Morten Overgaard & Thomas Alrik Sørensen

# *Introspection Distinct From First-Order Experiences*

## Introduction

As is the case with other concepts about mental affairs, the concept of introspection has many different interpretations. Some seem to consider introspecting a perceptive act (Lycan, 1996) and others see it as a thinking activity (Rosenthal, 1990). For the present purpose, we will claim it as a common understanding in all such theories that introspection (sometimes referred to as meta-awareness or second order consciousness) presupposes consciousness (e.g. Jack, 1998). States of consciousness, broadly discussed in the philosophical and empirical literature as first order states of consciousness, are states in which a subject is aware of some or other object, thought, or feeling. Introspective states, however, are states in which a subject directs his or her attention towards their own conscious state. According to this understanding — which we claim is a widespread one — introspection can exist only in conscious subjects, and, furthermore, it is by way of introspection that a subject can learn about having this or that experience. To avoid misunderstandings, we wish to underline that the claim is not that experiencing as such is dependent upon such acts of introspection. On the contrary, we believe that a subject can have all kinds of intero- and exteroceptive experiences, directing attention towards the represented object (be this an object in the surroundings, an object of thought, memory, etc.). It is only when the subject directs attention not towards the object as such but towards the very state of being conscious of the object that he or she is introspective.

One could now object that we always know what we experience given that by experience we understand some sort of awareness of something. Only very few accounts of consciousness attempt to make sense of experience as something that does not involve our having access to knowledge about it in, at least, a minimal

Correspondence:
Morten Overgaard, Danish National Research Foundation's Center for Functionally Integrative Neuroscience, Aarhus University Hospital & University of Aarhus, Building 30, Norrebrogade 44, 8000 Aarhus C, Denmark. *E-mail: Overgard@pet.au.dk*
Thomas A. Sørensen, Dept. of Psychology, University of Copenhagen, Njalsgade 80, 2300 Copenhagen S, Denmark.

*Journal of Consciousness Studies*, **11**, No. 7–8, 2004, pp. 77–95

sense of the word (Block, 1995). Some would even say that knowledge about the 'ownership' of a conscious state — that it belongs to 'me' — is an intrinsic property of any conscious state (Zahavi & Parnas, 1998). So how can one make sense of the difference, we claim between the awareness intrinsic to first order states of consciousness and of introspective states? One way to get grips with this delicate matter is by distinguishing between contingent knowledge and knowledge about facts. One might say that in first order experience there is knowledge about the conscious content. The subject is aware of and has access to this content. However, it does not follow from this awareness that there is knowledge of the fact that the conscious content 'is a conscious state had by me'. Such knowledge, we will claim, is not intrinsic in the experience itself, but demands a further mental act.

This difference is often treated with a surprising indifference in modern philosophy of mind, and even more so in empirical science. The latter is especially surprising considering that investigations of consciousness seem crucially dependent upon introspective methodology. When asking a subject about his or her experiences, clearly, it would demand introspection in order to reply. Of course, one can use so-called objective measures, although behavioural measures or studies of correctness do not seem very reliable in the study of consciousness: Experimental psychology has studied unconscious perception for years, and it seems well established that subjects can be correct about stimuli without experiencing them. Furthermore, without the subject's own descriptions, the information one can get about consciousness is very poor. On this basis, one could argue that a scientific approach to consciousness is based on the use of introspection, even though experimental psychology gave up on the introspectionism of Wundt and Titchener long ago.

Here, however, we shall not just explicitly use introspection to study consciousness, but make introspection itself the object of empirical investigation. So far, our distinction between first order and introspective consciousness is a conceptual distinction. Furthermore, we believe that there is an experiential difference between the two kinds of states. However, to conduct empirical investigations on introspection, one must be able to collect some sort of behavioural data (verbal reports included as such) in which such a difference shows up.

## Introspection As Experimental Variable

Jack & Shallice (2001), Schooler (2002) and Lutz *et al.* (2002) have all discussed introspective states (or 'meta-awareness') as separate processes from first-order conscious states. Against such a notion, it has however been claimed that we are never aware of our mental states themselves. Thus, introspection is no distinct state or process, according to this claim. E.g. John Searle (1992) has argued that since a conscious state can only be described in terms of what the state represents, awareness of a conscious state as such is always just awareness of the very object represented by the conscious state itself. According to this view, one cannot make sense of a distinction between being introspectively conscious and

having first-order conscious states. Similarly, Fred Dretske (1995) argues that introspection can be seen as an instance of what he calls 'displaced perception': We come to know that we are in a particular type of mental state by being aware of the objects represented by one's mental states.

We find it is possible to maintain the notion of introspection as distinct from first order experience without going against the critique raised by among others Searle and Dretske. Thus, one might say we never have experiences of our mental states in the way that we have experiences of, say, objects in the world. Yet, we might still be said to have access to the conscious state in such a way that makes us able to recognise and think about the state as being conscious. Although we consider it possible to defend this notion of introspection, it is not always hypothesised that such a process could be distinguished from 'normal' first-order states in an empirical study. We performed an experiment on visual perception to investigate if such a difference could be obtained.

## EXPERIMENT 1

### *Subjects*

Subjects were 16 healthy volunteers with normal or corrected to normal vision (age range 16–35).

