Andreas Roepstorff and Anthony I. Jack

Trust or Interaction?

Editorial Introduction

One of the best gimmicks on the cognitive science conference circuit is the demonstration of inattentional blindness. Many readers of this journal must have already been exposed to it. For the rest we will briefly describe a striking and popular demonstration.[1] It typically evolves during a conference talk, where the presenter provides the audience with a stimulus in the form of a small video clip of six people, three in white, three in black, who pass two basket balls around. The instruction is to count the number of passes made by the players in white. When the movie is over, the presenter asks for a response, someone provides it, and he then goes on to elicit a report: 'Did anyone notice something strange during the film?' Usually, only those who have seen the film before have, but they are not allowed to answer. The film is then replayed, but this time, the audience is instructed to look out for anything unusual. Muted, surprised laughter goes through the auditorium when someone dressed as a gorilla enters into the scene of kids playing, confronts the spectators, bangs his chest, and slowly walks out again. It seems incredible that it went unnoticed the first time around, and yet those who have been exposed to the experiment before can testify that the gorilla was indeed there in both showings of the film.

What does this say of reports? Should we trust those in the audience who claim that they did not see the gorilla the first time around? Probably — there is no motivation to deceive themselves or others. But then, if we do trust those reports, what does this force us to conclude about their experiences? Their very experiences appear untrustworthy, unfaithful to reality. This catch 22 seems a *reductio ad absurdum* on the very notion of an 'objective report'. The fact is that the man in a gorilla suit did appear, and anyone who tells you they didn't see what they

Correspondence:
Dr A. I. Jack, Washington University Campus Box 8225, 4525 Scott Avenue, St Louis, MO 63110, USA. *Email: ajack@npg.wustl.edu*
Dr A. Roepstorff, PET-Centre, Aarhus University, Nørrebrogade 44, DK-8000 Aarhus C, Denmark. *Email: andreas@pet.au.dk*

[1] If not, you may read the next line of instructions and then look for yourself at http://viscog.beckman.uiuc.edu/grafs/demos/15.html (see also Simons & Chabris, 1999, and Goldman, this issue) .

were looking directly at is obviously mistaken in one way or another. Yet this dismissal of the report would only seem adequate from a distant third- person perspective. If you are one of those persons, who experienced 'what it is like' not to see the gorilla the first time around, you will be acutely aware of how such a dismissive treatment fails to capture both the puzzle following the first call for a report (did I miss something?) and the astonishment when the ape suddenly enters, clearly visible. If we are to make progress on this puzzle, we shall have to be careful not to conflate the experience of reality with the reality of experience.

The Mind-Dependence of Experience

In the closing lines of the first volume of *Trusting the Subject?*, Tony Marcel (2003) argued that contrary to popular thinking in the seventeenth century, the content of consciousness does not have 'thing-like' qualities, it is not something to which one may simply have 'accurate' access. Rather, Marcel claims, it is embedded in an intrinsic subjectivity and no methodology, however objective it may seem, can get out of this. There is always what Marcel calls a 'mind-dependence' of phenomenal experience. Both inside and outside of a laboratory setting, experience is necessarily tied to concrete individuals, it involves a perspective, a mind and a body. The inattentional blindness experiment is one example of this. It apparently involves a dissociation between experience and reality shared by most subjects exposed to it for the first time, and the obvious explanation is that attention helps to shape the content of experience. At a recent demonstration of the phenomenon, one of our colleagues gave a different report. English was not her native language. The American presenter had given the instructions very rapidly, and she had not understood many words of it. Left without an idea of counting the number of ball passes, she badly failed the task. However, she clearly saw the gorilla the first time around, and she did not quite understand what the fuss was all about. In order to understand this lack of experience of inattentional blindness, we must factor the presenter into the equation. If one fails to understand the *script* for the scenario (count the ball passes), or decides not to follow the instructions, one may not be able to give a correct response to the stimulus. But the case of our colleague clearly demonstrates that not only the performance of the task but also the content of experience may co-vary dramatically with adherence to the script. We take this to be a token case of a slightly different phenomenon: the intersubjectivity involved in most cognitive experiments. It suggests that the behaviourist's ideal of a simple stimulus–response relation, visible from a third-person perspective, is usually embedded in a second-person interaction which involves exchange of frames of reference and of attentional focus. This, in turn, may affect the actual experience, putative on-line measurements, and a subsequent report (Jack & Roepstorff, 2002a).

We have previously argued that the interaction between experimenter and experimental subject should be factored into the understanding of a cognitive experiment (*op. cit.*), and that this allows both for assigning a putative role for reports in the data analysis and for understanding that apparently particular human ability to exchange and share interpretive frames of reference (Roepstorff

& Frith, 2004), just as the inattentional blindness case suggests. This phenomenon is an extension of the 'why trust the subject' question asked in the introduction to the first volume (Jack & Roepstorff, 2003). It is a discrete move from the inherently problematic question of 'trust' to the potentially descriptive level of 'interaction'. We shall generalise this discussion towards the end of this introduction. For now, we hope to have evoked in your mind two mental markers, a putative mind-dependence of phenomenal experience and a putative intersubjective element in setting up the mind-dependent context both in reports and experience. Both of these require a distinction between the experience of reality and the reality of the experience.

In the following, we will give a short overview of the content of the two special issues. We hope this will allow the reader to treat them as one ongoing dialogue rather than two distinct volumes.

