Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/20/$12.00

2020, Vol. 46, No. 2, 257–279
http://dx.doi.org/10.1037/xlm0000723

Gaze Data Reveal Individual Differences in Relational Representation Processes

Joshua Zonca
University of Trento

Giorgio Coricelli
University of Trento and University of Southern California, Los Angeles

Luca Polonio
University of Trento

In our everyday life, we often need to anticipate the potential occurrence of events and their consequences. In this context, the way we represent contingencies can determine our ability to adapt to the environment. However, it is not clear how agents encode and organize available knowledge about the future to react to possible states of the world. In the present study, we investigated the process of contingency representation with three eye-tracking experiments. In Experiment 1, we introduced a novel relational-inference task in which participants had to learn and represent conditional rules regulating the occurrence of interdependent future events. A cluster analysis on early gaze data revealed the existence of 2 distinct types of encoders. A group of (sophisticated) participants built exhaustive contingency models that explicitly linked states with each of their potential consequences. Another group of (unsophisticated) participants simply learned binary conditional rules without exploring the underlying relational complexity. Analyses of individual cognitive measures revealed that cognitive reflection is associated with the emergence of either sophisticated or unsophisticated representation behavior. In Experiment 2, we observed that unsophisticated participants switched toward the sophisticated strategy after having received information about its existence, suggesting that representation behavior was modulated by strategy generation mechanisms. In Experiment 3, we showed that the heterogeneity in representation strategy emerges also in conditional reasoning with verbal sequences, indicating the existence of a general disposition in building either sophisticated or unsophisticated models of contingencies.

Keywords: cognitive abilities, eye-tracking, individual differences, representation, strategy generation

Supplemental materials: http://dx.doi.org/10.1037/xlm0000723.supp

The main challenge we face in our everyday experience is adapting to the environment we live in. We need to foresee that some events might take place in the future and to be aware of the possible consequences of their occurrence (Schultz, Dayan, & Montague, 1997; Suddendorf & Corballis, 2007). However, our world is not always predictable: we can learn how to respond to a specific event, but we may not know whether this event will actually occur. For example, I know that I will have to take the bus if the train does not arrive, but the (non-) arrival of the train is in some way unforeseeable. In this context, the way we encode and organize relevant knowledge about the world (i.e., the type of environmental representation we generate) can affect our ability to respond to future events (Bar, 2007; Gilbert & Wilson, 2007). On the one hand, agents may build an exhaustive representation of the relational structure underlying interrelated contingencies and plan future behavior taking into consideration every predictable consequence of potential states. In our example, I am prepared for the possibility that the train does not arrive, and so I bring my bus pass to be ready to respond optimally to the occurrence of both states of the world. On the other hand, agents may learn only basic units of knowledge (e.g., binary associations between a state and an outcome), without building an explicit model of how these simple rules relate to each other. Only once a specific condition takes place, these latter agents would use stored knowledge to react to

This article was published Online First June 6, 2019.

Joshua Zonca, Center for Mind and Brain Sciences, University of Trento; Giorgio Coricelli, Center for Mind and Brain Sciences, University of Trento, and Department of Economics, University of Southern California, Los Angeles; Luca Polonio, Center for Mind and Brain Sciences, University of Trento.

Parts of the work described here were presented as an oral presentation at the 36th meeting of the European Group of Process Tracing Studies in Judgment and Decision Making (EGPROC), NUI Galway, June 24, 2017 and at the annual meeting of the Society for NeuroEconomics, Philadelphia, October 7, 2018. We thank Carlo Reverberi, Katya Tentori, Ben Timberlake, and Thibaud Gressinger for insightful comments. We gratefully acknowledge the financial support of the European Research Council (ERC Consolidator Grant 617629).