### *Stimulus presentation and materials*

Subjects were to focus on a fixation point (a white cross on black background) that was presented for a randomly selected duration, after which a stimulus was presented for a duration span ranging from 16 to 192 ms, with intervals about 16 ms (according to the monitor refresh rate). The stimuli were brief presentations of simple figures; in most cases they were shaped as triangles, circles or squares, and with one of three colours; red, blue or green. However, some variations of the shapes and colours occurred. The variations for shape were half, combined, or upside-down versions of the standard figures as well as a few completely different ones. For variations of colour, we used differences in luminance for each colour as well as yellow and grey as alternative colours. There were three possible locations where the presentation could appear: 1 cm. above the fixation cross, 1 cm. to the left or 1 cm. to the right. In approximately 1 out 10 trials, no stimulus was presented. The stimuli were presented on a black background, on a 15-inch SVGA colour computer screen (cathode ray tube; resolution 800 x 600) with a 16 ms. refresh rate, controlled by a 466 MHz CPU. The programme was made on Presentation version 0.40 in Windows 98. The viewing distance was fixed to 60 cm. The programme checked for uncertainties in timing, and the data was analysed accordingly. The sequence of stimulus presentations was randomised, and no combination of form, colour, position and duration was delivered more than once. In all, each subject underwent 200 trials. A mask consisting of all stimulus features merged together followed the presentation of a stimulus for 500 ms. at all three possible locations.

The subjects responded using three scales: One each for stimulus shape, colour and location. The scale of shapes consisted of a display of 34 different figures, some of which were included in the stimulus material. The scale of colours consisted of eight different luminance levels for the colours red, green, blue, yellow and grey. As for the shapes, only some of the colours were actually included as stimuli. The scale of positions displayed the fixation cross in the middle and the three different locations where the stimulus could occur.

*Experimental procedure*

The subjects were divided in two groups of eight subjects. One group was asked to guess what had been presented to them on the screen in a forced choice task by choosing between a total of 34 possible answers in the categories of stimulus shape and colour. The subjects were instructed to 'Report which shape, colour and location on the screen, the first presented figure had.[1] Report about this by pointing at these scales of colours, shapes and locations after each presentation. Point at the shapes, colours and locations that you believe were shown. Even if you are in doubt, you will have to guess what was on the screen' (quoted and translated from instruction manual). The subjects were told that if they thought no stimulus was presented, they could report 'blank screen' by pointing to a specific figure on the scale. The other group was given a different task. Here, the subjects were asked to introspect and then report about the experience by pointing to the same scale as used in the other task. Now, they were told 'Every time you see a picture, you are not to think about it as a figure out there on the screen. Instead, you are to think of it as an experience you are having' (quoted and translated from instruction manual). The subjects were told that if they had no experience of something on the screen, they could report 'no experience' by pointing to a specific figure on the scale. After each experiment, we carried out an open interview with the subjects, asking them to describe exactly what they did during the experiment. If the subjects convinced us that they either introspected or did not introspect (according to what they were asked to do in the instructions) they were included in the experiment. If the subjects were confused about what was meant with 'introspection' or did not perform the right task consequently, they were excluded. We performed such an interview with each subject in each of the four experiments here reported. Response times were collected manually, using a stopwatch.

*Results*

The responses of the subjects were treated as being either 'correct', 'incorrect' or 'near correct'. 'Near correct'-responses matched stimulus partially correct (e.g. when they pointed at the same colour as the one presented, but in a brighter or darker tone). With regard to the colours, a further category was used in the analysis of data, labelled 'near correct 2'. If subjects pointed at a colour immediately next to the correct one on the scale — that is, one with only little variation

[1] This wording was chosen in order that the subjects should not report about the mask.

compared to the colour presented — it was considered 'near correct'. If they mismatched a colour to a larger extent than plus/minus 1 luminance level at the category it was considered a 'near correct 2'. In table 1, the total distribution of correct, incorrect and near correct responses is shown with regard to the three stimulus features.[2]

We applied a Pearson's chi-square test to look for significant differences between the two subject groups. We looked for 2-tailed significance since the direction of the difference could not be predicted. There is an increased amount of 'near correct'-responses for shape and 'near correct 2' for colour during the introspective task. Accordingly, in the non-introspective condition, the subjects had a tendency to be more in the 'correct' or 'non-correct' categories. In the cumulated data, there is a 2-tailed significant difference for shape ($p=0.001$) and for colour ($p<0.001$) but not for position ($p=0.45$). This latter result is of less interest in that it does not make sense to have a 'near correct' category for position.

We also looked for significant differences between the individual subjects within each condition as well as between individual subjects in a comparison of the two conditions using Pearson's chi-square test. The range of P-values when comparing individual subjects in the non-introspective condition was 0.60–0.74 and 0.41–69 in the introspective condition. The average P-value when comparing individual subjects between conditions was less than 0.001. This indicates that the difference between the two groups is much more likely to be due to the different instructions rather than there being different subjects in the two conditions. Thus, our experiment shows a measurable difference between reports when subjects were given an instruction suggesting introspection than when the instruction suggests the subject to guess.

The response times were rather long, given that the subject was to identify the correct colour or shape from a large number of possibilities, and the manual recording of response times were of course not ideal. However, there seemed not to be significant differences between the introspective and non-introspective conditions (mean RT = 3.4 sec. for shape and 3.7 sec. for colour in the non-introspective condition, and 3.9 sec. for shape and 3.6 sec. for colour in the introspective condition).

### *Discussion*

Do our results in fact vindicate an objectively measurable difference between introspective and non-introspective reports? One line of criticism, our experiment seems open to, is that our two conditions do not only differ in terms of introspection and non-introspection. It has been argued based on our first reporting of results using this experimental design (Overgaard *et al*., 2001) that there is a further difference, in that our subjects perform a dual task in the introspective condition. We are unsure of this criticism ourselves, in that we do not find it obvious that subjects attend to the stimulus *and* to their experience of the

[2] In the 'no experience' category for the non-introspective trials, the 'blank screen'-responses are displayed.

stimulus. It could just as well be argued that subjects simply attend to the content of their experience and give their response on this basis. In fact, this is the way the subjects should respond to follow instructions. This issue is addressed experimentally in experiment 2, below.