Setting the Agenda

The opening article for the first special issue is written by **Anders K. Ericsson**, whose work in collaboration with Herb Simon and others since the 1980s has pioneered a careful, critical use of verbal reports in cognitive science. It is not reasonable, Ericsson argues, to ask the scientific community to trust verbal reports in general. Empirical research has firmly demonstrated that certain types of reports are reactive (reporting interferes with the thought processes it was meant to describe), and that in some tasks, there is no agreement between the reports and the performance. This does not, however, imply that reports should be eschewed as scientific evidence. Ericsson suggests that the methods of valid and non-reactive verbalization of tasks may be extended also to instances where the introspective condition is interfering with the task performance. In these situations, reports can be used as inspiration to set up non-reactive psychological experiments that allow for more 'traditional' methods of validation. The problem with reports thus becomes epistemic: how should one establish, in a given situation, the amount of trust the experimenter should extend to the subject? 'Trusting the subject' is therefore for Ericsson not an affirmative option, it is in itself an analytical question that other methods will have to decide.

A similar point is raised by **Daniel Dennett** in his discussion of heterophenomenology. Dennett has for decades been arguing that there is no difference in kind between investigating atoms and planets and investigating human consciousness. It is all about establishing objective, third-person methodologies, with the only difference that *utterances* are the basic entitites in the heterophenomenological examinations, and they serve to point to beliefs that may outline the subjectivity of the subject. The experimenter should not, however, literally 'trust' the subject more than one would trust any other object of a scientific investigation. The truth value of utterances is simply not an issue, at stake is only how they serve to outline a particular subjectivity. This places 'trust' in a double relation where, on the one hand, the experimenter should encourage an atmosphere of trust, but probably be hesitant about *actually trusting* the subjects, and

not forget to, 'quietly erect the usual barriers and foils that keep the subjects from too intimate an appreciation of what the experimenter have in mind'. Dennett's worry is that '[t]he experimenter who gets in a position where the subject can do the manipulation has lost control of the investigation'.

The Relevance of a First-Person Perspective

Whereas Ericsson and Dennett each in their own way place the epistemic burden of truth with third-person evidences, a standard objection of researchers with a phenomenological bent is to insist that there are characteristic properties of the first-person perspective, which a study of consciousness must take into account. In their contribution, **Antoine Lutz** and **Evan Thompson** provide a synthetic review of neurophenomenology, a term coined by the late Francisco Varela (1996).They discuss three standard objections to the use of first person reports: that they may be biased, that the process of generating reports interferes with experience, and that there is an explanatory gap between phenomenological reports and biobehavioural data. By way of a pilot study they purport to demonstrate that these three issues may be the result of ill-posed questions. The experiment demonstrates that reports of the subjective context of perception can be related to patterns in EEG measurements within subjects prior to the actual stimulation. This suggests, the authors claim, that first-person data 'may render intelligible some of the opacity of brain responses' and that the a neurophenomenological approach may, if not close, then potentially bridge the gap between different epistemological and methodological levels 'by working to establish strong reciprocal constraints between phenomenological accounts of experience and cognitive-scientific accounts of mental processes'.

Based on a critical review of the literature on infantile autism, **Dan Zahavi** and **Josef Parnas** outline an alternative position to the claim that our understanding of other people is, in principle, like the understanding of physical phenomena such as stars, clouds and geological phenomena. This tenet, they argue, underlies the 'theory-theory' account of autism where a failure in the development of the capacity to 'mind-read', which requires a theory of mental phenomena, underlies the three cardinal symptoms in infantile autism: impairment in social interaction and empathy, impairment in communication, and lack of creativity. Based on an analysis in the continental phenomenological tradition they suggest that this understanding misconstrues the human condition in that 'the very attempt to grasp the mental states of others is the exception rather than the rule. Under normal circumstances we understand each other well enough through our shared engagement in the common world'. From this perspective, high-functioning autists may be characterised by having much too much theory of mind, but no grasp of that immediate pre-reflexive understanding of social meaning that underlies normal intersubjective interaction. As noted by Zahavi and Parnas, it has become commonplace to describe autistic childen as practising behaviourists (Baron-Cohen, 1989). An extension of their argument would suggest that high

functioning autists could, perhaps, be better described as practising hetero-phenomenologists . . .

Patrick Haggard and **Helen Johnson** claim that most psychological studies with a phenomenological bent have focused on the describing perceptual states. In comparison, control of action has been regarded as automatic and experientially trivial. Based on a review of the history of studying the phenomenology of action and of recent psychophysical experiments they argue that voluntary action may be a privileged area for a connection between phenomenology and neuroscience. This holds true both for the design of psychophysical and brain imaging experiments and for the interpretation of neurobiological findings. A number of experiments suggests that the phenomenology of action is a post hoc phenomenon, and that cases of disputed agency, or error awareness, are particularly prone both to experiential effects and to strong brain activations.

These three phenomenologically inspired contributions all argue that taking the first-person experience seriously may play an important role in cognitive science both in restricting hypotheses, directing research and developing experimental paradigms. **Shaun Gallagher** provides a synthetic account of how phenomenology and neuroscience may benefit one another. One approach is exemplified in neurophenomenology where first-person data and experiental categories restrict the analysis of brain imaging data. A second approach, which is somewhat problematic, Gallagher argues, is to use phenomenology for a post hoc interpretation of scientific findings. Finally 'front-loaded phenomenology' attempts to use phenomenological evidence for generation of scientific hypotheses and experiments.