Correspondence concerning this article should be addressed to Luca Polonio, who is now at the Department of Economics, University of Minnesota, 4-101 Hanson Hall, Minneapolis, MN 55455. E-mail: lpolonio@umn.edu

that specific event. In our example, this representation process would lead me to realize that I need the bus pass only after apprehending that the train has not arrived, potentially catching me unprepared (i.e., I may have left the bus pass at home). These two types of representation processes express different degrees of sophistication: despite the latter behavior might be occasionally efficient, the former is more sophisticated because it is suitable for responding optimally to every predictable environmental contingency. Although this behavioral difference is reminiscent of the distinctions between rule abstraction and memorization in category learning (McDaniel, Cahill, Robbins, & Wiener, 2014), proactive and reactive cognitive control (Braver, 2012), model-based and model-free learning (Daw, Niv, & Dayan, 2005, 2011; Konovalov & Krajbich, 2016), and problem-model and direct-translation strategies in problem-solving (Boote et al., 2018; Mayer & Hegarty, 1996), it is still unclear how agents build internal contingency models starting from available relational knowledge. In particular, we should understand whether distinct processes of relational representation do exist, as well as the cognitive sources of this heterogeneity. To explore these issues, we ran three different eye-tracking experiments.

In Experiment 1, we designed a novel relational-inference task in which each trial was composed of two phases: representation and response. In the representation phase, participants had a limited amount of time to learn triplets of between-state rules connected by higher-order transitive relations (i.e., if the state X occurs, then the state Y follows; if the state Y occurs, then the state Z follows; if the state Z occurs, then the state W occurs as well). These pieces of information established the conditional relations regulating the occurrence of states, but did not provide information about their actual occurrence (i.e., participants know that state Y follows from the occurrence of state X, but do not know if state X will actually occur). In the response phase, the occurrence of a specific state was disclosed, and participants had to infer which other states necessarily followed given the relational model acquired in the representation phase. In the relational-inference task, we used eye-tracking to explore top-down attentional mechanisms including search, selection and binding of relevant information, which can reveal how agents spontaneously build representations of the current relational environment. Specifically, in the representation phase, we expect some (sophisticated) participants to explore the environment searching for all possible relational information to construct a representation that explicitly expresses all the existing relations between states. Conversely, unsophisticated agents should not explore the relational properties of the current relational set, because they do not aim to build a comprehensive model of the relational structure of the environment.

Results of a cluster analysis on early gaze data in the representation phase confirmed the existence of two distinct groups of participants that respectively exhibited sophisticated and unsophisticated behaviors, and showed marked differences in task performance.

To explore the cognitive mechanisms driving heterogeneity in representation behavior, in Experiment 2 we collected data on a new pool of participants performing the relational-inference task in two different sessions (pre- and post- treatment). In the pretreatment session, participants performed the relational-inference task with the same modalities of Experiment 1. At the beginning of the posttreatment session, the same participants were informed about the existence of sophisticated and unsophisticated strategies and their respective average efficiencies. Then they were asked to complete again the relational-inference task in the way they preferred. We therefore compared the representation strategy implemented by participants in the two sessions. We found a notable strategy switch from the unsophisticated toward the sophisticated strategy, suggesting that the implementation of a specific strategy is not driven by cognitive capacity or motivation, but rather by strategy generation mechanisms.

In Experiment 3, we investigated whether the heterogeneity in Experiment 1 and 2 could generalize to a verbal-inference task requiring conditional reasoning in real-life scenarios. The verbal-inference task differed from the relational-inference task in different ways. First, it included verbal instead of symbolic content, setting conditional reasoning in a more naturalistic context; second, task resolution was not dependent on short-term memory (STM) components and encoding time constraints. The verbal-inference task was completed by participants of Experiment 2, because we aimed to compare individual representation strategies in the two tasks. Results show that sophisticated participants, as defined in the relational-inference task, spontaneously adopted sophisticated representation behavior in the verbal-inference task, suggesting the existence of general, context-independent processes of encoding, integration, and representation of relational information between hypothetical states of the world.

