

MAGICONTROL: A Normed Video Database of Magic Effects and Matched Control Events for Experimental Research

Miguel De Lucas and Jon Andoni Duñabeitia * 

Centro de Investigación Nebrija en Cognición (CINC), Universidad Nebrija, 28043 Madrid, Spain; mromeroluis@alumnos.nebrija.es

* Correspondence: jdunabeitia@nebrija.es

Abstract

Magic effects are staged events in which a performer uses concealed conjuring methods to produce an apparently impossible outcome, such as an object vanishing, appearing, or transforming. Because these events violate observers' ordinary causal expectations, they provide useful stimuli for investigating epistemic emotions and expectation violation. Yet, existing stimulus databases often lack carefully matched control conditions and contain uncontrolled social variables. To address these methodological gaps, this study introduces MAGICONTROL, an open-access video database comprising 100 short clips divided into 50 genuine magic effects and 50 closely matched non-magical control events. All stimuli were recorded from a fixed frontal perspective by a single professional performer, intentionally eliminating speech and audience reactions to minimize social confounds. The database was normed via an online study with 112 participants, assessing trick recognition alongside subjective ratings for curiosity, interest, and surprise. Descriptive analyses showed comparable mean durations and file sizes across conditions, while pair-level analyses quantified the degree of temporal matching within each pair. Normative results demonstrate high overall trick classification accuracy alongside substantial, interpretable item-level variability across the 10-point Likert ratings for curiosity, interest, and surprise. Ultimately, MAGICONTROL offers a standardized, ecologically valid, and highly controlled paired stimulus resource that supports controlled comparisons intended to isolate the contribution of the magical outcome. This database supports reproducible experimental research in cognitive science, specifically concerning causal reasoning, attention, prediction error, and affective responses to impossible events.

Keywords: magic tricks; magic effects; misdirection; neuroscience of magic; illusionism



Academic Editor: Athanasios S. Drigas

Received: 26 August 2026

Revised: 8 October 2026

Accepted: 10 October 2026

Published: 13 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Summary

Magic is one of the most universally recognized elicitors of surprise and wonder in human history, insofar as it consists of a systematic violation of the physical regularities that normally guide observers' predictions about the world [1,2]. The scientific study of magic has expanded considerably in recent decades [3]. Neuroimaging research has shown that the perception of causally impossible events recruits neural systems associated with expectation violation and cognitive conflict, including the dorsolateral prefrontal cortex and the anterior cingulate cortex [4]. These findings have positioned magic-based stimuli as uniquely valuable tools for investigating the cognitive and affective consequences of prediction error under relatively naturalistic viewing conditions [1,5].

The emotional responses elicited by magic effects extend beyond mere surprise to encompass positively valenced epistemic states such as interest and curiosity. Within contemporary emotion theory, interest is conceptualized as a knowledge-oriented emotion, functionally characterized by sustained attentional engagement and an intrinsic drive toward exploration [6]. Curiosity, although closely related to interest, is functionally distinguishable from it. It involves a motivational state generated by a perceived gap in one's knowledge, that is, a discrepancy between what observers know and what they perceive they could know, which promotes information-seeking behavior aimed at reducing uncertainty [7]. In the context of magic, these constructs operate in tandem: the apparent physical impossibility of an effect creates an immediate epistemic gap that the observer cannot resolve, thereby simultaneously eliciting curiosity about the hidden method and sustained interest in the stimulus itself. This relationship is supported by robust behavioral and neuroimaging evidence; for instance, Lau et al. [8] demonstrated that magic tricks reliably induce curiosity, to the extent that participants were willing to receive electric shocks in order to learn the solution, and that this motivational effect shares neural substrates with the desire to satisfy basic needs, with especially relevant activation in the ventral striatum.

Surprise is conceptualized as a discrete emotional state triggered by events that are discrepant with the observer's cognitive schema. Functionally, this state is characterized by an immediate and automatic interruption of ongoing processing, a reorientation of attentional resources toward the unexpected stimulus, and the activation of causal analysis and schema-updating processes [9]. The magnitude of surprise is determined by the degree of discrepancy between the event and the observer's schema, rather than by the novelty or valence of the eliciting event per se [9]. At the computational level, this framework aligns precisely with predictive coding accounts, in which surprise can be operationalized as a signal proportional to prediction error, namely the discrepancy between the prior expectations of the observer's generative model and the incoming sensory evidence [5]. Under this formulation, magic effects constitute a privileged class of stimuli: they systematically violate highly automatized physical priors, generating strong prediction errors under naturalistic observation conditions [1,2].

The empirical investigation of these constructs requires ecologically valid and standardized stimulus sets that reliably elicit surprise and related epistemic emotions across participants. Several studies have generated their own magic-video stimuli for this purpose [4,10], but these materials have not generally been made available to the scientific community, thereby limiting replicability and cumulative progress in the field. The main open-access database of magic stimuli currently available is the Magic Curiosity Arousing Tricks database, or MagicCATs [11], which comprises 166 short video clips of magic tricks validated for their capacity to elicit epistemic emotions such as curiosity, interest, and surprise. The MagicCATs database has subsequently been validated in additional cultural and linguistic contexts, including Italian samples of young and middle-aged adults [12] and a broader Italian sample of participants aged between 18 and 86 years [13]. Consequently, this validated stimulus set represents a significant opportunity for the scientific community to conduct rigorous and replicable research on epistemic emotions.

Despite its relevance, MagicCATs presents two methodological limitations for some experimental purposes. First, its stimulus set includes videos performed both in front of live audiences and directly to camera, which may introduce social contagion as an uncontrolled source of variance in observers' emotional responses. Audience reactions, laughter, applause, or visible spectator engagement can modulate the perceived intensity, relevance, and affective impact of the observed event. Second, MagicCATs does not include matched control versions for each magic effect, which prevents the direct experimental isolation of expectancy disconfirmation as the causal component of interest. This limitation

is not unique to MagicCATs. Previous experimental studies using magic stimuli have often developed ad hoc control conditions in which the same general sequence of actions is presented without the impossible outcome, precisely to isolate the cognitive and neural correlates specifically attributable to violations of physical expectation [4,10]. Thus, the reliance on such heterogeneous, custom-made control conditions across different laboratories severely hinders direct comparisons between studies and limits the cumulative progress of the discipline.

The present study addresses these gaps by developing and publishing an open-access database of controlled video stimuli for the study of surprise, curiosity, interest, and related responses to magic. The database consists of 100 short video clips equally distributed between a magic condition and a matched control condition. All videos were recorded directly to camera, without a live audience, speech, or interpersonal interaction, thereby minimizing social and linguistic confounds. Each magic video depicts a genuine conjuring effect, whereas its corresponding control video reproduces the same performer, setting, frontal perspective, approximate duration, object context, and action structure without including the magical outcome. The stimulus set spans multiple domains of close-up and stage magic, including card magic, coin magic, silk magic, cup magic, bottle magic, oversized-card effects, and other object-based routines. These routines encompass varying levels of technical complexity, presenting an inherent execution challenge due to the inability to deploy traditional misdirection techniques that rely on direct spectator interaction [14].