Evidence from Neuroscience

These discussions demonstrate that one of the major methodological problems in applying introspective and retrospective methods is independent control of reliability. The advent of novel brain imaging techniques may be changing this. **Bernard Baars** reviews seven cases where functional imaging experiments targeted at inherently subjective phenomena show patterns of activity that may, if not prove, then at least lend support to a convergence between introspective reports and objective measurements. It can be argued that some of the most influential brain imaging experiments use carefully controlled subjective states of the volunteering subject as the experimental variable, either because they manipulate the 'script' or use the 'report' as an analytical category (Jack & Roepstorff, 2002a; Roepstorff & Gjedde, 2003). This allows for a process of triangulation, where behavioural data, neurobiological findings and subject reports are related to each other.

One of the most interesting examples of triangulation is to be found in recent animal studies. **David Leopold**, **Alexander Maier** and **Nikos Logothetis** discuss several ways to take subjective reports of perception from non-human primates. By subjecting primates to visual stimulus paradigms whose effects are known in humans, developing methods to let the monkeys to report on their

perceptions, and relate these to measurements of behavioural, physiological neuronal parameters, the authors are able to triangulate on the underlying perceptive state. This may establish when reports of the animals are to be trusted. The interspecies comparison of particular cognitive and perceptual tasks is an area of increasing importance (Nakahara *et al.*, 2002), and it may reveal crucial knowledge both about the generalities of the mammalian and primate brain, and the functional consequences of the particularities of the human brain (Roepstorff & Frith, 2004).

The Nature — and Culture — of Introspective Reports

Reading across the contributions discussed so far suggests that two interconnected issues may, in fact, be at stake here. One pertains to the 'nature' of introspective reports, that is, establishing their potentials and pitfalls. However, it appears as important to understand the particular characteristics of those scientific traditions and practices that either rely on or are sceptical about introspective evidence. We propose to call this 'the culture' of introspective reports.

The systematic collection of introspective evidence from animals requires a number of highly developed techniques. In interactions with people, the exchange of reports of mental states is so commonplace that it is almost unnoticed, and we constantly and almost automatically engage in an implicit evaluation of their truth value and evidence. **Gualtiero Piccinini** suggests that this everyday experience may be a starting point for a scientific use of reports. The approach is pragmatic. Piccinini argues that evidence in science is public and intersubjective, but so are introspective reports. Once they are uttered, they may be treated as raw data for a scientific inquiry on a par with any other kind of information — be that neuronal measurements or tokens of behaviour — given that one, as with all other scientific evidence, takes into account the particularities of the material.

The starting point for using reports both commonsensically and scientifically is that understanding is based on common perception and common language, and it is against this background that one evaluates informally the epistemic properties. However, for a scientific use one must — as with other modes of data collection - establish more specific criteria. There should be *process reliability* (the data must correlate with a genuine phenomenon); *extraction publicity* with regard to how the raw data are transformed into evidence; and a *framework validity* that systematically analyses and explicates how the data are framed and which phenomena they correlate with. These issues are, Piccinini claims, common problems in all scientific investigations, and they have already been solved for certain types of introspective reports, but each type of reports has its particular characteristics that must be worked out through experimental practice.

Introspective reports are not only widely used in everyday life. **Timothy Wilson**, whose work with Nisbett (Nisbett & Wilson, 1977) set the standards for a generation of work on verbal reports, examines the apparent conundrum that while introspective methods are used successfully in almost every subdiscipline

of psychology they have been under constant attack ever since psychology was established as an empirical science more than a century ago. By a sweeping review of a number of different subfields (personality research, emotions research, attitudes research, memory research and developmental research) he demonstrates that in each area, reports yield crucial and useful information, but at the price of structurally similar flaws. The problem is twofold; on one level the pragmatics of the interaction between the psychologist and the subject are such that the latter may not be willing to report what the former is keen on measuring. More fundamentally the subject may not have conscious access to the mental processes that the researcher is investigating. This issue is crucial in determining when one should ask for introspective evidence, and although the area is currently being investigated, we have only a very rough map of the relation between what is consciously accessible, and the adaptive unconscious processing, which appears to underlie much of perception and acting. Wilson hereby effectively connects the question of reportability with more fundamental problems in consciousness research; issues that may only be solved by carefully relating reports to other sorts of evidence.

The work on synaesthesia over the last decades by **Richard Cytowic** and colleagues is one prominent example of a mental phenomenon that was first identified via introspective reports. Although at first the claim that subjects like MW would get tactile perceptions from taste and smell seemed incredible, Cytowic managed through careful and critical use of reports and retesting to establish that 'synesthetic concepts are generic and consistent'. It was crucial to this research process neither to take the reports at face value nor consider them as metaphoric. Instead Cytowic interpreted the reports of the synaesthetes and used this understanding to establish a behavioural correlate. This involves an ongoing dialogue between the clinician and the subject on the experience of the condition that, on one hand, allows them to establish novel objective tests of the condition, and on the other hand allows the synaesthete further self-observation. This interplay between first- and third-person accounts mediated by a second-person relation of shared knowledge and experience is pivotal in Cytowic's claim that when working with introspective accounts, both the introspectors and the examiners need to be trained. The former needs to report the actual experience unbiased without first interpreting it; the latter needs to get rid of theory-laden biases in interpreting and understanding mental phenomena, which they may not have encountered in themselves. The careful use of introspective evidence, which relies on interpretation and contextualisation, rather than blind trust, is essential for clinicians who operate within a paradox: they rely on patients' reports but cannot trust them fully. This pragmatic use of reports, which blurs a strong subjective– objective dichotomy, is, Cytowic argues, in sharp contrast to much current research practice where a denial of subjective reports misconstrues the epistemic object.