Cognitive Drivers of Sophisticated and Unsophisticated Representation Processes

To date, we lack evidence about the contribution of cognitive abilities in modulating representation-building mechanisms. We can hypothesize that high working memory is necessary for the generation of sophisticated representations, because it constitutes the workspace where relational representations are constructed (Doumas, Hummel, & Sandhofer, 2008; Halford, Wilson, & Phillips, 2010), and guarantees that agents can build, retain and update representations (Oberauer, 2009). However, it is possible that working memory sustains active maintenance and manipulation of representations without affecting the type of representation that is generated. To investigate the role of working memory in these processes, we collected four different working memory measures: digit span forward and backward (Wechsler, 2008) and the *n*-back task (in two versions of increasing difficulty, 2-back and 3-back, Kirchner, 1958). The forward version of the digit span assesses simple short-term maintenance and recall of elements in working memory, whereas the backward version requires the additional component of mental manipulation of digits (Baddeley, 1996; Koenigs, Barbey, Postle, & Grafman, 2009; Monaco, Costa, Caltagirone, & Carlesimo, 2013). The *n*-back task tests the ability to maintain and update a dynamic set of information, targeting processes related to cognitive control, such as inhibition and interference resolution (Kane, Conway, Miura, & Colflesh, 2007).

Another cognitive ability that could intervene in the representation process is fluid intelligence, which expresses the ability to adapt to unknown contexts and reason on abstract information with minimal dependence on crystalized knowledge (Cattell, 1963). However, we do not know whether fluid intelligence intervenes in an early stage of representation generation or simply sustain updating and inferential mechanisms, as recently suggested by Shipstead, Harrison, and Engle (2016). To collect individual measures of fluid intelligence, we tested participants on the Raven Advanced Progressive Matrices Test (APM; Raven, Raven, & Court, 1998).

Finally, we investigated whether cognitive reflection, measured by the Cognitive Reflection Test (CRT; Frederick, 2005), could be a potential candidate to predict the existence of distinct representation processes. The CRT traditionally assesses the individual tendency to implement either reflective or reflexive cognitive processes. In particular, a high cognitive reflection level expresses the ability to reason exhaustively about the characteristics of a problem, inhibiting intuitive but incorrect responses. Conversely, a low cognitive reflection level indicates an aptitude for generating heuristics on salient information at the expense of problem understanding (Toplak, West, & Stanovich, 2011, 2014). In recent years, several studies have underlined the relevance of the CRT beyond the classical deliberation-intuition trade-off (Baron, Scott, Fincher, & Metz, 2015; Mata et al., 2014; Szaszi, Szollosi, Palfi, & Aczel, 2017). In particular, it has been linked to the tendency to use more thorough search processes (Cokely & Kelley, 2009; Cokely et al., 2009) and to the ability to accurately process and represent task-relevant information (Mata et al., 2014; Sirota, Juanchich, & Hagmayer, 2014). Furthermore, recent evidence pointed out that the CRT is related to analytical thinking (Hoppe & Kusterer, 2011), behavioral biases (Oechssler, Roider, & Schmitz, 2009), probabilistic reasoning (Koehler & James, 2010; Liberali, Reyna, Furlan, Stein, & Pardo, 2012), and rule abstraction (Don, Goldwater, Otto, & Livesey, 2016), suggesting a broader involvement of cognitive reflection in intelligent behavior.

Experiment 1

Method

Relational-inference task. In this novel task, participants were presented series of three conditional statements of the form “if A, then B” connecting pairs of symbols. Symbols represented states of the world whose occurrence was uncertain, whereas conditional relations between symbols prescribed the necessary occurrence of a state (e.g., B) upon the occurrence of another state (e.g., A). Importantly, conditional relations could be linked by transitive relations (e.g., given the two conditionals “if A then B” and “if B then C”, you can conclude that “if A then C”). Henceforth, we will refer to the three conditional statements as C1, C2 and C3. Four abstract symbols (square, circle, triangle and cross) were used to represent states (Figure 1, left panel). Using this set of items, we created 80 different relational sets. From all the possible combinations of symbols and relations, we excluded those including a specific symbol simultaneously repeated in all three antecedents or in all three consequents of the conditionals. Each configuration could contain zero, one, two, or three transitive relations connecting conditionals in up-down or down-up directions.