The database was normed through an online study implemented in Gorilla Experiment Builder [15], with participants recruited through Prolific. The norming protocol replicated the core epistemic-emotion dimensions assessed in MagicCATs, including surprise, interest, and curiosity, and complemented them with binary measures of trick detection and comprehension. By combining matched TRICK-CONTROL pairs with item-level normative ratings, the resulting database provides a reusable experimental resource for researchers interested in expectation violation, causal reasoning, attention, curiosity, surprise, memory, affective responses to impossible events, and the cognitive science of magic. The following sections describe the development of the video stimuli, the construction of the matched control clips, the norming procedure, the structure of the released dataset, and the technical validation analyses supporting its reuse.

2. Data Description

The MAGICCONTROL dataset is deposited in Figshare under the Digital Object Identifier (DOI) <https://doi.org/10.6084/m9.figshare.32603220>. The repository contains the final video stimuli and two comma-separated values (CSV) files describing the technical, item-level, and normative properties of the database. The video files are organized in a folder named MAGICCONTROL Videos, which includes all final clips in MPEG-4 Part 14 (MP4) format with their definitive stimulus names. The two associated data files are MAGICCONTROL_video_data.csv and MAGICCONTROL_norming_data.csv.

The MAGICCONTROL Videos folder contains the 100 final video stimuli. The stimuli are organized into 50 matched TRICK-CONTROL pairs. TRICK videos are named using the structure TRICK_XX.mp4, and CONTROL videos are named using the structure CONTROL_XX.mp4, where XX identifies the pair number. Thus, for example, TRICK_01.mp4 and CONTROL_01.mp4 correspond to the two members of the same matched pair. Each TRICK video depicts a genuine magic effect, whereas the corresponding CONTROL video reproduces the same general action context without the magical outcome. All videos are provided in MP4 format. Pair identifiers preserve the labels assigned during construction of the original 54-pair stimulus set. Consequently, identifiers in the final release are not

fully consecutive: pairs 19, 24, 27, and 51 were excluded during final quality control and are not included in the released video files or data records (see below for details).

The file `MAGICCONTROL_video_data.csv` provides item-level information for the 50 TRICK-CONTROL pairs. The file is organized in a wide format, with one row per pair. Each row contains the file name and technical properties of the TRICK video, the normative scores associated with that TRICK video, the file name and technical properties of the corresponding CONTROL video, and a pair-level duration-matching index. This file is intended to allow researchers to select individual stimuli or complete pairs according to their technical and normative characteristics. The columns in `MAGICCONTROL_video_data.csv` are as follows. `TRICK_VIDEO` contains the file name of the TRICK video. `TRICK_SIZE_Mb` contains the file size of the TRICK video in megabytes. `TRICK_DURATION_s` contains the duration of the TRICK video in seconds. `ACCURACY` contains the percentage of participants who correctly identified the TRICK video as depicting a magic trick. `CURIOSITY`, `INTEREST`, and `SURPRISE` contain the mean normative ratings for the corresponding TRICK video on the three epistemic-emotion scales. `CONTROL_VIDEO` contains the file name of the matched CONTROL video. `CONTROL_SIZE_Mb` contains the file size of the CONTROL video in megabytes. `CONTROL_DURATION_s` contains the duration of the CONTROL video in seconds. `DEVIATION_RATIO` contains the pair-level duration mismatch between the TRICK and CONTROL videos. `DEVIATION_RATIO` was calculated to quantify temporal matching within each TRICK-CONTROL pair. For each pair, the absolute difference between the duration of the TRICK video and the duration of the CONTROL video was divided by the duration of the longer video. The index can therefore be expressed as:

$$\text{DEVIATION_RATIO} = \frac{|\text{duration_TRICK} - \text{duration_CONTROL}|}{\max(\text{duration_TRICK}, \text{duration_CONTROL})}$$

This metric is bounded between 0 and 1. A value of 0 indicates identical durations for the two videos in a pair, whereas higher values indicate larger relative duration differences. Because the denominator is the duration of the longer video, the index expresses the mismatch as a proportion of the maximum possible duration within that pair. This makes the value directly interpretable across video pairs of different absolute lengths. The file also includes two descriptive columns. `MATERIAL` lists the objects used in each pair, separated by semicolons. `BASE_TRICK` classifies the magical effect of each pair according to the taxonomy of Lamont and Wiseman [16], which distinguishes nine basic types of effect: appearance, vanish, transposition, transformation, penetration, restoration, extraordinary feats, telekinesis, and extrasensory perception.

The file `MAGICCONTROL_norming_data.csv` contains the trial-level responses from the final included participants in the norming study. The file is organized in long format, with one row per participant, video, trial, and question. This structure preserves the original response-level organization of the norming task and allows researchers to reconstruct participant-level, item-level, or question-level summaries according to their own analytical needs, and it derives from the online norming study described above and in <https://app.gorilla.sc/openmaterials/1288748> (accessed on 30 July 2026).

The columns in `MAGICCONTROL_norming_data.csv` are as follows. `PARTICIPANT` contains the anonymized participant identifier. `AGE` contains the participant's age in years. `GENDER` contains the participant's self-reported gender. `VIDEO` contains the name of the video presented on that trial. `TRIAL` indicates the randomized presentation order of the video for that participant. `QUESTION` identifies the question presented after the video. `RESPONSE` contains the participant's response to that question.

The QUESTION column contains five possible values. MAGICTRICK corresponds to the binary question “Did you see a magic trick?”, with responses coded as YES, NO, or TIMEOUT. UNDERSTAND corresponds to the binary question “Did you understand the intention of the magic trick?”, with responses coded as YES, NO, or TIMEOUT. SURPRISE corresponds to the question “How surprised were you at the magic trick?”. INTEREST corresponds to the question “How interesting was the magic trick?”. CURIOUS corresponds to the question “How curious were you about how the magic trick was done?”. The three rating questions were answered using a 10-point Likert scale ranging from 1, “Not at all”, to 10, “Very much”. In the RESPONSE column, these scale responses are stored as values from 1 to 10, with TIMEOUT indicating that no response was provided within the allowed response window.

The TRIAL column should be interpreted as the position of the video within each participant’s randomized sequence, not as a fixed property of the video. Thus, the same video may appear at different trial positions for different participants. This information is retained in the dataset to allow users to examine potential order effects, fatigue effects, or presentation-sequence effects if required.