In the contribution that closes the first volume, **Anthony Marcel** describes how 'trust', 'self-knowledge' and 'science' have been intertwined concepts for centuries, ever since the beginning of those intellectual traditions that explicitly labels themselves scientific. Indeed, the roots of the current take on the problem

of consciousness have to be found in the seventereenth century, particularly in the notion that phenomenal experience has an objective, thing-like existence. This understanding is wrong, he argues, since consciousness is not singular, unified and transparent; it is better seen as disunified, as opaque and as affected by, perhaps even a product of, paying attention to it. This has important implications for the use of introspective reports, they cannot be considered as a neutral probing. The measurement affects — even constitutes — the phenomena (Jack & Roepstorff, 2002b), but this is not an argument against using introspective evidence. Rather, Marcel claims, one must accept the intrinsic subjectivity and mind-dependence of phenomenal experience and develop research methodologies that cope with this condition. By way of four coping strategies, it is demonstrated how a careful interrelation of different forms of evidence within a conceptually open framework may allow for constructive research practices.

Epistemology

In the paragraphs opening this second volume, **Alvin I. Goldman** succinctly states that 'the question of trusting subjective reports . . . is a question of epistemology'. This allows him to examine the credibility of introspective reports on a par with other sources of evidence such as deduction, induction, memory and perception. In Goldman's careful analysis, introspective reports seem to do pretty well. He cogently argues that introspection for the introspector can be considered a basic evidential source, but that the scientific use of reports is complicated because reports are a species of 'testimony'. Reports therefore exhibit certain inherentp roperties: importantly their credibility varies over topics. As introspective reports are not reliable across all conditions or descriptions of mental life, a crucial problem becomes to establish the range of introspective reliability. In spite of these pitfalls, Goldman summarises his position as 'the provisional attitude of scientists should be wary acceptance of subjects' verbal reports, with the understanding that this default attitude should be reversed if there are experimental or theoretical reasons to do so' (see also Piccinini's contribution to the first issue, 2003).

This careful stance is echoed in the paper by **Jonathan W. Schooler** and **Charles A. Schreiber.** In their analysis, a relation between experience, which in itself is indisputable, and reports introduces a distinction between the contents of an experience, consciousness, and the knowledge of the content of experience, meta-consciousness. This introduces a paradox, since it is possible to have distinct experiences, that at best can be given a very imprecise rendering, just as meta-consciousness may, they claim, misrepresent the underlying experience. They therefore propose having confidence in introspections mainly when they systematically co-vary with experience, behaviours, and/or physiological responses.

This analysis suggests that introspection is not a unitary phenomenon, and this is examined further in **Jesse J. Prinz** paper. Starting out from a somewhat paradoxical Titchener statement, Prinz claims that the many forms of mental processes, which go under the heading of introspection, may not be more related

than the many forms of environmental processes under the heading of 'natural disaster'. This may be one explanation for the many concurrent theories of introspection, which may indeed all be correct but only in respect of a subset of the phenomena. Instead, Prince argues, the different 'species' of introspection may have only a very abstract feature in common: they all involve access to mental states. However, this does not entail an underlying unity at the level of mechanism.

Given the doubts raised, how would Titchener then use introspection as a method? Through a careful reading of Titchener's 1600 page lab manual, **Eric Schwitzgebel** demonstrates that not just any introspection would do. It was of paramount importance that the subjects, or observers as Titchener usually called them, had received 'some significant degree of introspective training'. For a psychologist to use untrained observers was likened to a chemist working with an untrained assistant that could not even read the simplest measure. Therefore, Titchener had his observers train for hundreds or even thousands of hours, 'to observe without bias, preconception and theory'. But an interesting tension is at work here, which may be paradigmatic for much early work on introspection. Against Titchener's implicit argument *if only you train sufficiently you will be able to report your experiences the way they really are*, the obvious counter interpretation seems to be *if you train long enough, your experience changes*. It is as if there is a Janus face to introspective observation; should it aim at being true to distinctions in the outer world or should it aim at the experience itself? Does it make sense to dissociate an experience as such from more or less skilled attentional control? Similar paradoxes reoccur in recent work on neurophenomenology (Lutz & Thompson, 2003), which addresses the question whether processes of the mind should be studied mainly with highly trained observers (i.e. people experienced in certain forms of meditation; see also Cytowic, 2003).

Methodology

A putative difference between attending to perception and attending to the experience of perception is examined experimentally in the paper by **Morten Overgaard** and **Thomas Alrik Sørensen**. In a number of perceptual experiments, they demonstrate significant differences in report accuracy depending on whether subjects attend to the percept or the experience of the percept. This demonstrates, they claim, that introspection is not only a retrospective process. Rather, an introspective stance may interfere with the process of perception itself, and this lends further support to Tony Marcel's claim for the mind dependence of phenomenal experience. In order to identify differences between modalities, Overgaard and Sørensen developed a classification scheme that allowed them to separate correct, uncorrect and near-correct reports. This sets reporting on a par with other epistemic technologies such as various modes of brain imaging, where particular methods have sensitivity to some phenomena but not to others (Jack & Roepstorff, 2003; Roepstorff, 2003c), and it calls for the

development of report modes that allow for fine grained representation of experiences. This is the topic of the following two papers.

In their paper, **Derek J. Snyder**, **Katharine Fast** and **Linda M. Bartoshuk** argue that cautious sampling may allow one to get beyond a simple detection of perceptual threshold. They carefully demonstrate how the construction of suprathreshold scales with ratio properties, and the systematic comparison of experiences different modalities may quantify experiential intensity. This allows for a much more precise mapping of phenomenal experience. Potentially, such an approach paves the way for sophisticated triangulations between brain activities, psychophysics, and experience both within and between groups.