Another problem to the results is the fact that we do not know how often 'guessing', in the non-introspective condition, reflected conscious processes and how often it reflected non-conscious (subliminal) perception. We found no way to control for this factor in the experiment, since the subjects would have to take on an introspective attitude to their perceptual task if they were to answer this question. One could of course choose to assume that the number of 'no experience'-responses in the introspective condition were more or less equivalent to the number of not experienced trials in the non-introspective condition, given that stimulus conditions were identical. However, such an assumption would be dubious given the fact that the experiment indicates that introspection affects correctness, thus probably also the ability to detect whether a stimulus was presented or not. To make up for this problem, we included the possibility to report 'blank screen' in the non-introspective condition. Subjects experiencing nothing at all on the screen thus had this possibility.

The measured response times indicate no interesting difference between the two conditions in terms of memory decay, which both conditions must be hypothesised to be influenced by. It would however be interesting to see a replication of the study with a set-up that yielded faster responses, thus less memory decay.

## EXPERIMENT 2

We repeated the set-up of experiment 1, but instead of varying introspection, we gave 16 new subjects the non-introspective task coupled with another task, thus inducing a dividing of attention. We believe that this experiment would test the criticism stated above.

### *Subjects*

Subjects in experiment 2 were 16 healthy volunteers with normal or corrected to normal vision (age range 22–31).

### *Experimental procedure*

The subjects were divided in two groups of eight subjects. Both groups were given the instructions of the non-introspective task of experiment 1. One group was asked to visualise a three-syllable word that was read to them app. 2 sec prior to each stimulus and to keep that visual image while performing the task. A new word was read for each stimulus. The other group was asked to repeat the same set of words in their head while performing the task.

### *Results*

The results of the two dual tasks are presented below together with the two original conditions.

| | Shape | | Colour | | Position | |
|---|---|---|---|---|---|---|
| | **Non-intro** | **Intro** | **Non-intro** | **Intro** | **Non-intro** | **Intro** |
| **Correct** | 61 | 42 | 53 | 40 | 82 | 78 |
| **Incorrect** | 21 | 17 | 19 | 15 | 12 | 16 |
| **Near correct** | 9 | 27 | 19 | 12 | 0 | 0 |
| **Near correct 2** | - | - | 3 | 26 | - | - |
| **Blank / No experience** | 9 | 14 | 6 | 7 | 6 | 7 |

| | Shape | | Colour | | Position | |
|---|---|---|---|---|---|---|
| | **Word repetition** | **Visualisation** | **Word repetition** | **Visualisation** | **Word repetition** | **Visualisation** |
| **Correct** | 47 | 34 | 40 | 33 | 73 | 71 |
| **Incorrect** | 30 | 46 | 26 | 36 | 20 | 23 |
| **Near correct** | 11 | 12 | 22 | 18 | 0 | 0 |
| **Near correct 2** | - | - | 5 | 7 | - | - |
| **Blank / No experience** | 12 | 8 | 7 | 6 | 7 | 6 |

*Table 1*. Distribution of responses in the four experimental conditions of experiment 1 and 2.

We analysed all four conditions for significant differences for both shape, colour and position responses using Pearson's chi-square.

| | | **Non-introspection** | **Introspection** | **Visualisation** |
|---|---|---|---|---|
| **Introspection** | **Shape** | ** | | |
| | **Colour** | ** | | |
| | **Position** | NS | | |
| **Visualisation** | **Shape** | NS | ** | |
| | **Colour** | ** | ** | |
| | **Position** | * | NS | |
| **Word repetition** | **Shape** | ** | ** | NS |
| | **Colour** | NS | ** | NS |
| | **Position** | NS | NS | NS |

*Table 2*. Test for significant differences between the four conditions.
** = significant, P<0.01, * = significant, P<0.05, NS = not significant

*Discussion*

The two conditions, inducing dual tasks, do not at any of the parameters above differ significantly from each other. Seemingly, the visualisation task is somewhat more difficult than the word repetition task, which makes good sense

because the other task is visual. Compared to the simple non-introspective task, the two dual tasks differ in that the subjects generate more mistakes, and they differ significantly at some parameters (non-introspection vs. visualisation for shape and colour). The introspective condition, however, differs significantly from all other conditions at all parameters (except for position). Furthermore, and more importantly, the difference between the non-introspective condition and the two dual tasks seems to be of a different kind than is the difference between the non-introspective and the introspective task. In the dual task, subjects generate more mistakes. We do not find the tendency to respond in a way that seems informed by the stimulus yet that is 'slightly off target' as it is the case with 'near correct' reports.

Does this result then validate experiment 1 in showing that the difference between the introspective and the non-introspective conditions does in fact reflect introspection alone? One could attempt to argue that our two ual task conditions are not a 'real test' of the kind of divided attention that hypothetically is going on in the introspective condition. One could argue that they are different because the subjects perform two completely new tasks. However, this argument could be taken as supportive of our claim in that we exactly hypothesise that the subjects do not perform two different tasks when being introspective.

It could, nevertheless, still be argued that with a more parsimonious account, one would find a gradient of interference between the tasks, with introspection eliciting the least interference, and visualisation eliciting the most. More experimental work could help to resolve this issue, but it would require a testing of a large amount of tasks.

Another line of interpretation would be the possibility mentioned above that subjects in the introspective condition were slower in generating reports, so that the decreased ability to give correct reports reflects a decay of memory. A further argument against this possibility, apart from the RT data, would be that it could not explain why subjects also were less incorrect in the introspective condition.

### *Dissociations between response modalities*

Anthony Marcel (1993) describes a result comparable with the results of experiment 1. Marcel carried out a series of experiments with the blindsight patient G.Y. who was instructed to (a) blink with his right eye, (b) press a button with his right hand, or (c) say 'yes' if he *felt* that a light was shown in his blind field. The experiment showed a dissociation where the patient gave the most correct answers when using eyeblinks as response, and the fewest correct answers when using verbal responses.