Together, the two CSV files provide complementary levels of information. MAGICCONTROL_video_data.csv offers a compact item-level description of the paired stimulus set, including technical properties, normative scores, and the duration-matching index. MAGICCONTROL_norming_data.csv provides the full response-level normative data from the final included sample, allowing independent aggregation, reanalysis, quality control, or alternative scoring procedures. The combination of the final MP4 files, item-level metadata, and trial-level norming data is intended to support transparent and reproducible reuse of the MAGICCONTROL stimulus set.

3. Methods

3.1. Stimulus Development

The initial stimulus set consisted of 108 short video clips equally distributed between two experimental categories: 54 magic videos, hereafter referred to as TRICK videos, and 54 matched non-magic videos, hereafter referred to as CONTROL videos. Each TRICK video depicted a genuine conjuring effect, whereas its corresponding CONTROL video reproduced the same general scenario, performer, setting, frontal viewing perspective, and approximate temporal structure without including the magical outcome. This matched-pairs architecture was intended to isolate the violation of expectation produced by the conjuring effect as the critical manipulated component. Similar magic-control contrasts have been used in previous neuroimaging studies examining expectation violation through magic tricks, including work contrasting magic trick clips with control clips in which the expected action-outcome relationship was preserved [10]. The present database extends this logic to a normed stimulus resource specifically designed for reuse in experimental research. While the initial stimulus set comprised 54 matched TRICK-CONTROL pairs, after completion of the norming study and before public release, a final item-level quality-control review was conducted. Two pairs (19 and 27) were removed because their TRICK effects were highly similar to effects already represented in the stimulus set, thereby reducing the diversity of the final resource. Two further pairs (24 and 51) were removed because their CONTROL videos showed false-positive magic-trick classification rates exceeding 20% of valid responses. In each case, both members of the pair were excluded to preserve the one-to-one matched structure of the database. The final released dataset therefore comprises 50 TRICK-CONTROL pairs, corresponding to 100 videos.

The videos were recorded during two filming sessions conducted in a professional magic-learning studio under controlled recording conditions. A multicamera system was

initially used to capture the performances from different viewpoints. After inspecting the available recordings, the final stimulus set was created using a fixed frontal camera angle that provided the clearest view of the magical effects and approximated the perspective of a spectator seated in front of the performer. The camera remained fixed across recordings, and the relevant actions were performed directly toward the lens, as if the camera occupied the position of the observer. Crucially, this recording setup was not an arbitrary choice, but a highly deliberate methodological decision. By maintaining this static, direct-to-lens perspective, the observer is afforded the opportunity to witness the magic effects clearly, free from extraneous movements or the attention-diverting misdirection typical of standard performances.

Regarding the technical setup, although the recording setup included additional cameras, only the fixed frontal camera was used to produce the stimuli; footage from the other cameras was not used, and no camera switching was applied, so that changes in shot or viewpoint could not act as an additional source of attentional cueing. This frontal camera was a Sony FDR-AX33 4K Handycam (Sony Group Corporation, Tokyo, Japan) with an integrated ZEISS lens (Carl Zeiss AG, Oberkochen, Germany), mounted on a fixed tripod. During recording, the camera signals were routed via HDMI to a Blackmagic Design ATEM Mini Pro switcher (Blackmagic Design Pty. Ltd., Melbourne, Australia) and recorded directly to the computer's hard drive at a resolution of 1920×1080 pixels and 50 frames per second. The scene was illuminated frontally with three Neewer NL660 bi-color LED panels (40 W, 3200–5600 K; Shenzhen Neewer Technology Co., Ltd., Shenzhen, China), set to a warm color temperature of approximately 3500–4000 K to optimize the visibility of the objects and actions, and lighting conditions were kept constant across recording sessions. Video editing and cropping were performed in Adobe Premiere Pro (version 26.5.1, build 1), and the final files were exported with Adobe Media Encoder (version 26.5, build 85).

The magic effects were performed by a Spanish professional magician, who is also the first author of the article. The stimulus set covered several domains of close-up and stage magic, including card magic, coin magic, silk magic, cup magic, bottle magic, oversized-card effects, and other object-based routines. This diversity was intentional, as different types of magical effects may vary in causal structure, perceptual salience, object familiarity, apparent impossibility, and emotional impact. Including a heterogeneous set of effects therefore increases the potential utility of the database for researchers interested in expectation violation, attention, curiosity, surprise, memory, causal reasoning, and affective responses to impossible events. Importantly, the selection of these stimulus objects was strictly guided by their immediate familiarity to the observer; thus, no obscure or unconventional items requiring prior explanation or introduction were utilized. In addition, the effects were selected according to the following explicit inclusion criteria: (1) they had to be comprehensible through visual information alone, (2) executable without speech or audience interaction, (3) clearly visible from a fixed frontal camera, (4) suitable for the construction of a non-magical counterpart, (5) representative of a variety of phenomena and materials, and (6) performable with high technical quality. The selection did not aim to equate surprise across effects, but rather to cover a useful range of intensity, so that item-level variability constitutes an intended property of the database.

All recordings were designed to maintain a high degree of visual homogeneity. The magician wore black clothing in all videos, either a black shirt or a black shirt with a black jacket. No video included verbalizations, dialogue, or interaction with additional people. The absence of a live audience was also deliberate, as it prevents observers' responses from being influenced by spectator reactions, laughter, applause, or other forms of social contagion. This differentiates the present stimulus set from the MagicCATs database [11]. In the present database, the removal of audience feedback, speech, and social interaction

was intended to focus attention on the unfolding visual event itself. We acknowledge that many magicians use patter to build the expectations that are subsequently violated, and that its absence may reduce the impact of some effects at the cost of some ecological richness. For this reason, the database does not aim to represent the full repertoire of magic. It includes only visually self-sufficient effects and excludes routines whose efficacy depends essentially on patter, audience interaction, or verbal framing, in line with previous work using silent magic stimuli [10]. Moreover, omitting speech avoids linguistic and prosodic differences between TRICK and CONTROL videos and makes the stimuli independent of any particular language, facilitating their use across linguistic and cultural contexts. Additionally, all video clips were completely muted to eliminate any incidental auditory cues, such as the physical handling of objects, that could inadvertently betray the magic method or introduce unwanted sensory variance.

The general scene was kept constant across videos. The performer was positioned in front of the camera in a controlled studio environment, and the critical action took place in the central visual field. The relevant objects and gestures were presented from the same frontal perspective, without changes in viewpoint, camera movement, or narrative context. Facial visibility was minimized whenever possible during editing to reduce the potential influence of facial expressions and other social cues. This design choice was intended to concentrate participants' attention on the objects, manual actions, and transformations occurring in the central region of the scene. This visual restriction is particularly advantageous for eye-tracking and neuroimaging paradigms, as it prevents the confounding activation of face-processing brain networks and reduces extraneous exploratory gaze behaviors.