Russell T. Hurlburt and **Christopher L. Heavey** argue that there are clever ways to get around the problems of sampling experience outside a laboratory setting. They review the use of a timed beeper as a consistent method to probe conscious states across a number of experimental and experiential conditions. This shows that using the beeper may offer an access to awareness, which bypasses some of the obvious problems related to the meta-conscious stance of 'soon having to give a report'. The beeper may thereby indicate a possible course between the fundamentalist position that introspection is impossible, and the naïve position that introspection is easy.

Taken together, these three papers clearly demonstrate how it is possible, using different techniques, to get a systematic, epistemological handle on subjects' reports without treating them as mind-independent entities. On the contrary, rather than seeing the subjectivity of reports as a problem to be avoided, they treat them as a resource useful both in understanding the content and the process of perception and cognition. The four papers closing this volume demonstrate that taking 'the subjectivity of the subject' for granted have implications that extend beyond mere pragmatic questions of generating data.

The Problem of the Other

In a fundamental way, the questions of introspection and the understanding of other persons — 'the knowing of me and the knowing of you' — appear if not identical then closely intertwined. This is clearly demonstrated in **Philip Robbins**' comparative analysis of recent research on mentalising — both with regards to the self and to the other. He proposes that neither theory-of-mind nor theory-theory models of mentalising can give a mechanistic account of all properties of introspection. Backed by recent neuroimaging evidence, he suggests that introspection also utilises a distinct monitoring mechanism, which is not necessarily implicated in evaluating the mental states of others. This expansion on the theory-theory model allows the capture of a number of known dissociations between self-other conditions, but it also indicates, as discussed by other contributions to this volume, that different truth values may have to be assigned to different introspective reports: access to beliefs, desires and intentions seems to rely more on complex inferences than do access to judgments and perceptions.

Elisabeth Hill, **David Sally** and **Uta Frith** argue that there are also strong methodological links between introspection and the understanding of mentalising. They provide a detailed parallel analysis of behavioural experiments and of subject reports from people inside and outside of the autistic spectrum. This demonstrates that successfully solving social dilemma tasks, hitherto considered an almost paradigmatic case of mentalising, does not seem to require a theory-of-mind based approach, it can also be due to a rapid, efficient strategy based on logical interactions. Obviously, this finding has implications both for the construction of clinical tests for autism and for the experimental investigations of neuronal correlates of 'theory of mind'. However, it also brings to mind an alternative explanation for autism, put forward by Zahavi and Parnas (2003) in the first issue of *Trusting the Subject?*, that high-functioning autists may be characterised by having much too much theory of mind, but no grasp of that immediate pre-reflexive understanding of social meaning that underlies normal intersubjective interaction.

Free Will and Agency

The final two papers in the volume address the issue of free will and the experience of agency in the light of recent cognitive research. **Eddy Nahmias**, **Stephen Morris**, **Thomas Nadelhoffer**, and **Jason Turner**, demonstrate how many philosophical discussions of 'free will' claim to derive arguments from phenomenological analysis. However, on a closer look, the evidence to be gained from actual investigations in phenomenology and psychology appear very scarce at best. Through an analysis of three classical disputes between compatibilists and libertarians (the 'could have done otherwise' case, the experience of agency and the 'closed call' vs 'confident' decisions), the authors claim that the burden of proof lies with the libertarians. There is, they argue, a need for systematic data collection on phenomenology to get at the 'folk intution' of free will, and a need to supplement this with properly designed experiments. They demonstrate the point by conducting a pilot study of the phenomenology of free will.

Jakob Hohwy and **Chris Frith** use an examination of being in control of one's own actions as a test case for the explanatory potentials of contemporary cognitive neuroscience (see also Haggard & Johnson, 2003). They claim that in this field, which is at the heart of human self-understanding, neurocognitive explanations are doing much better than most sceptics and many researchers alike, would hold. Taken on their own, the various approaches may give partial, selective and contrastive understanding, but seen in context they demonstrate a structural isomorphism between neural mechanisms and conscious states, which moves beyond a simple correlation. It provides, perhaps, a rough sketch of an explanation of conscious states, which integrates introspective reports, behavioural data and brain imaging data (see also Piccinini, 2003). This suggests that doubts about using introspective reports may reflect the current paucity of

explanations of conscious states, rather than any inherent unreliability of introspective reports.

The Primacy of Interacting Persons

We believe these volumes have indicated several ways in which scientists may, in fact, use introspective evidence as a central part of the research practice, not only in the validation of data, but also by implementing it in the very formulation of research questions, by frontloading phenomenology into experiments, as Shaun Gallager (2003) cogently puts it. Ever since the scientific revolution, it has been the societal role of scientists to make sure that colleagues get the epistemology right, to establish why they and their results can be trusted (Shapin & Schaffer, 1985). As is evident from these volumes, there have been understandable reasons to exclude introspective reports from that domain, but we also see the contours of the consequences of this epistemological apartheid.

The articles on the problem of the other and on free will and agency demonstrate why the topic of these special issues has implications far beyond the narrow field of cognitive science. Science is not only about making facts, it is also about construing models and understandings of the world and of ourselves, which at any point in time seems right and plausible. Therefore embedded in — and growing out of — the works discussed here, is a larger story of what it is to be a human, what it is to know, and what there is in the world. Such topics form what anthropologists for decades have known as a cosmology (Roepstorff *et al*., 2003), and there is no doubt to us that scientific laboratories and conferences are one of the most important sites for 'cosmology making' (Barth, 1987) in the contemporary world.