Following the initial experiment, Marcel showed that this kind of dissociation between response modality is not restricted to blindsight patients, but can be found in normal subjects. Here, Marcel used two different instructions, one in which the subjects had to guess whether a light was presented, and one in which they had to report if they 'had an impression' of the light. All stimuli were presented at subjective threshold. According to our previous discussion, the first instruction induces a non-introspective task: They did not report about their

experiences. The second instruction, however, does induce an introspective task according to our definition.[3] The subjects are to judge whether they had 'an impression' of the stimulus as opposed to making guesses about it. Such a question can only be answered by consulting what is consciously perceived. So, though the subjects can be 'just as conscious' in the first task, they need to attend to their experiences only in the second one.

Using back-projection as above, a luminance increment in a designated location was presented that was so small that each subject only had a vivid experience of it 25% of the time, Marcel found that 9 out of 10 normal subjects were less correct using verbal responses compared to the other modalities when giving introspective reports. When guessing, the subjects showed the same kind of dissociation, but to a much smaller extent.

| | | **Introspection / Stimulus** | | **Non-introspection / Stimulus** | |
|---|---|---|---|---|---|
| | | **Present** | **Absent** | **Present** | **Absent** |
| **Blink** | **False** | 32 | 33 | 15 | 15 |
| | **Correct** | 68 | 67 | 85 | 85 |
| **Button press** | **False** | 32 | 35 | 15 | 23 |
| | **Correct** | 68 | 65 | 85 | 77 |
| **Verbal** | **False** | 55 | 38 | 25 | 20 |
| | **Correct** | 45 | 62 | 75 | 80 |

*Table 3.* Results for normal subjects in Marcel's detection task using three response modalities.

Marcel's results, like those of experiment 1, show an empirical difference between subjects' reports when giving non-introspective and introspective reports. To go beyond a descriptive look at Marcel's data, we decided to reanalyse his results using signal detection theory (Green & Swets, 1966) to disentangle variations in sensitivity (measured by parameter $d'$) from variations in response bias (measured by the natural logarithm ). The calculations show a *d'* value of 0.91 for eye blinks, 0.85 for button press and 0.18 for verbal responses in the introspective condition. Contrary to what Marcel states in his paper, the difference between eye blinks and button press does not seem very convincing. In the non-introspective condition, $d'$ is 2.07 for eye blinks, 1.76 for button press and 1.52 for verbal responses. The difference between the two kinds of report condition is quite apparent: The subjects have more correct responses in the non-introspective condition. Furthermore, the reduction in performance for verbal responses compared to the other is more convincing in the introspective condition. Response bias is generally low: 0.01 for eye blinks, –0.04 for button press and 0.18 for verbal responses in the introspective condition. In the

[3] Marcel does not refer to introspection in his 1993 paper, in that the primary goal was to explore differences between conscious and unconscious perception. However, he finds an interpretation in terms of introspection in accordance with the design and results of the experiment (Marcel, personal correspondence).

non-introspective condition: 0 for eye blinks, –0.26 for button press and 0.13 for verbal responses.

This result, as such, suggests that introspection as a factor is an important part of the cause of the phenomenon that Marcel observed. Furthermore, the experiments generate two hypotheses about the nature of introspection. One hypothesis Marcel mentions, is that the nature of an intended response influences the experience. Another hypothesis, mentioned by Marcel, is that different ways of reporting have differential access to an experience. Both of these hypotheses are consistent with Marcel's findings.

Marcel's findings have implications for our understanding of the relation between introspection and (other) cognitive processes. Not only does the result confirm that introspection seems to give rise to recordable differences compared to non-introspection as mentioned above. It could even be used as a point of departure to narrow down which cognitive processes that relate to (i.e. are affected by) introspection. The two hypotheses that Marcel mentions, could be re-stated as a question whether introspection should be seen as part of/affecting the act of perception or the act of reporting. That is, it could be argued that if it is the case that the way the subject orients him — or herself towards the stimulus changes the perceptual experience itself, introspection should not be seen as some sort of retrospective reflection. Rather, it should be seen as an activity that occurs already while perceiving the object, or maybe even before that. If that is indeed the case, the equation often made between introspection and retrospective reports is wrong. Of Marcel's two hypothesises and the two operationalised re-statements, here mentioned, are not fully identical. However, they do overlap when it comes to their most central aspects: Whether introspection is part of accessing the conscious content in order to report about it or whether it is part of an expectation to report about something before it has been actually perceived. A clarification of which of the two hypotheses that is most likely, would give indications about which cognitive processes introspection is related to. To explore this, we decided on an experimental approach.

## EXPERIMENT 3

In the original experiments by Marcel, the method of back-projection was used to present the stimulus with a brightness that made each subject see it clearly 25% of the time, see it 'almost clearly' 50% of the time, and not see it 25% of the time. When this level of brightness was achieved, it was used as background condition for the experiments. For replication, we used a computer with a Pentium II 350 Hz processor and a screen with a refreshment rate of 85 Hz. With the programming tool E-prime, we created a design as close to Marcel's original design as possible based on his 1993 paper and personal correspondence. A loop function could assess the ideal level of stimulus hue, so that if the number of correct responses exceeded 25% for a block of 20 trials, the stimulus hue would automatically decrease for the next block, and vice versa. This process was to continue until the subject was at the same level of correctness as in the experiments

by Anthony Marcel. The stimulus brightness in the first block was found by letting the subject look at a dot on the screen before the experiment started. The experimenter adjusted the brightness manually until the subject reported that the stimulus had disappeared. We used a blue dot as stimulus throughout the experiment, in that this allowed for a finer control when finding the baseline for each subject: Each level of brightness is closer to the previous for the blue colour than for red and green.

However, with this methodology, we were unable to replicate Marcel's initial level of correctness. On the contrary, our subjects were either close to 100% correct about the stimulus, or, at the next level of brightness, at the chance level of 50%.