Each magic effect was recorded repeatedly until the magician, the second author, and an additional professional magician acting as an expert observer agreed that the performance was clear, fluent, and suitable for inclusion in the database. These selected recordings constituted the initial pool of TRICK videos. Once a satisfactory version of a given magic effect had been obtained, the magician produced a corresponding CONTROL version. In these control recordings, the performer attempted to emulate the general movements, timing, handling style, object configuration, and visual structure of the original magic effect, while avoiding any procedure or outcome that could be perceived as a trick, illusion, or impossible event. This procedure yielded a paired database rather than two independent sets of clips. Each CONTROL video was constructed as the closest possible non-magical counterpart of a specific TRICK video. To confirm the validity of the CONTROL recordings, the independent expert verified that they successfully depicted natural actions and were entirely free of any lingering deceptive visual cues or unresolved anomalies. The resulting stimulus set therefore allows researchers to compare responses to magical and non-magical events while controlling, as far as possible, for performer identity, recording context, visual framing, object category, action structure, and video duration.

3.2. TRICK-CONTROL Pairing Procedure

The pairing procedure was designed to preserve the perceptual and event-level characteristics of each magic video while removing the defining magical outcome. The critical difference between each TRICK-CONTROL pair was the presence or absence of the conjuring effect. This pairing followed a one-to-one architecture by which each CONTROL video was recorded as the specific counterpart of a single TRICK video, using the same objects, initial configuration, sequence of manipulations, and approximate duration, but with a physically expected outcome. Therefore, the 50 CONTROL videos do not constitute a mere generic pool of non-magical clips. Quite the contrary, each of them was required to preserve item-level matching, so that any TRICK video could be selected together with

its corresponding visual baseline. The similarity between the two videos of each pair was verified by an independent professional magician, and their temporal matching was quantified for each pair using the Relative Deviation Ratio (see below). For example, in TRICK_03 (a vanish effect), the magician shows two coins, places one slightly overlapping the other, passes his hand over them, and the two coins merge into a single coin. In the corresponding CONTROL video, the magician performs the same sequence of actions, but when the hand passes over the coins, he shows both coins to the camera again. Thus, the movements in both videos are virtually identical, with the only difference being that the CONTROL video contains no violation of expectation.

After the initial recordings had been completed, all TRICK and CONTROL videos were edited in pairs. The beginning of each clip was trimmed so that the video started with the first task-relevant action, avoiding the inclusion of preparatory movements, pauses, or events unrelated to the effect itself. The editing process also aimed to equate the duration of each TRICK-CONTROL pair as closely as possible, while preserving the natural temporal structure of the action. To ensure the scientific integrity of the stimuli, all post-production and editing adjustments were strictly administrative and never altered the continuous, real-time performance of the magic effects; thus, the database is entirely free of camera tricks, hidden cuts, or digital manipulation.

All final videos were exported in MP4 format using H.264 encoding. The standardized video dimensions were 720 pixels in width and 600 pixels in height, corresponding to an aspect ratio of 1.20. This format was selected to provide a stable and homogeneous presentation frame across the whole database. During editing, the visual field was centered on the relevant action area.

Thus, the final stimulus set contained 100 videos: 50 TRICK videos and 50 matched CONTROL videos. The TRICK videos had a mean duration of 9.78 s, with a range from 3.53 to 24.30 s and a standard deviation of 4.28 s. The CONTROL videos had a mean duration of 9.21 s, with a range from 3.90 to 26.47 s and a standard deviation of 4.54 s. Thus, the two video categories were closely matched in their overall temporal characteristics. To quantify duration matching at the pair level, we used the Relative Deviation Ratio (DEVIATION_RATIO; see Section 2 for its definition). Across the 50 retained pairs, the mean Relative Deviation Ratio was 0.12, with a minimum of 0, a maximum of 0.400, and a standard deviation of 0.10.

File size was also documented as part of the technical characterization of the database. TRICK videos had a mean file size of 4.76 MB, ranging from 1.78 to 11.63 MB, with a standard deviation of 2.04 MB. CONTROL videos had a mean file size of 4.42 MB, ranging from 1.96 to 12.62 MB, with a standard deviation of 2.06 MB. These values indicate that the two stimulus categories were comparable not only in duration and format, but also in basic digital properties relevant for experimental presentation, storage, and online deployment.

3.3. Norming Procedure

The norming study was implemented online using Gorilla Experiment Builder [15], and participants were recruited through Prolific. The study was designed to obtain approximately 100 classification responses or exposures per video. This target was defined a priori based on previous normative work with magic-trick video stimuli. In the original MagicCATs database, each video was evaluated by approximately 95 to 107 participants. The present study therefore aimed to match the upper range of the original MagicCATs norming density, while extending the design to a paired TRICK-CONTROL database.

Before launching the full study, the experimental procedure was piloted with five members of the research team. Based on this pilot, the task duration was fixed at 30 min, starting from the presentation of the first video. The pilot indicated that participants

could evaluate approximately 100 videos within this time window. Given that the full database contained 108 videos, the initial sampling plan estimated that slightly more than 100 participants would be required to obtain the desired number of ratings per video. Accordingly, the Prolific study was initially completed by 125 participants before exclusions were applied.

All participants provided informed consent before starting the task. The study protocol was approved by the Ethics Board of Universidad Nebrija under approval code UNNE-2025-0119 and was conducted in accordance with standard ethical principles for behavioral research. All Gorilla materials needed to reproduce the norming study are openly available at: <https://app.gorilla.sc/openmaterials/1288748>.

3.3.1. Participants

Participants were eligible for inclusion if they were born in the United States, currently resided in the United States, held US nationality, and reported English as their native and primary language. Additional inclusion criteria required normal or corrected-to-normal vision, no reported dyslexia or reading problems, no reported neurodiversity, and an age between 25 and 50 years.

After applying the predefined exclusion criteria (see below for details), the final sample consisted of 112 participants. The sample was gender-balanced, with 56 women and 56 men. The mean age of the final sample was 39.09 years, with a standard deviation of 6.81 years.

3.3.2. Task Structure

Each trial started with a 3-s countdown, which was included to orient participants to the upcoming stimulus and focus their attention on the screen. Immediately after the countdown, the video was presented with autoplay. Each video was presented only once, and participants could not replay it before responding. This single-exposure procedure was adopted to capture participants' immediate responses to each effect, as in a live performance, and to prevent repeated viewing from facilitating the detection of the method, which would reduce the experience of expectation violation [17]. It also ensured that all participants received the same amount of exposure to each stimulus. Once the video ended, the response screens were displayed sequentially. Participants had a maximum of 5 s to respond to each question. If no response was entered within the allotted time, the response was coded as a timeout. A blank screen of 500 ms was inserted between consecutive response screens.