To many scientific researchers the question of whether or not they should collect and make use of their subjects' reports may seem to be a purely methodological matter, a matter of which heuristic works better. Yet as the results of this choice propagate into wider societal circles, its consequences become much broader. The decision not to take notice of the subjects' perspective, to leave out their experience, helps creates a cosmology in which people are seen as truly nothing more than mere mechanisms whose individual perspectives are of little consequence. To restate our final words in the introduction to the first issue: 'Cognitive scientists should not fear that introspective evidence will impugn the scientific creditability of their work. They should fear the Frankenstein science they will create without it'.

What would a non-frankensteinian science look like? We will at this stage dare to propose that it should incorporate at least two extensions to that third-person understanding of the human subject, which has dominated much of the behavioural sciences in the twentieth century (Roepstorff, 2003b). First of all, it needs to be able to encompass the fact that the existence of particular human experiences, although a troublesome epistemic entity, is difficult to talk away. In other words, it needs to be able to incorporate the existence, if not the usefulness, of a first-person perspective. Secondly, and this is perhaps where humans set

themselves most drastically aside in the animal kingdom, it must factor in the human ability to exchange these experiences with each other and, importantly, to influence the very content of experiences by providing context, explanation, understanding and bewilderment. This entails being able to encompass a second-person perspective (Prætorius, in press; Thompson, 2001). Both of these extensions are a version of Marcel's dictum on the 'mind-dependence of phenomenal experience', but we need to consider that the mind having the experience is not necessarily the same as the one that affects the dependence.

Epistemologically — if not ontologically — this suggests that one should take as a starting point, both as scientist and as human, that concrete persons are a basic unit in the world. On the one hand, persons are the anchor points for experiences and brain processes, on the other persons are the meeting point for interpersonal processes, which affect both the experience and the brain activities in concrete individuals. When the presenter, who opened this introduction, told the audience to look out for the ball, he did not communicate with the secondary and tertiary visual areas of their brains; he talked to concrete persons, equipped with particular experiences and competences. They may, in turn, be described as experimental subjects, or even brains, but neglecting the ontological primacy of the person and conflating the different levels into one are two of the greatest dangers of contemporary science.

Conclusion: The Trajectory and the Rocket

It has been a great privilege to edit these two volumes on *Trusting the Subject?*. We have been allowed to include contributions that cover the issue from many perspectives: historical, methodological, conceptual, philosophical, phenomenological to mention just a few. We would like to thank all the authors for their excellent work. Not only did they draft fine papers, they also managed to enter into the larger dialogue of this project. We believe that this effort will allow the reader, through a superposition of the different perspectives, to see the contours of a complex, evolving field of research. More than anyone else this realisation owes a great debt to Anthony Freeman, in his patient and competent role as managing editor of this journal.

The particular course of this project owes a lot to our respective backgrounds. Tony Jack had the peculiar experience, while doing a PhD in experimental psychology on 'perception without awareness' (Jack, 1998), of discovering that the dominant view in psychology was that actually asking people what they experienced, whether or not they were aware of a stimulus, could provide no useful information about their state of awareness. This had caused him to re-examine the theoretical motivation behind psychological methodology and to engage with the philosophical issues surrounding experience — resulting in the claim that introspection is necessary for studying consciousness (Jack & Shallice, 2001). He then went on to explore whether brain imaging could provide an additional basis on which to ground our understanding of the mind.

Andreas Roepstorff had been educated in neurobiology and in social anthropology and he had worked as an 'anthropologist of knowledge' in Greenland, studying how different groups of people had varying knowledge and different ways of knowing environmental processes. This had demonstrated how theoreticians could completely ignore valuable 'knowledge on the ground', and how particular ways of representing seemingly natural processes had wide implications for a more general cosmology (Roepstorff, 2003a). He had then gone to do ethnography in functional imaging labs (Roepstorff, 2004), where he became interested in how the subjectivity of experimental subjects got transformed into the objectivity of scientific papers (Roepstorff, 2002).

When we met in a brain imaging lab we had, in other words, complementary approaches and backgrounds, but each had in them the seeds of this project, and we could see how we brought something different to it. Critically, we both felt that the very success of interdisciplinary work depended on people being able to take different perspectives, and to overcome the barriers of their own practice and methodology. Hence the importance of bringing together researchers from different disciplines to explore the issue, to open a discussion, and hopefully to begin to find points of agreement. However, we had never anticipated that there would be so many others out there interested in contributing, including representatives of most of the major fields that relate to the project: Goldman (this issue) who had made some early remarks about introspection as a source of knowledge in science, Dennett with his rather different view expressed by heterophenomenology (Dennett, 2003), Wilson who gained fame for having introspection bashed but kind of recaptured it (Wilson, 2003), Ericsson who had set the standards for verbal protocol analysis (Ericsson, 2003), Leopold and Logothetis whose excellent and careful report paradigm has given rise to the neural correlate of consciousness debate, but which had never been discussed much itself (Leopold *et al.*, 2003), Hurlburt (this issue) who has introduced methods for refined experience sampling, Bartoshuk (this issue) whose excellent methodological work is bringing experience back into psychophysical investigations, Cytowic who had shown the clinical implications of report use (Cytowic, 2003) and Baars who before most others had insisted on taking consciousness seriously (Baars, 2003).

