss here whether these sates are neural or not, but I think that verbal reports refer to things that exist. If I say I'm seeing a colour or I'm feeling a pain I don't understand what it would mean to say that "it only seems to me that I'm seeing a colour" or that "I'm feeling a pain". The utterances refer to something that is happening in me. I think the mental state happens prior to the verbal report, even if the verbal report gives a certain perspective about it. There is no difference between this and the perceptive descriptions of states of the world. We use language to describe these states, but the language implicated isn't the state itself, even if language describes the state in a certain way.

Secondly, I think it is true that the experimenter deals with verbal reports but what he is really interested in are the sates that these verbal reports refer to. In the same way, the paleontologist wants to know the facts that his data are a sign of. Of course he isn't directly in contact with these facts but he can reconstruct them. Likewise, what the scientist of consciousness wants to know is the states that the verbal reports refer to.

And that's why we can conceive a progressive approximation to the reality of things: because the thing is relatively independent of the descriptions done about it.

Now if we admit that there are subjective states and that these sates are not necessarily identical with third-person data, can we equate introspective data with scientific data? Or does the fact that they refer to subjective sates prevent its utilization as scientific data, capable of being crossed with behaviouristic and neurological data?

I think we must not be more demanding with the truth of introspective data than we are with any other scientific data.

It's true that subjective states are by definition experienced only by one subject and they can't be experienced by anybody else. Others can only experience similar states but not the same state.

That is why thought-experiments like those of the inverted spectrum or the zombie were created. It's always possible that the same objective behaviour corresponds to a different subjective experience or to no experience at all.

However, we can assume that creatures of the same species or phylogenetically close species have the same subjective states if they display the same behaviour and neurology. This kind of assumption is common in other sciences. The physicists assume that there is an outside world even if it's possible to put it in question. The aim of those thought-experiments - the inverted spectrum and the zombie - is to demonstrate the existence of *qualia*, but these thought-experiments should not interfere with scientific research.

Moreover, we can also assume that we have the capacity of empathy that allows us to understand the experiences of human beings and even other mammals. When the experimenter hears and records the introspective verbal reports of a subject, she can understand the experiences that these verbal reports express without losing objectivity. One way to see this is that she can change the place with the subject. The experimenter could be the subject and vice-versa, and so we have evidence that they have an intersubjective understanding.

The experimenter is not passive as if she was observing a physical object. She can ask the subject in order to make the subject improve her verbal report.

This brings us to the question of the training of the subjects and of the experimenter: a problem that was extensively dealt with by the first psychologists supporting the validity of introspective data (see for example **Schwitzgebel** **The Unreliability of Naive Introspection" (2008), *Philosophical Review,* 117, 245-273**.).

In a recent article **The Unreliability of Naive Introspection *Philosophical Review.*2008; 117: 245-273,** Schwitzgebel is very sceptical in relation to spontaneous introspection.

He thinks that we can't rely in introspective verbal reports about the conscious occurrence of emotions, sensations of pleasure and pain or vision.

And this doesn't refer exclusively to occasional errors, but to the great majority of verbal reports. The only exceptions would those of the foveal experiences of colour and to presence or absence of traditional experiences of pain.

The problem isn't in the barrier of language but in introspection itself. It's not a question of, for example, how to describe the colours of a sunset, but of the capacity to introspect our conscious states.

The reason he presents as the likely cause for the failure of introspection is that we deploy our judgments and beliefs about the world when we do introspection. The reason is that we try to identify solid and permanent objects and we are not aware of the fleeting current conscious states.

For example, if I know that my hamburger has cheese in it, when asked, I tend to verbalize that I taste cheese but I don't really pay much attention to the real taste I'm having. Evolutionarily, we were built to identify permanent characteristics of the world and the multiplicity of minimal sensations that traverse our consciousness.

I think the author exaggerates the errors of introspection and that the difficulties he puts forward can be compensated through training of the subject and of the experimenter.

The experimenter can ask questions in order to lead the subject to deepen his capacity of paying attention to the details of is experience.

And the subject, in time, can improve this capacity.

Both can improve the language in order to understand each other better.
The language can be formalised by the experimenters in order to reach a greater measure of objectivity.

So, it's possible to overcome this difficulty with training, but the experimenter must beware not to influence the subject too much with his/her preconceived theories. Of course, the subject already has a certain point of view but the goal is that she doesn't incline the subject towards this point of view.

One problem are the experiences that are not common to all human beings.

For example, a mystical state could be reach only by special techniques that a common person has no knowledge of.

In this case, the experimenter can approach the subjective experience of the subject by talking about common experiences in order to form an idea of the mystical subjective state. For example, we can imagine compassion towards all creatures by thinking about compassion we feel for a creature we love, etc.

In this way, we can understand the state of universal compassion that the Buddhists claim to reach with long practice.

It's also possible to study the brain states of a person that is experiencing a mystical state. This was done by Lutz and others **(Lutz A, Brefczynski-Lewis J, Johnstone T, Davidson RJ 2008 Regulation of the Neural Circuitry of Emotion by Compassion Meditation: Effects of Meditative Expertise. PLoS ONE 3(3): e1897 doi:10.1371/journal.pone.0001897)** who neurologically studied the state of Buddhist compassion meditation in a group of long-term meditators and novice ones.

In this study, we can see that the researchers must begin by relying on the introspective analysis done by the subjects, and only then can they infer from the brain state to the meditation state. We can clearly see, in this case, the integration of introspection and neurology in the study of consciousness.

I think I have given sufficient reasons to conclude that introspective data can be scientific data in the study of consciousness if we are careful when using them. We must not rely too much in the first spontaneous introspective reports, and must conduct interviews in order to perfect the introspective capacity of the subject. But I'm convinced that even spontaneous introspective data that are not so bad as some researchers think it is.

After all, we are accustomed to rely on introspective judgements about ourselves in everyday life.

We know that sometimes we are mistaken, especially if we try to describe the processes of our mental states, it means if we try to explain why we have the mental states that we have.

But if we limit ourselves to introspecting current conscious states, I believe we will frequently good descriptions.

When used as scientific data, introspective data aren't infallible. Like perceptive data, they could be mistaken and they are contaminated by language.

Language has a history and it's never pure of connotations, but it allows time the communication between persons at the same, even if it demands an empathetic understanding that reaches beyond words.

The fact that introspection isn't infallible isn't an argument against its use in a science of consciousness. Other scientific methods aren't infallible either.

What scientific methods can do is to amplify the probability of a given conjecture being right, but can never offer absolute certainty.

EEG and imagioly also are a point of view about the brain and not the brain itself. They are the best techniques we have to know neural activity and it's with it that scientists work.

Therefore, even if introspection has its limitations, we can use it as data among other kinds of data.