The full task lasted 30 min from the onset of the first video and was terminated automatically once this time limit had elapsed. Participants therefore evaluated as many videos as possible within the fixed task duration. In the final included sample, participants viewed an average of 97.95 videos, with a range from 65 to 108 videos and a standard deviation of 7.43.

3.3.3. Normative Questions

The norming protocol combined binary classification and comprehension questions with Likert-type ratings of epistemic emotions. The first question was "Did you see a magic trick?", with the binary response options "Yes" and "No". This question was used to determine whether participants classified the video as depicting a magic trick. For videos classified as not containing a magic trick, the trial ended, and the participant proceeded to the next video. For videos classified as containing a magic trick, participants completed the remaining questions. This conditional branching was strategically implemented to mitigate participant fatigue and to ensure that ratings of epistemic emotions were exclusively grounded in active experiences of expectancy violation.

The second binary question was “Did you understand the intention of the magic trick?”, with the response options “Yes” and “No”. This item did not assess whether participants had discovered the method or solution of the trick, but whether they had understood what happened in the video, that is, which magical effect was intended to be perceived. Accordingly, we hereafter refer to this variable as comprehension of the intended magic effect. This item was included following the logic of the clarity measure used in MagicCATs and its later validations, where participants indicated whether they understood the intention or content of the trick. In those studies, this binary clarity question was used both as a descriptive measure of stimulus comprehensibility and as a data-quality criterion. Ozono et al. [11] described this measure as assessing whether participants understood what happened after seeing the video clip, and later Italian validations retained the same conceptual structure. By retaining this item, our database allows researchers to separate genuine surprise resulting from expectation violation from cognitive noise caused by mere ambiguity or lack of comprehension of the magical narrative.

Participants then answered three rating questions using 10-point Likert scales, where 1 indicated “Not at all” and 10 indicated “Very much”. The first rating question was “How surprised were you at the magic trick?”; the second question was “How interesting was the magic trick?”; and the third question was “How curious were you about how the magic trick was done?”. These three questions were selected because surprise, interest, and curiosity are the core epistemic-emotion dimensions assessed in the original MagicCATs database and in subsequent validation studies.

3.3.4. Exclusion Criteria

Three exclusion criteria were defined a priori. First, participants were excluded if more than 10% of their responses were coded as timeouts. This criterion was used to remove participants who failed to respond consistently within the task constraints. In the final included sample, the mean percentage of timed-out responses per participant was 0.35% with a range from 0% to 5.80% and a standard deviation of 0.84%. Second, participants were excluded if they classified fewer than 70% of the videos correctly as TRICK or CONTROL in response to the question “Did you see a magic trick?”. This criterion was included to ensure that participants were attending to the critical manipulation distinguishing magic from non-magic videos. Of the 125 participants who completed the experiment, 10 participants showed classification accuracy below 70% and were excluded on this basis. Among the final included participants, mean classification accuracy was 93%, with a range from 73% to 100% and a standard deviation of 6%. Third, participants were excluded if they answered “No” to the comprehension question “Did you understand the intention of the magic trick?” for all videos that they had classified as TRICK. This criterion parallels the use of the clarity question in MagicCATs and its validation studies as an indicator of whether participants understood the intended event structure of the trick. After removing participants who failed the classification criterion, 3 additional participants met this 0% comprehension criterion and were excluded. After applying all exclusion criteria, the final normative sample comprised the 112 participants described above.

3.3.5. Rating Coverage per Video

The final sampling density was consistent with the a priori goal of obtaining approximately 100 ratings per video. Across the 100 videos, the mean number of participants who responded to each video was 102, with a minimum of 95, a maximum of 109, and a standard deviation of 2.97. Thus, the final dataset achieved highly balanced rating coverage across videos and closely reproduced the item-level rating density of the original MagicCATs normative procedure.

3.4. Technical Validation

The technical validation of the MAGICCONTROL database focused on confirming the normative characterization of the TRICK videos and ensuring participant response quality and technical matching (as detailed in Sections 3.2 and 3.3). The original norming study was conducted on 54 TRICK-CONTROL pairs. Before release, four complete pairs were excluded during final item-level quality control, as described in Section 3.1. Pairs 19 and 27 were removed because their TRICK effects were highly similar to other retained effects, whereas pairs 24 and 51 were removed because their CONTROL videos exceeded a 20% false-positive magic-trick classification rate. The final normative values reported below therefore correspond to the retained 50 pairs.

The normative distributions of the 50 TRICK videos were inspected for classification accuracy and for the three epistemic-emotion ratings: curiosity, interest, and surprise. Figure 1 shows the distribution of item-level means for these four variables. Across valid classification responses, the 50 TRICK videos were identified as magic tricks with a sensitivity of 89% (95% Wilson CI [89%, 91%]). The 50 CONTROL videos were correctly classified as non-magic with a specificity of 96% (95% Wilson CI [96%, 97%]). Balanced accuracy was 93% (bootstrap 95% CI [92%, 94%]). The overall false-positive rate for CONTROL videos was 4%.