We succeeded in replicating Marcel's initial condition of 25% correctness, adding to the design that a fixation cross would come up on the screen, and the stimulus then would appear 3 cm. to the left or right from the cross. Corresponding with Marcel, we later realised that the stimulus was presented peripheral to the fixation point in the 1993 article. This probably explains our initial failure, replicating his data, and our success using a fixation cross and peripheral stimuli.

*Subjects*

Subjects were 10 healthy volunteers (5 males and 5 females, age between 20 and 44 years of age). All had normal or corrected to normal vision.

*Stimulus presentation and materials*

The visual stimulus was a blue dot appearing 3 cm. from the middle of the screen (either left or right) with a size of 2° diameter visual angle.

*Experimental procedure*

We used two conditions for the experiment. In one condition, the subjects were informed that they should give a report if 'they had a clear experience of something being shown on the screen'. As in the experiment by Marcel, the subjects were to report by eyeblinks, by a moving of the hand, or verbally. For eyeblinks, the subjects were to blink their right eye for 'yes, I experienced something' and their left eye for 'no'. For hand movements, they were to lift their right hand for 'yes' and left hand for 'no'. For verbal responses they were to say 'yes' or 'no'. The subjects were told which kind of response to use by a word on the computer screen ('eye', 'hand' or 'mouth') in a randomised order. The subjects were trained in this rather complicated task (paying attention to the stimulus as well as the kind of response to use) for 2–300 trials, during which their individual criterion for stimulus hue was established. In 50% of the trials, the word of instruction came on before the stimulus, in the other 50% it came on after the stimulus. After each presentation of stimulus and cue, the computer gave an auditory signal, after which the subject was to give his or her response as quickly as possible. If the dissociation described by Marcel could be replicated and if it were only present in the pre-instruction condition, we reasoned that it could serve as argument that the effect was related to an anticipatory activity. In such a case, it

would support the hypothesis that the intention to report in a specific way, when being introspective, changes the experiences of the subject. Or, according to our restatement, introspection can be part of a perceptive act. By contrast, if the effect is the same in both conditions, it could serve as argument against this hypothesis.

The two conditions were as follows:

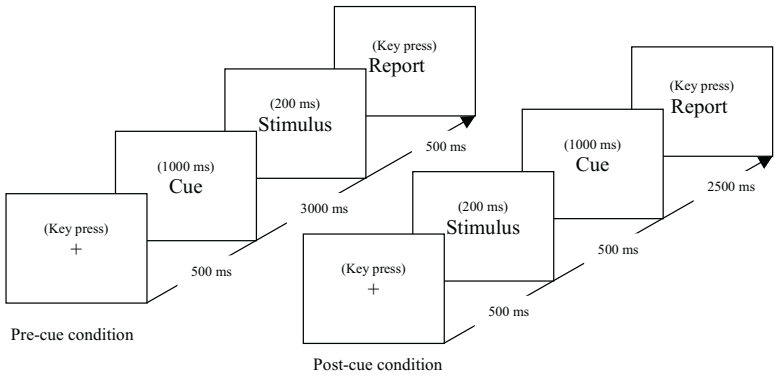



*Figure 1.* Stimulus design for experiment 3.

The logic of the experimental procedure has certain presuppositions. First of all, it is assumed that introspection coupled with knowledge about response type is the cause of the dissociation between response modalities. Furthermore, it is assumed that introspection has a certain duration in time so that it in principle would be possible to investigate when the dissociation is present and when it is not. Results indicating that such dissociation is present in the pre-cue condition but not in the post-cue condition could be interpreted to show that subjects introspect when perceiving the stimulus, and not when issuing the report[4]. Or, at least, it could be said that introspection has its maximum effect on behaviour during the perception of the stimulus. On the other hand, results showing that the dissociation is present in both conditions could be taken as an indication of introspection being part of an activity of reflection or reporting. This is so, because the knowledge of how to report would be present during introspection in both conditions. Based on Marcel's two hypothesises, a situation in which dissociation is present in the post-cue condition but not in the pre-cue condition would not be expected.

### *Results*

At a first look at the distribution of correct and false responses of the subjects, there did not seem to be an effect of the time of the stimulus cue, as shown below.

[4] This latter possibility does not say that subjects are introspecting while actually reporting, but rather immediately prior to the report. However, several theorists would consider consciousness and introspection as related to reporting – one claim, issued by David Chalmers (1997) for one, is that introspection requires that a mental state is 'available for report'.

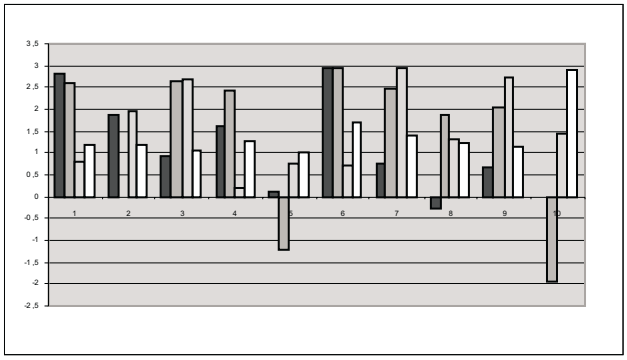
| | | Post cue condition | | Pre cue condition | |
|---|---|---|---|---|---|
| | | **Present** | **Absent** | **Present** | **Absent** |
| **Blink** | **False** | 35 | 90 | 50 | 10 |
| | **Correct** | 65 | 10 | 50 | 90 |
| **Button press** | **False** | 35 | 7 | 47 | 12 |
| | **Correct** | 65 | 93 | 53 | 88 |
| **Verbal** | **False** | 36 | 10 | 48 | 11 |
| | **Correct** | 64 | 90 | 52 | 89 |

*Table 4*. Distribution of responses for experiment 3.
The table shows correct and false responses when stimulus was present and absent.

The graphs below, however, indicate different tendencies for pre- and post-cue instruction conditions when looking at the variance of responses for each subject individually.