When one late night in Skövde in August 2001 we got the idea for this examination of the use of introspective evidence and subject reports in cognitive science, we thought of it as a multi-stage rocket. We had recently spent a week of an unusually sunny Danish June in front of a huge old blackboard, which we had propped precariously atop an old picnic table against the back of Andreas's house. Each evening, we covered the blackboard with plastic so that we would not wake up to a blank slate. Each morning, we added further scribblings, the layers of partly erased ideas becoming increasingly dense, still faintly visible on the crowded canvas. In the end, only two things stood out from the fuzzy background. The first was a figure indicating that cognitive activation experiments could not be understood in terms of a simple stimulus–response relationship. We realised that a 'script' was essential for setting up the subject for the experiment,

and that a report was the only way to get access, however problematic, to the subject's subjective understanding of that script. Both of these instances were, we thought, inherently second-person, that is, they relied on an interaction between the experimenter and the subject, which involved some element of mutual trust. The other was the notion 'triangulation', which we took to imply that cognitive brain imaging experiments could be described at three levels; by way of behavioural measurements, by way of brain activations and by way of subject reports. It seemed to us essential to relate these three levels of description, both looking out for instances of accord and discord.

The first major stage in the rocket was to identify between the two of us a common conceptual framework that would encompass our joint intuition that reports were important both in construing and in understanding cognitive experiments (Jack & Roepstorff 2002a). The first issue of *Trusting the Subject?* represented the second stage in the rocket. We will conclude this introduction by a countdown for the third stage: the launch of the second volume of *Trusting the Subject?*. We hope that the rocket is now ready to take off on its own, and that it will provoke discussions and comments in the years to come.

Seven ways in which objective behaviour is different from subjective reports

- In terms of inference to mechanism
- In terms of relevance to internal subjective states
- In terms of 'immediate public availability' of the data
- Behaviour is third-person, reports are second-person — behaviour depends on goals of an organism concerning the world, reports reflect a motivation to engage in a social and communicative interaction.
- Behaviour can and does dissociate from its supposed links to experience — these links need to be justified and established.
- Effective communications about experience require a shared reference frame.
- Behaviour is easily seen to be right/wrong/successful/unsuccessful. Reports can't be so easily verified — their accuracy/success must be inferred based on a combination of independent evidence and assessments of trust.

Six characteristics of reports

Introspective reports are the result of several interpretive steps:

- they depend on subject's categories and conceptual frameworks for understanding the world — many of which will be idiosyncratic to the subject
- they depend on the communicative situation, that is, on the relation of trust and expectations between experimenter and experimental subject
- the concepts for reliably categorizing mental states may or may not need to be learnt
- at times we need to look for thinner rather than thicker descriptions (i.e. get subjects to describe their experiences in as unembroidered a manner as possible)

- objective measures are also interpretive — do not suppose this is a unique or overwhelming problem with reports only
- the critical process is a re-interpretation of the subjects' reports in terms of a model that coheres with what we know about brain function and a re-interpretation of models of brain functions in terms of what we know about experience.

Five characteristics of introspection

Introspection is a cognitive process which

- can look at its operations
- affects other ongoing processes to a greater or lesser extent (i) through redirection of what we are attending to; (ii) through central resources being required to form and generate concepts, for keeping in mind reference points to allow comparisons, to recall information both of source experience and of comparison experiences; (iii) depending on the extent to which the cognitive operation in question itself involves or does not involve introspective judgments
- occurs all the time, whether or not it is elicited directly by the experimenter
- may be brought under greater control by eliciting reports
- involves meta-cognition, and involvement of introspective judgments in cognitive tasks.

Four things that introspection is NOT

With introspection we

- do not 'see' causation, the impression that we can 'see' intentional action is illusory
- do not get to know the actual properties of mental states — these are inferred
- have to realise that rationalising is not introspecting!
- have to realise that what we have internal access to, and how we can accurately categorise internal states is not settled once and for all: is it just perceptual in basis, or does it involve higher order processes (e.g. attention, non-perceptual aspects of emotion, etc).

Three instances of 'intersubjectivity by default'

- The third person stimulus–response situation in a cognitive experiment should be analytically framed by script-report scenario mediated by a second-person interaction.
- This stresses the intersubjective dimension inherent in cognitive experiments and it stresses a necessary relation of mutual trust, but not blind faith, between the experimenter and the subject.
- Rightly framed, intersubjectivity is a resource for experimental design, validation and data collection rather than a hindrance to be avoided.

Two outcomes of reports

Introspective reports can provide

- the primary and most direct source of evidence concerning consciousness and the nature of experience
- insight into the operation of the processes introspected.

One call for research

- Triangulate! Correlations between physiology, behaviour and experience can provide us with some clues, but real progress will occur when we investigate, and begin to explain, disparities. We won't understand the mind until we can integrate these seemingly disparate and incommensurable perspectives.

Acknowledgment

Andreas Roepstorff was supported by a grant from Danish National Research Foundation to Centre for Functionally Integrative Neuroscience.

References

Baron-Cohen, Simon (1989), 'Are autistic children "behaviourists"?', *Journal of Autism and Developmental Disorders,* **19** (4), pp. 579–600.

Barth, Fredrik. (1987), *Cosmologies in the Making: A Generative Approach to Cultural Variation in Inner New Guinea* (Cambridge: Cambridge University Press).

Baars, Bernard (2003), 'How brain reveals mind', *Journal of Consciousness Studies,* **10** (9–10), pp. 85–99.

Cytowic, Richard E. (2003), 'The clinician's paradox: Believing those you must not trust', *Journal of Consciousness Studies,* **10** (9–10), pp. 157–66.

Dennett, Daniel C. (2003), 'Who's on first? Heterophenomenology explained', *Journal of Consciousness Studies,* **10** (9–10), pp. 53–71.

Ericsson, K. Anders (2003), 'Valid and non-reactive verbalization of thought during performance of tasks', *Journal of Consciousness Studies,* **10** (9–10), pp. 1–19.