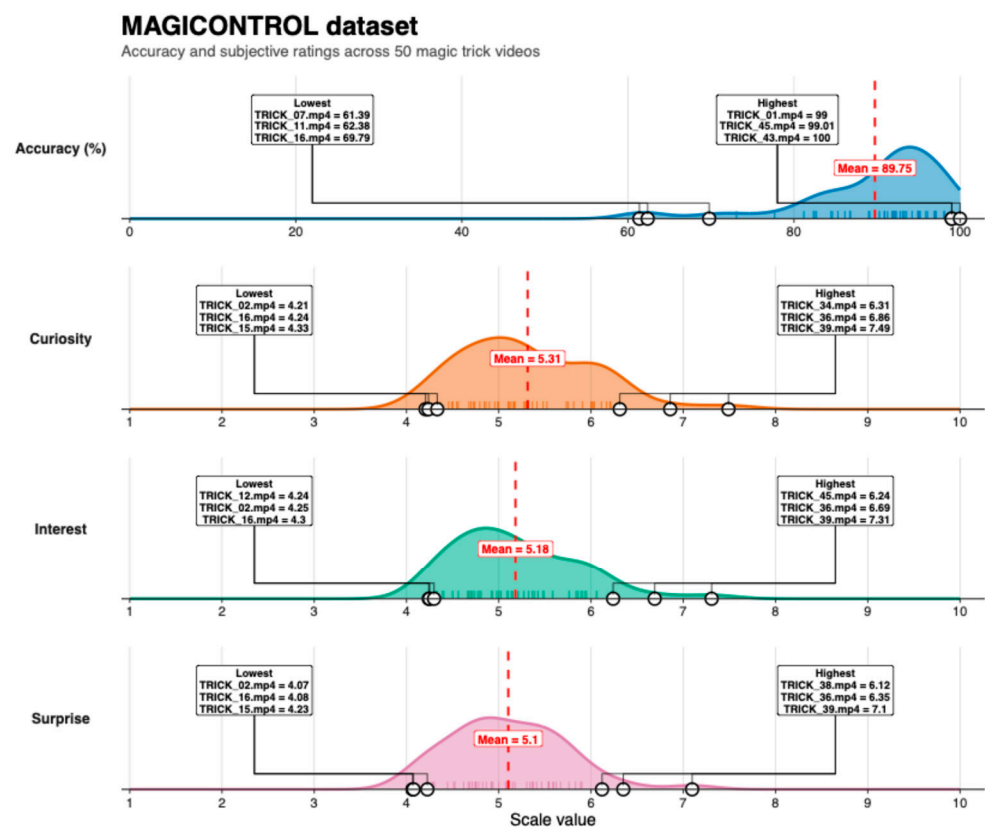


Figure 1. Distribution of normative accuracy and epistemic-emotion ratings across the 50 TRICK videos. Density plots show the distribution of item-level means for classification accuracy, curiosity, interest, and surprise. Vertical dashed red lines indicate the mean value of each distribution. Rug marks represent individual TRICK videos. Boxed annotations identify the three videos with the lowest and highest values for each variable. Accuracy corresponds to the percentage of participants who correctly identified each TRICK video as depicting a magic trick. Curiosity, interest, and surprise correspond to mean ratings on 10-point Likert scales, where 1 = “Not at all” and 10 = “Very much”.

Item-level false-positive rates for the retained CONTROL videos are shown in Figure 2. Rates were generally low, with a median of 3.0% and a maximum of 12.1%. Only two

of the 50 CONTROL videos exceeded a 10% false-positive rate. A mixed-effects logistic regression with condition as a fixed effect and random intercepts for participant and matched pair confirmed a strong classification difference between TRICK and CONTROL videos (OR = 449.0, 95% CI [362.0, 559.0], $p < 0.001$). This result was robust in a sensitivity model using random intercepts for participant and video (OR = 542.0, 95% CI [373.0, 788.0], $p < 0.001$). These results are useful for stimulus selection, as they allow researchers to choose highly recognizable magic effects with their correctly matched controls.

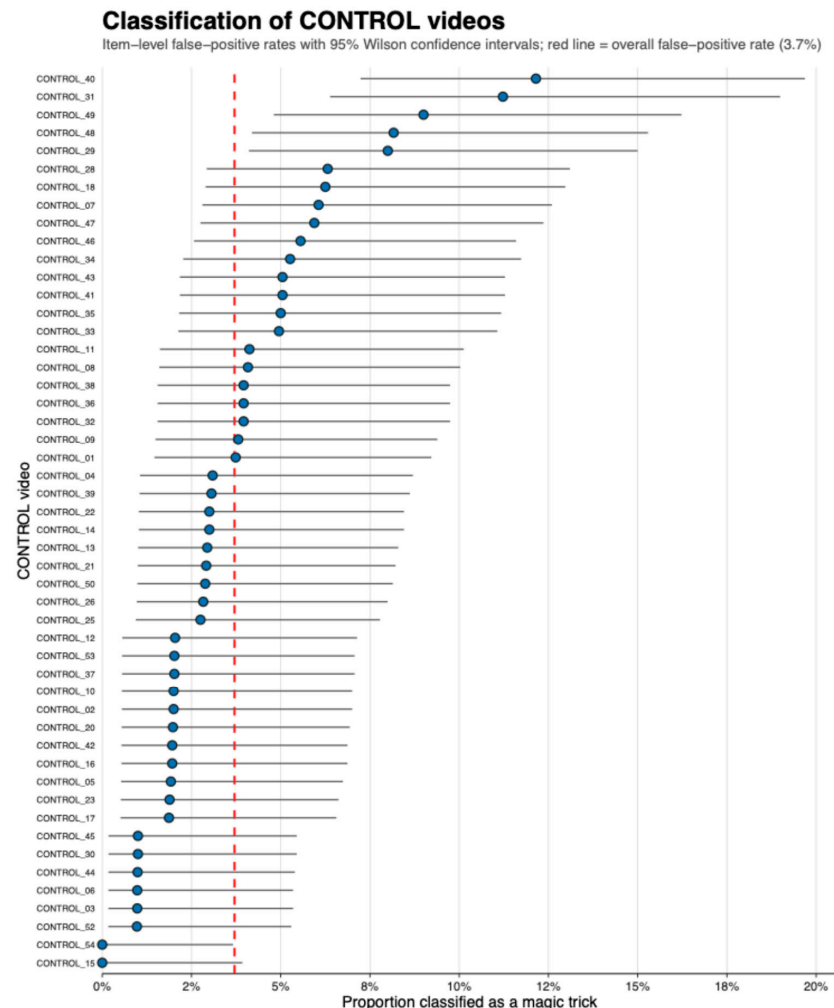


Figure 2. Item-level false-positive magic-trick classifications for the 50 retained CONTROL videos. Points represent the proportion of valid classification responses in which each CONTROL video was classified as a magic trick. Horizontal error bars indicate 95% Wilson confidence intervals. The vertical dashed red line indicates the overall false-positive rate across CONTROL videos.

The three subjective rating dimensions showed moderate mean values and substantial item-level variation. Mean curiosity was 5.31 on the 10-point scale (SD = 0.72), mean interest was 5.18 (SD = 0.67), and mean surprise was 5.10 (SD = 0.62). The lowest item-level means were approximately 4.21 for curiosity, 4.24 for interest, and 4.07 for surprise, whereas the highest item-level means reached 7.49 for curiosity, 7.31 for interest, and 7.10 for surprise. These distributions indicate that the database contains stimuli spanning a broad range of epistemic-emotion intensity rather than a restricted set of uniformly high- or low-intensity effects. This range is valuable for experimental designs requiring stimulus selection based on affective intensity, item variability, or parametric modulation.

Exploratory mixed-effects analyses were conducted to examine whether normative responses varied according to age and gender, considering the sufficiently distributed age

of the participants and the gender-balanced nature of the sample. For TRICK classification, a mixed-effects logistic regression included condition, age (centered and scaled in 10-year units), gender, and the interactions of condition with age and gender, with random intercepts for participant and video. The TRICK–CONTROL classification difference increased with age (condition $\times$ age: OR = 1.85 per 10-year increase, 95% CI [1.41, 2.42], $p < 0.001$). There was no clear evidence that this classification contrast differed by gender (condition $\times$ gender: OR = 1.27, 95% CI [0.87, 1.83], $p = 0.214$).