Figure 2 shows that the subjects generally perform equally well with all response modalities in the post-cue trials but not in the pre-cue trials. Furthermore, we see that subjects generally perform better in the post-cue trial, but not consistently so. The differences between the two conditions do not show up in

**A**



**B**

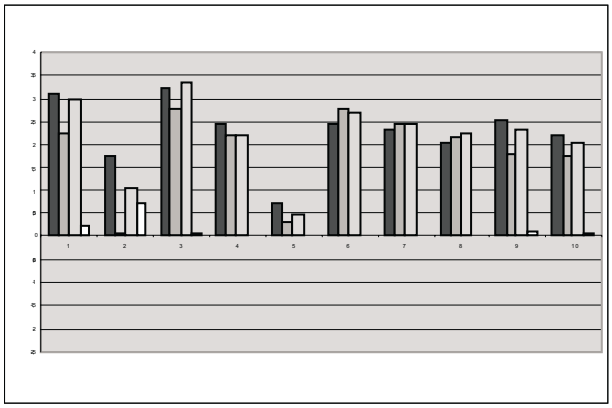


*Figure 2.*

Distribution of responses in experiment 3 calculated with *d′* using pre-cue instruction (A) and post-cue instruction (B).

For each subject, four bars are displayed showing responses using

(1) hand movement,

(2) verbal reports, and

(3) eye blinks.

The last bar for each subject shows the variation in *d′* between the other three categories.

table 5 because the subjects, contrary to Marcel's subjects, differed individually with regards to which response modality they performed better and worse with. We used signal detection theory to analyse the data for each subject. Looking at the variance of $d'$, subjects are significantly different in the pre- and post-cue conditions.

*Discussion*

Our results support one of Marcel's two hypothesises: That the intention to respond in a specific way is a crucial factor for the dissociation between response modalities. With our design we did not find the general tendency that Marcel describes. In fact, more subjects performed better verbally than with eyeblinks than the other way around. Also, the subjects were not consistently less correct in the pre-cueing conditions: It rather looks as if some kind of noise is introduced.

Marcel explained the variance between response modalities in the introspective vs. non-introspective condition with reference to the way in which the subjects were oriented towards the stimulus. Accordingly, our finding of a significantly increased variance in the pre-cue condition can be seen as due to ongoing introspection. Thus, we suggest that introspection gives rise to significant variance between response modalities only in the pre-cueing condition because the subjects are introspective during the perception of the stimulus. That is, when the subjects are not yet informed how to respond when seeing the stimulus in the post-cueing condition, their introspective attitude does not cause any significant differences in their responses between modalities. This, we believe, indicates that introspection is more related to the pre-stimulus preparedness or the actual perception rather than a retrospective reflective activity or the report.

One could argue that subjects differ in the two conditions because there is a different duration from their seeing the stimulus to their report. This could lead to a different level of memory decay in the two conditions. Furthermore, the duration in experiment 3 could have modified the variability in the pre-cue condition and removed an effect that would otherwise be present. We decided to check for this possibility by replicating experiment 3 with the variation that the two conditions were identical in the duration from stimulus to report (instead of being almost identical in the duration from cue to report).

## EXPERIMENT 4

*Subjects*

Subjects were 10 healthy volunteers (5 males and 5 females, age between 22 and 32 years of age). All had normal or corrected to normal vision.

*Visual presentation and materials*

The visual stimulus was a blue dot appearing 3 cm. from the middle of the screen (either left or right) with a size of 2° diameter visual angle.

*Experimental procedure*

The experimental procedure was identical to that of experiment 3, except the duration from stimulus onset to report was now identical in the two conditions.

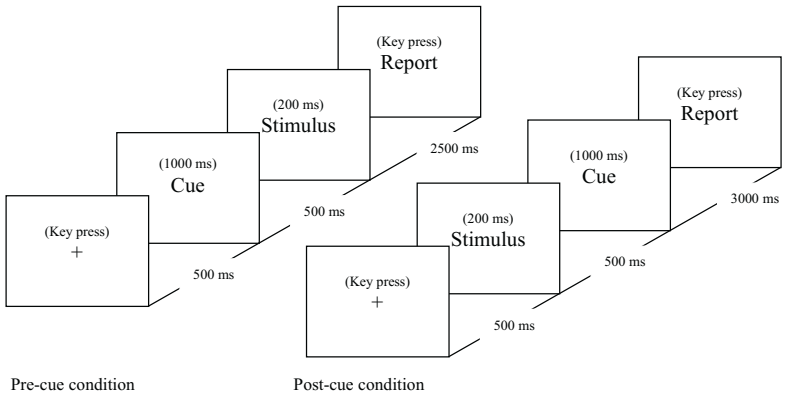



*Figure 3*. Stimulus design for experiment 4.

*Results*

At a descriptive level, data from the two conditions look more similar than in experiment 3, as shown below.

| | | Post cue / Stimulus | | Pre cue / Stimulus | |
|---|---|---|---|---|---|
| | | **Present** | **Absent** | **Present** | **Absent** |
| **Blink** | **False** | 40 | 1 | 40 | 3 |
| | **Correct** | 60 | 99 | 60 | 97 |
| **Button press** | **False** | 45 | 1 | 52 | 11 |
| | **Correct** | 55 | 99 | 48 | 89 |
| **Verbal** | **False** | 44 | 1 | 45 | 13 |
| | **Correct** | 56 | 99 | 55 | 86 |

*Table 5*. Distribution of responses for experiment 4.
The table shows correct and false responses when stimulus was present and absent.

However, as before, when looking at the variance for each response type for the two conditions, the difference is very remarkable.

When comparing figures 2 and 4, it is revealed that the experiment 3 and 4 generate identical tendencies. There is a high variance between response modalities in the pre-cue condition, and almost none in the post-cue condition.

*Discussion*

Experiment 4 leads to the same conclusions as experiment 3. This fact underlines that an introspective preparation to act in a specific way influences a subject's report. Furthermore, the experiment finds no significant effect of varying the duration from stimulus or cue to report.

