Background

In a much-publicized article, Bem (2011) claimed that people can look into the future. The experiment that attracted most media attention was Experiment 1, in which, on each trial, participants first viewed two curtains on either side of the computer screen. Participants had been told that behind one of the curtains was a picture, and that it was their job to try and predict, on each trial, which curtain was hiding a picture. After participants selected one of the curtains, the computer randomly determined whether that curtain should be replaced by a picture or not. The content of the pictures could be neutral, positive, negative, romantic but not erotic, or erotic. For all conditions, chance performance is 50%. However, Bem (2011) reported that participants were able to predict the location of erotic pictures (and only erotic pictures) with above-chance accuracy, namely 53.1%.

In a critique, Wagenmakers et al. (2011) suggested that the result may have been due to exploratory and suboptimal elements in the design of the study and the analysis of the data. To assess the validity of this explanation we will attempt to replicate Bem's (2011) Experiment 1. Our more general goal is to demonstrate the requirements of strictly confirmatory design and the advantages of Bayesian hypothesis testing.

In order to ensure that the data analysis and hypothesis tests are 100% confirmatory, this document describes an experiment and the associated data analysis before a single participant has been tested.

Method

Participants. We test only Dutch-speaking women. Participants have to speak Dutch because they are requested to fill out a personality questionnaire worded in Dutch (see below); participants have to be female because Bem (2011) claimed that the erotic stimuli from the IAPS database are affective only for women.

We will test a minimum for 52 participants (i.e., 13 groups of four). As explained below, we monitor the Bayes factor as the participants come in (in groups of four) and stop whenever the result is sufficiently clear. For instance, we could stop whenever the Bayes factor in favor of H_1 or H_0 is 30, but we could stop at any time we feel out point has been proven or disproven, or whenever we run out of time, money, or patience (see Edwards, Lindman, & Savage, 1963).

Stimulus Materials. The questionnaire for extraversion is part of a personality questionnaire that is commonly used at the University of Amsterdam (i.e., the "5PFT", Elshout & Akkerman). For the main experiment we used the IAPS picture data base to select 45 neutral pictures and 15 erotic pictures (as in Bem, 2011). The neutral pictures represent tools or utensils such as a towel, a rolling pin, or a drinking glass. The erotic pictures feature naked bodies but do not show genitalia.

Design. The experiment features two sessions of 60 trials each. Each trial entails the possible presentation of a picture, for a maximum of 45 neutral pictures and 15 erotic pictures. Session 2 features the same stimuli as session 1 but has a different random sequence of presentation. With respect to presentation, the important design variables are picture location (left or right) and picture type (neutral or erotic). Random presentation of the pictures may result in undesirable sequences; for instance, consider a hypothetical four-trial experiment with sequences <left> <left><left><left> for picture location or <neutral><neutral><neutral><erotic> for picture type. The latter case is particularly troublesome because effects of practice may be misattributed to effects of picture type. It is true that by means of purely random presentation the effects of these potential confounds eventually washes out. However, stricter control over these variables is possible, allowing for a more efficient test. Here we use yoking to minimize the impact of random sequences of picture location and type. Specifically, for participant *i* we randomize both picture location and picture type, separately for two sessions of 60 trials. This random sequence forms the basis for the sequences presented to three other participants. That is, participant *i+1* was confronted with exactly the same random sequence, but with opposite picture location (i.e., in the above example: <right> <right><right><right>). Participant *i+2* was also confronted with the same random sequence as participant *i*, but now with picture type sequence in reverse order (i.e., the sequence for participant *i* was presented backwards for participant *i+2*: <erotic> <neutral><neutral><neutral>). Finally, participant *i+3* was show a random sequence that is the opposite from participant *i* both for picture location and picture type. This yoking procedure creates groups of four participants in which the unwanted effects of picture location and type are experimentally controlled.

Procedure. Participants first sign an informed consent form; the form states that participants may be confronted with erotic pictures. Next, participants fill out the personality questionnaire. For the experiment itself, participants are tested in individual booths. All instructions are shown on the computer screen. Participants first experience a three-minute video featuring scenic landscapes and piano music. The goal of this is to put participants in a relaxed state of mind, as in Bem (2011). Next,

participants receive on-screen instructions for the main task, that is, they are asked to guess which of two curtains hides a picture. For each trial, the curtains are displayed until the participant makes a selection. Participants press the "z" key to select the left curtain and the "?/" key to select the right curtain. Immediately after the key press the selected curtain is replaced either with a picture (neutral or erotic) or with a white patch. Picture or patch are visible for 1 second. Next, a blank screen with an asterisk in the middle is presented for 1 second, after which the next trial starts. The experiment features two sessions of 60 trials each. Session 2 immediately follows session 1 without notification to the participant. The entire experiment takes approximately 30 minutes.

Data Analysis

Data analysis proceeds by a series of Bayesian tests. For the Bayesian t-tests, the null hypothesis H0 is always specified as the absence of a difference. Alternative hypothesis 1, H1, assumes that effect size is distributed as Cauchy (0,1); this is the default prior proposed by Rouder et al. (2009). Alternative hypothesis 2, H2, assumes that effect size is distributed as a half-normal distribution with positive mass only and the 90th percentile at an effect size of 0.5; this is the "knowledge-based prior" proposed by Bem et al. (submitted). We will compute the Bayes factor for H0 vs. H1 (BF_01) and for H0 vs. H2 (BF_02).

We will conduct the following analyses:

(1) Based on the data of session 1 only: Does performance for neutral pictures differ from performance for erotic pictures? To address this question we compute a paired t-test and monitor BF_01 and BF_02 as the data come in.

(2) Based on the data of session 1 only: Does performance for erotic pictures differ from chance (in this study 50%)? To address this question we compute a one-sample t-test and monitor BF_01 and BF_02 as the data come in.

(3) Based on the data of session 1 only: Is there a positive correlation between extraversion scores and performance for erotic pictures? This possibility was suggested by Bem (2011), and we assess this claim using the default Bayesian test for correlation proposed by Jeffreys (1961).

(4) If participants have ESP, this trait should be related from session 1 to session 2. In other words, individual differences in ESP express themselves statistically as a positive correlation between

performance on erotic pictures for session 1 and session 2. This prediction will again be tested using the default Bayesian test for correlation proposed by Jeffreys (1961).

(5) Same as (1), but now for the combined data from session 1 and session 2.

(6) Same as (2), but now for the combined data from sessions 1 and 2.

The final version of this document was made available online the 9th of June, 2011. Experimentation starts later the same day.

Researchers responsible: Adam Sasiadek, Boris Pinksterboer, Esther Lietaert Peerbolte, Rebecca Schild, Ruud Wetzels, Rogier Kievit, and Eric-Jan Wagenmakers.