
Participants
Participants were recruited from humanities undergraduates (in various disciplines excluding psychology) at Université Lumière, Lyon. They took part in this experiment as the first part of a paid one-hour session. There were 18 participants (12 women, 6 men), aged 18 to 33 (\(M = 22, \text{SD} = 3\)).

Materials and norming
The recall material consisted in a two-page story adapted from Barrett (1996). A diplomat is about to be sent as an ambassador to a distant galaxy. He goes to the local museum to get a better idea of what to expect over there. Between introduction (arrival at the museum) and end (return home), the main part of the text is a list of 24 short descriptions of exhibits in the museum. We used four versions of the stories, with identical items in different orders.

The story included 12 items for each of the Artifact and Person categories. In each category, six items described a breach of intuitive expectations (“Br” items) and the other six an expected association (“Sn” items). Person items consisted in colloquial descriptions of psychological features taken from the “theory of mind” literature, for example, “there were people who could see what was in front of them” (Sn) or “there were people who could see through a wall” (Br). The artifact items described physical features taken from the “intuitive physics” literature, for example, “pieces of furniture that you can move by pushing them” (Sn) and “pieces of furniture that float in the air if you drop them” (Br). Order of presentation was counterbalanced.

All items were pretested with 18 students and staff of Université Lumière in three different conditions: asking them to rate the items as “normal v. abnormal,” “banal v. surprising” and “familiar v. unfamiliar.” Items that reached less than 90% consensus were discarded. The remaining items were slightly modified to result in a similar sentence structure and roughly similar word-count. In the story, all items were described in two sentences, the second one being a straightforward paraphrase of the first.

A first questionnaire form (Questionnaire 1–1) included all the items in the recall text, in a different order. Instructions read as follows: “The following sentences describe the different exhibits Mr. Wurg saw in the museum. For each sentence, indicate whether you find the object or person described similar or different from what we usually encounter in reality, by checking the appropriate box.”

A second questionnaire (1–2) had the same list of items and different instructions: “The following sentences describe the different exhibits Mr. Wurg saw in the museum. Indicate whether each sentence describes an object or person that you have previously encountered, either in reality or in films, stories, cartoons, and so forth by checking the appropriate box.”

Design and procedure
This was a 2 (category) x 2 (level) design with both category (artifact vs. person) and level (standard items vs. breaches of expectations) as within-subject variables. The participants were all tested individually in an experimental booth. They were given the printed text of the story. The instructions were to read the story very carefully and try to imagine each situation described in the story. When they reached the end of the text they were instructed to hand back the text to the experimenter. As a distraction task, they were then asked to do some mental arithmetic and to multiply the number of vowels in various words. After this 5 min distraction task, they were then given ruled sheets and instructed to write down as many of the exhibits in the museum as they could recall without regard for item order but with as much detail as they could recall. This part of the test was limited to ten minutes.

They were then given the two questionnaires and instructed to give spontaneous, literal responses and avoid metaphorical interpretations of the items. For the second questionnaire (1–2),
the experimenter first asked the participant to explain the difference with questionnaire 1–1 and used two “training” questions to check that both fictional and real familiar items were assigned to the same category. The participants were then de-briefed about the aims of the experiment and asked whether they saw any connection between the items and religious notions.