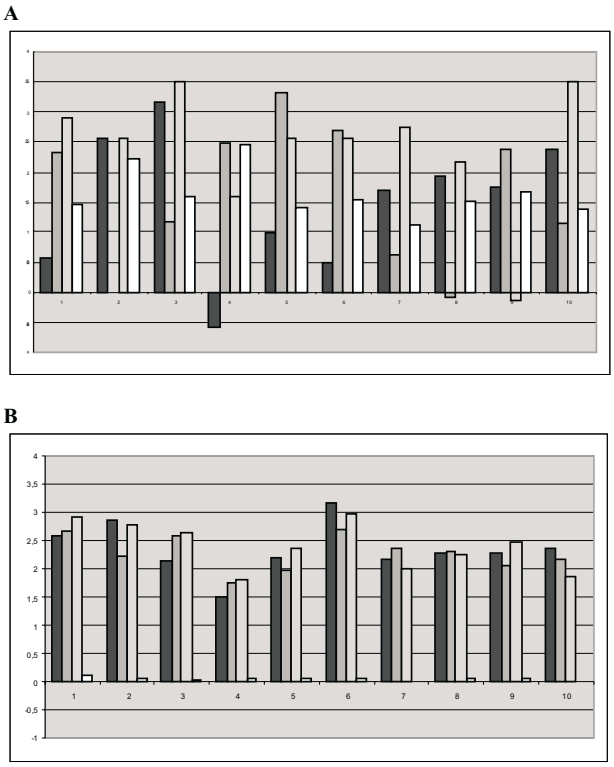



*Figure 4*. Distribution of responses in experiment 4 calculated with *d'* using pre cue instruction (A) and post cue instruction (B). For each subject, four bars are displayed showing responses using (1) hand movement, (2) verbal reports, and (3) eye blinks. The last bar for each subject shows the variation in d' between the other three categories.

One serious critique that could be raised against our third and fourth experiment is that the validity of our interpretation depends on whether they can be compared to Marcel's results. That is, we assume that introspection coupled with knowledge of how to respond is the cause of the observed effect. However, our observed dissociation between response types is not identical to the dissociation by Marcel. This could be taken as an indication that we cannot simplistically compare our results with Marcel's. We see this as an open question demanding further research.

Another possible critique is that detection threshold can vary greatly within subjects over a large number of trials (Howarth & Bulmer, 1956). Thus, one could argue that the observed effect is the result of one condition (post or pre cue) being prior to the other rather than an effect of the order of events (the cue being before or after the stimulus). We analysed for this possibility by looking at the number of correct responses for each subject the first and last 50 trials in each condition. If the effect is as we claim, we should expect little difference between the first and last trials within each condition, and certainly no systematic difference. If the results are confounded in the here mentioned way, we should expect an increasing difference as a result of trial number and little difference between

the last 50 trials of the first condition and the first 50 of the second, the first being immediately followed by the latter.

As shown in table 6, there is no systematic difference in correctness over time, indicating that variation of detection threshold has had no relevant influence on our results.

| | | Set of trials | | | |
|---|---|---|---|---|---|
| | | **1A** | **2A** | **1B** | **2B** |
| **Subject** | **01** | 44 | 41 | 46 | 42 |
| | **02** | 25 | 26 | 31 | 30 |
| | **03** | 40 | 43 | 46 | 37 |
| | **04** | 30 | 37 | 38 | 27 |
| | **05** | 28 | 33 | 21 | 31 |
| | **06** | 43 | 46 | 42 | 47 |
| | **07** | 44 | 41 | 45 | 40 |
| | **08** | 41 | 42 | 30 | 40 |
| | **09** | 38 | 39 | 39 | 37 |
| | **10** | 35 | 41 | 28 | 27 |

*Table 6.* Variation of detection threshold over time.
The table shows the number of correct responses for the first 50 stimuli in the pre-cue condition (1A), the last 50 stimuli in the pre-cue condition (2A), and the same stimulus sets in the post cue condition (1B and 2B).

## General Discussion

In this paper, in the course of 4 experiments and an analysis of an experiment conducted by Anthony Marcel, we first of all argue for the possibility of separating introspective states of consciousness from non-introspective ones. Second, we argue that the introspective instruction changes the perceptual process when the subject is informed about which way to respond while perceiving a stimulus introspectively, compared to perceiving without a preparedness to respond in a specific way or when not introspecting. We claim that this result indicates that subjects are introspective when perceiving the stimulus, and that this is the cause of what looks like noise.

There are important alternative interpretations that come to mind when inspecting the data. One possibility is that subjects in the introspective condition somehow are slowed in their generating a report, thus giving rise to a difference in effect of memory decay. However, RT data presented for experiment 1 as well as other aspects of the results (e.g. there being a decrease of incorrect reports in the introspective condition) argue against this possibility. Another alternative interpretation is that the introspective task simply puts ‘extra load’ on attention. With this interpretation, also, it seems intuitively strange that subjects have less incorrect responses in the introspective condition, while the two dual tasks in experiment 2 certainly do generate more incorrect responses. However, one may defend the possibility that the introspective condition puts only a discreet weight on attention and that our two dual tasks demanded much more.

If our interpretation is valid, it has certain consequences. Firstly, it has certain theoretical consequences. As indicated earlier, there seems to be more to introspection or 'meta-awareness' than a mere 'seeing in by looking out' as Dretske writes (1995). Thus, the conceptual difference between introspection and first order states of consciousness is empirically validated. Our results go against the broadly conceived notion going back to William James (1898) that introspection is always retrospection. This view states that one cannot be introspective about ongoing mental events but only inspect one's conscious states in memory's 'rear view mirror'. Our findings indicate the opposite: That introspection occurs during perception. This amounts to the reality of different kinds of perceptual attitude.