Gallagher, Shaun (2003), 'Phenomenology and experimental design: Toward a phenomenologically enlightened experimental science', *Journal of Consciousness Studies,* **10** (9–10), pp. 85–99.

Haggard, Patrick & Johnson, Helen (2003), 'Experiences of voluntary action', *Journal of Consciousness Studies,* **10** (9–10), pp. 72–84.

Jack, Anthony I. (1998), *Perceptual Awareness in Visual Masking*. Unpublished doctoral dissertation (London: University College London).

Jack, Anthony I. & Shallice, Tim (2001), 'Introspective physicalism as an approach to the science of consciousness', *Cognition,* **79** (1–2), pp. 161–96.

Jack, Anthony I. & Roepstorff, Andreas (2002a), 'Introspection and cognitive brain mapping: from stimulus-response to script-report', *Trends in Cognitive Sciences,* **6** (8), pp. 333–9.

Jack, Anthony I. & Roepstorff, Andreas (2002b), 'The "measurement problem" for experience: damaging flaw or intriguing puzzle?', *Trends Cogn Sci,* **6** (9), pp. 372.

Jack, Anthony I. & Roepstorff, Andreas (2003), 'Why trust the subject?', *Journal of Consciousness Studies,* **10** (9–10), pp. vi–xx.

Leopold , David, Maier, Alexander & Logothetis, Nikos K. (2003), 'Measuring subjective visual perception in nonhuman primate', *Journal of Consciousness Studies,* **10** (9–10), pp. 115–30.

Lutz, Antoine & Thompson, Evan (2003), 'Neurophenomenology: Integrating subjective experienc and brain dynamics in the neuroscience of consciousness', *Journal of Consciousness Studies,* **10** (9–10), pp. 31–52.

Marcel, Anthony J. (2003), 'Introspective report: trust, self knowledge and science', *Journal of Consciousness Studies,* **10** (9–10), pp. 167–86.
Nakahara, K., Hayashi, T., Konishi, S. & Miyashita, Y. (2002), 'Functional MRI of macaque monkeys performing a cognitive set-shifting task', *Science,* **295** (5559), pp. 1532–6.
Nisbett, R.E. & Wilson, T.D. (1977), 'Telling more than we can know: Verbal reports on mental processes', *Psychological Review,* **84**, pp. 231–59.
Piccinini, Gualtieri (2003), 'Data from introspective reports: upgrading from common sense to science', *Journal of Consciousness Studies,* **10** (9–10), pp. 141–57.
Prætorius, Nini (in press), 'Intersubjectivity of cognition and language', *Phenomenology and Cognitive Science*.
Roepstorff, Andreas (2002), 'Transforming subjects into objectivity: An ethnography of knowledge in a brain imaging laboratory', *FOLK, Journal of the Danish Ethnographic Society,* **44**, pp. 145–70.
Roepstorff, Andreas (2003a), 'Clashing cosmologies: Contrasting knowledges in the Greenlandic fishery', in *Imagining Nature. Practices of Cosmology and Identity*, ed. A. Roepstorff & N. Bubandt & K. Kull (Aarhus: Aarhus University Press).
Roepstorff, Andreas (2003b), 'A double dissociation in twentieth century psychology?', *Journal of Consciousness Studies,* **10** (1), pp. 62–7.
Roepstorff, Andreas (2003c), 'Outlining the sandpit of consciousness studies: a question of foundations or of style?', *Bulletin fra Forum for Antropologisk Psykologi,* **13**, pp. 44–52.
Roepstorff, Andreas (2004), 'Mapping brain mappers: An ethnographic coda', in *Human Brain Function*, ed. R. S. e. a. Frackowiak (Second Edition) (San Diego & London: Elsevier).
Roepstorff, Andreas, Bubandt, Nils, & Kull, Kalevi (ed. 2003), *Imagining Nature: Practices of Cosmology and Identity* (Aarhus University Press).
Roepstorff, Andreas, & Frith, Chris D. (2004), 'What's at the top in the top-down control of action? Script-sharing and "top-top" control of action in cognitive experiments', *Psychological Research,* **68**, pp. 189–98.
Roepstorff, Andreas & Gjedde, Albert (2003), 'Subjectivity as a variable in brain imaging experiments [in Danish]', in *Subjektivitet og Videnskab. Bevidsthedsforskning i det 21. Århundrede*, ed. D. Zahavi & G. Christensen (Roskilde: Roskilde Universitetsforlag).
Shapin, S. & Shaffer, S. (1985), *Leviathan and the Air-pump: Hobbes, Boyle and the Experimental Life* (Princeton, NJ: Princeton University Press).
Simons, Daniel J. & Chabris, Christopher F. (1999), 'Gorillas in our midst: Sustained inattentional blindness for dynamic events', *Perception,* **28**, pp. 1059–74.
Thompson, Evan (ed. 2001), *Between Ourselves: Second Person Issues in the Study of Consciousness* (Exeter: Imprint Academic; also published as a Special Issue of *Journal of Consciousness Studies*).
Varela, Francisco J. (1996), 'Neurophenomenology. A methodological remedy for the hard problem', *Journal of Consciousness Studies,* **3** (4), pp. 330–49.
Wilson, Timothy D. (2003), 'Knowing when to ask: Introspection and the adaptive unconscious', *Journal of Consciousness Studies,* **10** (9–10), pp. 53–71.
Zahavi, Dan & Parnas, Josef (2003), 'Conceptual problems in infantile autism research: Why cognitive science needs phenomenology', *Journal of Consciousness Studies,* **10** (9–10), pp. 53–71.