Separate linear mixed-effects models were fitted for curiosity, interest, and surprise ratings among TRICK trials correctly classified as showing a magic effect. Age was associated with modestly lower surprise ratings ($\beta = -0.55$ points per 10-year increase, 95% CI [−1.10, −0.00], $p = 0.0498$). Estimates for curiosity ($\beta = -0.56$, 95% CI [−1.18, 0.06], $p = 0.077$) and interest ($\beta = -0.50$, 95% CI [−1.06, 0.06], $p = 0.077$) were similar in direction but statistically uncertain. As shown in Figure 3, gender differences were small across curiosity ($\beta = 0.08$, 95% CI [−0.76, 0.92]), interest ($\beta = 0.21$, 95% CI [−0.54, 0.97]), and surprise ($\beta = 0.08$, 95% CI [−0.67, 0.83]). These analyses are exploratory and are limited to the sampled age range. Hence, they should not be interpreted as evidence for developmental effects outside adults aged 25–50 years.

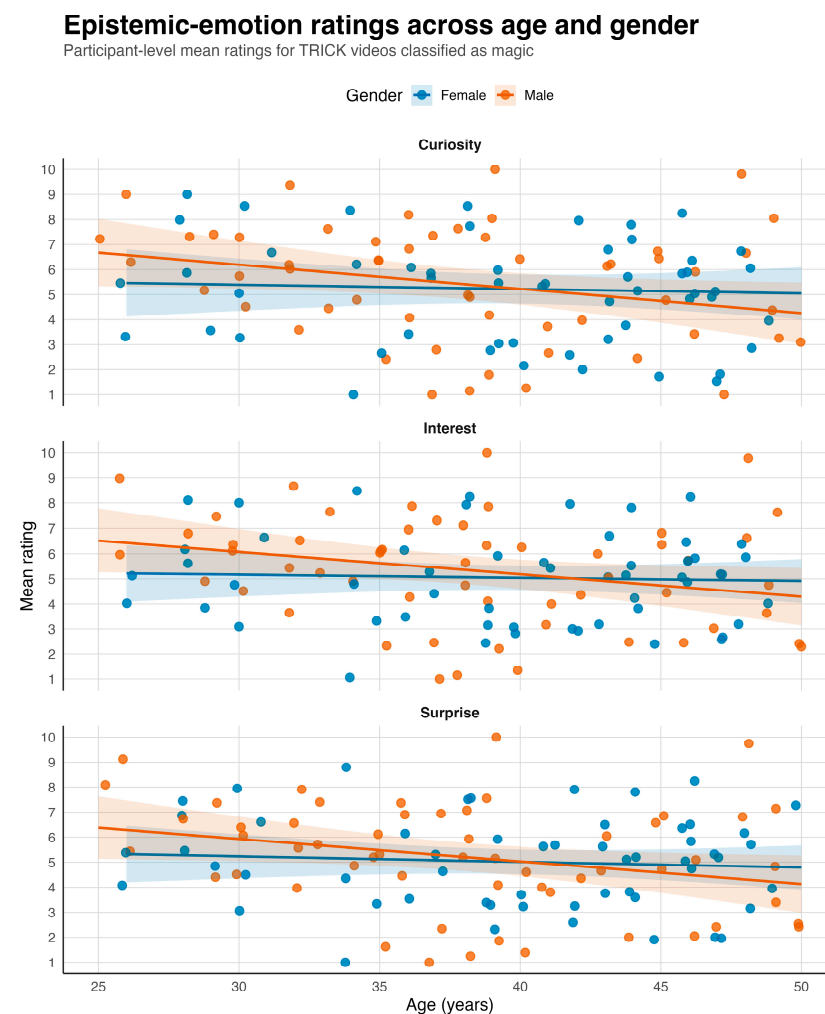


Figure 3. Participant-level epistemic-emotion ratings across age and gender. Each point represents a participant's mean curiosity, interest, or surprise rating across TRICK videos that the participant classified as magic tricks. Colors denote self-reported gender. Lines show separate linear trends for female and male participants, with shaded bands representing 95% confidence intervals.

Overall, the technical validation supports the use of MAGICONTROL as a reusable paired stimulus database. The final set of 50 TRICK videos displayed interpretable variability in recognition and epistemic-emotion ratings, and the TRICK-CONTROL pairs were technically standardized and temporally matched after removal of redundant effects and CONTROL videos with elevated false-positive classification rates. Researchers can therefore select stimuli according to video-level recognition accuracy, curiosity, interest, surprise, duration, file size, or pair-level duration matching.

4. User Notes

The MAGICONTROL dataset can be used in several ways depending on the goals of the study. Researchers interested in robust magic recognition may select TRICK videos with high classification accuracy, whereas researchers interested in ambiguity, individual differences, or uncertainty may select videos with intermediate accuracy values. Researchers interested in affective intensity may select stimuli according to their normative curiosity, interest, or surprise scores. Because these scores are provided at the item level in `MAGICONTROL_video_data.csv`, users can select subsets of videos that maximize or minimize specific normative dimensions, or create balanced stimulus sets across several dimensions simultaneously.

The paired structure of the database is central to its intended use. Each TRICK video has a corresponding CONTROL video, matched in the features described in Section 3.1 but without the magical outcome. Researchers interested in expectation violation can therefore compare responses to TRICK and CONTROL videos while controlling, as far as possible, for multiple perceptual and contextual features. This structure may be especially useful in studies examining prediction error, causal reasoning, attention, memory, curiosity, surprise, interest, and affective responses to apparently impossible events. The normative scores for curiosity, interest, and surprise should be interpreted in relation to the structure of the norming procedure. These ratings were collected after participants indicated that they had seen a magic trick. Consequently, the item-level affective scores primarily characterize the subjective response to the TRICK videos as perceived magic events. They should not be interpreted as fully symmetrical TRICK-CONTROL ratings unless users derive and justify additional scoring procedures from the raw norming data.

The open Gorilla materials associated with the norming task can also be used to reproduce, adapt, or extend the normative procedure, as in recent validations and cross-cultural extensions of the original MagicCATs by Ozono et al. [11] (see [12,13]). Researchers may use the same task architecture to validate the stimuli in other languages, age groups, cultural contexts, or experimental settings. The database can also be combined with additional measures, including eye tracking, pupil dilation, electroencephalography, functional neuroimaging, physiological recordings, memory tests, causal explanation tasks, confidence ratings, or post-video curiosity-resolution measures.