The findings have methodological consequences also. One such consequence relates to the search for neural correlates of consciousness ('NCCs'). In experiments concerning NCCs, subjects are generally asked to give reports about their conscious states. This is so because a report is the only means at our disposal by which we can objectively distinguish between conscious and non-conscious instances of a particular mental content in a subject. Reports could thus be said to act as a necessary methodological tool when looking for NCCs. The necessity of introspective reports in consciousness research is especially underlined in our experiment 1, showing that they differ from non-introspective reporting techniques as employed in so-called objective measures of consciousness (Merikle & Daneman, 1998b). If the two types of reporting do not result in comparable data, it seems more difficult to defend the use of guessing in the study of consciousness. However, introspective techniques give rise to a possible serious source for confounding in NCC research. Given that introspection is a separate mental state – a view our data could be taken to support – how is one to identify which brain activations are related to first order conscious states and which are related to introspection? We believe our experiments 3 and 4 take the first step towards one possible solution to this problem. If it proves possible to track down introspective processes in their relation to cognitive processes and specific brain activations, one may be able to distinguish brain activations related to introspection from those related to first order consciousness. This, however, would demand much further research.

It is, of course, essential to the interpretation above that our experimental subjects actually introspected and that the difference in introspective attitude is the factor behind the results. We have had no objective way to check for this, and simply had to rely on two factors, i.e. that 1) subjects knew themselves whether they followed instructions, and 2) that they told the truth in the interview subsequent to each experiment. We find it impossible to avoid that one makes use of subjective data and that one takes it seriously if one is to do experimental research on conscious states of any kind.[5]

[5] Examples of recent work that does give constructive guidelines for the serious use of subjective reports would be Jack & Roepstorff (2002) or Price & Aydede (draft).

*Acknowledgements*

Morten Overgaard was supported for this work by a grant from The Carlsberg Foundation.

Both authors wish to thank Tony Jack, Tony Marcel, Søren Kyllingsbæk, Claus Bundesen, Kim Mouridsen and Oliver Kauffmann for important discussions. Also thanks to Thomas Z. Ramsøy, Christine d'Auchamp, Patrick Renault and Robin Døngart for their help collecting a portion of the data. Finally, we wish to thank three reviewers for their constructive suggestions.

## References

Block, N. (1995), 'On a confusion about a function of consciousness', *Behavioral & Brain Sciences,* **18** (2), pp. 227–47.

Chalmers, D.J. (1997), 'Availability: The cognitive basis of experience', *Behavioral and Brain Sciences,* **20** (1), pp. 148–9.

Dretske, F. (1995), *Naturalizing the Mind* (Cambridge, MA: MIT Press).

Green, D.M. & Swets, J.A. (1966), *Signal Detection Theory and Psychophysics (Chichester:* John Wiley & Sons).

Howarth, C.I. & Bulmer, M.G. (1956), 'Nonrandom sequences in visual threshold experiments', *Quarterly Journal of Experimental Psychology,* **8***,* pp. 163–71.

Jack, A. (1998), 'Perceptual awareness in visual masking', Ph.D. dissertation, http://www.nil.wustl.edu/labs/corbetta/personnel/ajack.html

Jack, A. & Roepstorff, A. (2002), 'Retrospection and cognitive brain mapping: From stimulus-response to script-report', *Trends in Cognitive Science,* **6** (8), pp. 333–9.

Jack, A. & Shallice, T. (2001), 'Introspective physicalism as an approach to the science of consciousness', *Cognition* special issue edited by S. Dahaene; *The Cognitive Neuroscience of Consciousness* (Cambridge, MA: MIT Press).

James, W. (1898), *Principles of Psychology* (New York: Dover Publications).

Lane, R. (2000), 'Neural correlates of conscious emotional experience', in *Cognitive Neuroscience of Emotion*, ed. R. Lane *et al.* (Oxford: Oxford University Press).

Lutz, A., Lachaux, J.P., Martinerie, J. & Varela, F. (2002), 'Guiding the study of brain dynamics by using first-person data: synchrony patterns correlate with ongoing conscious states during a simple visual task', *Proceedings of the National Academy of Sciences USA,* **99** (3), pp. 1586–91.

Lycan, W.G. (1996), *Consciousness and Experience* (Cambridge, MA: MIT Press).

Marcel, A. (1993), 'Slippage in the unity of consciousness', in *Experimental and Theoretical Studies of Consciousness*, ed. G.R. Bock & J. Marsh, J. (Chichester: John Wiley & Sons).

Marcel, A. (2003), 'Introspective report: Trust, self knowledge and science', *Journal of Consciousness Studies*, **10** (9-10), pp. 167–86.

Merikle, P.M., & Daneman, M. (1998), 'Psychological investigations of unconscious perception', *Journal of Consciousness Studies*, **5** (1), pp. 5–18.

Overgaard, M., Kauffmann, O. & Ramsøy, T.Z. (2001), 'Consciousness and introspection', *Consciousness Research Abstracts* (A service provided in cooperation with the *Journal of Consciousness Studies*), Abstract no. 182.

Price, D.D. & Aydede, M. (draft), 'The experimental use of introspection in the scientific study of pain and its integration with third-person methodologies: The experiential-phenomenological approach'.

Rosenthal, D. (1990), 'A theory of consciousness', Report no. 40, Research Group on Mind and Brain, Zentrum für Interdisziplinäre Forschung, Bielefeld, Germany.

Schooler, J. (2002), 'Re-representing consciousness: Dissociations between experience and meta-consciousness', *Trends in Cognitive Sciences*, **6** (8), pp. 339–44.

Searle, J. (1992), *The Rediscovery of the Mind* (Cambridge, MA: MIT Press).

Zahavi, D. & Parnas, J. (1998), 'Phenomenal consciousness and self-awareness: A phenomenological critique of representational theory', *Journal of Consciousness Studies,* **5** (5–6), pp. 687–705.