Finally, several limitations should also be considered when reusing the MAGICONTROL dataset. The normative data were obtained from US participants aged 25–50 years who reported English as their native and primary language, so additional validation may be needed for other linguistic, cultural, or age groups. Also, all effects were performed by a single professional magician, ensuring consistency while limiting performer variability. Although TRICK-CONTROL pairs were matched in performer, setting, camera perspective, action context, and approximate duration, they were not computationally matched for low-level visual features such as luminance, contrast, optical flow, or visual salience. In addition, participants' prior familiarity with each specific effect was not measured, which may account for part of the item-level and individual variability in the ratings. Curiosity, interest, and surprise ratings were collected only after participants had indicated that

they had seen a magic trick. Consequently, these ratings primarily characterize responses to videos perceived as magic events and should not be interpreted as fully symmetric TRICK–CONTROL affective measures. We recommend that researchers using the database assess such familiarity when it is relevant to their research goals. Lastly, because the videos contain no speech, the findings may not generalize to magic routines whose psychological impact depends on the performer’s patter or verbal framing.

In sum, MAGICONTROL provides an open-access, normed, and paired video stimulus database for the experimental study of magic, expectation violation, and epistemic emotions. The final dataset includes 50 genuine magic effects and 50 matched control events, together with item-level metadata, pair-level duration-matching information, and trial-level normative responses from a final sample of 112 participants. By combining controlled TRICK–CONTROL pairs with normative ratings of classification accuracy, curiosity, interest, and surprise, MAGICONTROL offers a reusable resource for research on prediction error, causal reasoning, attention, memory, curiosity, surprise, and the cognitive science of magic. The database is intended to support transparent stimulus selection, reproducible experimental design, and cumulative research using magic-based stimuli under controlled conditions.

Author Contributions: Conceptualization, M.D.L. and J.A.D.; methodology, M.D.L. and J.A.D.; software, J.A.D.; validation, M.D.L. and J.A.D.; formal analysis, J.A.D.; investigation, M.D.L. and J.A.D.; resources, M.D.L. and J.A.D.; data curation, J.A.D.; writing—original draft preparation, M.D.L.; writing—review and editing, M.D.L. and J.A.D.; visualization, J.A.D.; supervision, J.A.D.; project administration, M.D.L. and J.A.D.; funding acquisition, M.D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033), grant number PID2024-161331NB-I00.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Universidad Nebrija (protocol code UNNE-2025-0119 and date of approval 15 October 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data supporting reported results can be found in Figshare: <https://doi.org/10.6084/m9.figshare.32603220>.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (GPT-5.6 Sol Pro; OpenAI, San Francisco, CA, USA) only as a language-editing aid to improve clarity, grammar, and readability. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Kuhn, G.; Amlani, A.A.; Rensink, R.A. Towards a science of magic. *Trends Cogn. Sci.* **2008**, *12*, 349–354. [[CrossRef](#)] [[PubMed](#)]
2. Rensink, R.A.; Kuhn, G. A framework for using magic to study the mind. *Front. Psychol.* **2015**, *5*, 1508. [[CrossRef](#)] [[PubMed](#)]
3. De Lucas, M.; Duñabeitia, J.A. Magic and cognition: A scoping review of empirical evidence on cognitive and neural processes. *Front. Psychol.* **2026**, *in press*. [[CrossRef](#)] [[PubMed](#)]
4. Parris, B.A.; Kuhn, G.; Mizon, G.A.; Benattayallah, A.; Hodgson, T.L. Imaging the impossible: An fMRI study of impossible causal relationships in magic tricks. *NeuroImage* **2009**, *45*, 1033–1039. [[CrossRef](#)] [[PubMed](#)]
5. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138. [[CrossRef](#)] [[PubMed](#)]
6. Silvia, P.J. Interest—The curious emotion. *Curr. Dir. Psychol. Sci.* **2008**, *17*, 57–60. [[CrossRef](#)]
7. Loewenstein, G. The psychology of curiosity: A review and reinterpretation. *Psychol. Bull.* **1994**, *116*, 75–98. [[CrossRef](#)]

8. Lau, J.K.L.; Ozono, H.; Kuratomi, K.; Komiya, A.; Murayama, K. Shared striatal activity in decisions to satisfy curiosity and hunger at the risk of electric shocks. *Nat. Hum. Behav.* **2020**, *4*, 531–543. [[CrossRef](#)] [[PubMed](#)]
9. Reisenzein, R.; Horstmann, G.; Schützwohl, A. The cognitive-evolutionary model of surprise: A review of the evidence. *Top. Cogn. Sci.* **2019**, *11*, 50–74. [[CrossRef](#)] [[PubMed](#)]
10. Danek, A.H.; Öllinger, M.; Fraps, T.; Grothe, B.; Flanagan, V.L. An fMRI investigation of expectation violation in magic tricks. *Front. Psychol.* **2015**, *6*, 84. [[CrossRef](#)] [[PubMed](#)]
11. Ozono, H.; Komiya, A.; Kuratomi, K.; Hatano, A.; Fastrich, G.; Raw, J.A.L.; Haffey, A.; Meliss, S.; Lau, J.K.L.; Murayama, K. Magic Curiosity Arousing Tricks (MagicCATs): A novel stimulus collection to induce epistemic emotions. *Behav. Res. Methods* **2021**, *53*, 188–215. [[CrossRef](#)] [[PubMed](#)]
12. Padulo, C.; Ponticorvo, M.; Fairfield, B. The Magic Curiosity Arousing Tricks (MagicCATs) database in Italian younger and middle-aged adults: Descriptive statistics and rule-based machine learning. *Behav. Res. Methods* **2026**, *58*, 1. [[CrossRef](#)] [[PubMed](#)]
13. Marascia, E.; Di Crosta, A.; La Malva, P.; Ceccato, I.; Prete, G.; Palumbo, R.; Mammarella, N.; Di Domenico, A. Exploring curiosity, interest, and surprise: Normative ratings of a magic trick video dataset in Italy. *Sci. Data* **2026**, *13*, 578. [[CrossRef](#)] [[PubMed](#)]
14. Garcia-Pelegrin, E.; Schnell, A.K.; Wilkins, C.; Clayton, N.S. Beyond the tricks: The science and comparative cognition of magic. *Annu. Rev. Psychol.* **2024**, *75*, 269–293. [[CrossRef](#)] [[PubMed](#)]
15. Anwyl-Irvine, A.L.; Massonnié, J.; Flitton, A.; Kirkham, N.; Evershed, J.K. Gorilla in our midst: An online behavioral experiment builder. *Behav. Res. Methods* **2020**, *52*, 388–407. [[CrossRef](#)] [[PubMed](#)]
16. Lamont, P.; Wiseman, R. *Magic in Theory: An Introduction to the Theoretical and Psychological Elements of Conjuring*; University of Hertfordshire Press: Hatfield, UK, 1999.
17. Danek, A.H.; Fraps, T.; von Müller, A.; Grothe, B.; Öllinger, M. Working wonders? Investigating insight with magic tricks. *Cognition* **2014**, *130*, 174–185. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.