Each trial of the task consisted of two phases: representation and response. In the representation phase (Figure 1, left panel), participants had 9 seconds to learn all the relevant pieces of information in a series before their disappearance. In the response phase, one of the symbols presented in the representation phase (source state) was highlighted, meaning that that state had indeed occurred. Given this novel information and the conditional relations shown in the representation phase, participants had to select all the states (i.e., symbols) that necessarily followed the occurrence of the source state (Figure 1, right panel). There was no delay between the two phases. In the response phase, each of the four symbols was paired with a specific response key. An intuitive interface supported the Response phase (Figure 1, right panel). Key-symbol associations remained stable along the entire experiment.[1] Symbols could be pressed in any order. Participants had the opportunity to repress the same response-key to deselect or reselect a specific symbol. Participants were instructed that deselecting and reselecting symbols would not have affected their score; final selection was confirmed by pressing the space bar and only this response was taken into account for evaluation. In sum, a trial was classified as correct if participants selected all and only the states that necessarily followed the occurrence of the source state and as incorrect in all other cases. Participants had unlimited time in the response phase, and they were instructed that RTs would not influence their final score.

We created two different categories of relational set: linear and nonlinear. In linear sets, the order of the presented triplet of conditionals was aligned with the latent relational structure (i.e., with the ordered sequence of concatenated events; Figure 2, left panel). In nonlinear sets, the underlying relational structure did not match with the order of the presented triplet of conditionals (Figure 2, right panel). The presence of nonlinear trials allowed us to disentangle sophisticated from unsophisticated representation processes: sophisticated participants should indeed search for all possible relations between states in every potential direction and location, whereas unsophisticated participants should encode binary conditional rules independently of their higher-order relations.

All these aspects of the task were carefully explained to participants with examples, control questions, and training trials (we report full instructions and control questions in the online supplemental materials). In particular, we ensured that participants correctly understood all the conditional, transitive and spatial properties of the task. Participants were provided with three 2-min breaks (one every 20 trials). The order of trials was randomized across participants. Each trial was preceded by a fixation-point positioned in one of four possible locations outside the symbol space.

The task was made incentive-compatible by paying participants according to their proportion of correct responses (minimum 0, maximum 14 euros).

Visual search control task. The Visual search task served as a control for individual differences in visual scan efficiency. In this task, participants had to detect a target among a variable number of distractors. They were instructed to be as accurate and fast as possible, and they were reimbursed based on a scoring formula combining accuracy and RTs (see the Visual Search Control Task: Experimental Design section in Appendix A).

Cognitive measures.

Raven Advanced Progressive Matrices Test (APM). Participants performed the Raven Advanced Progressive Matrices Test (APM).

[1] We checked for possible effects attributable to the position of symbols and corresponding keys in the response interface and we did not find any effect of source state (see Table A1, Appendix A).

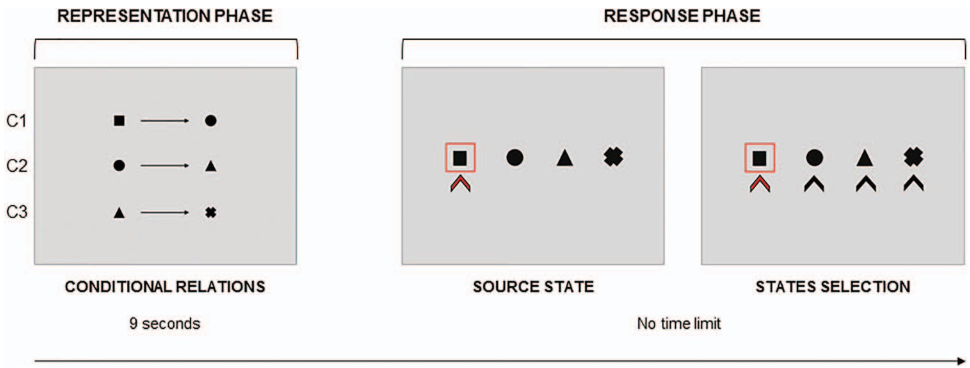

Figure 1. Relational-inference task. In the representation phase (left panel), participants observed for 9 seconds three conditional statements (C1, C2, C3) connecting abstract symbols (states). In the response phase (right panel), they had to select all the states that necessarily followed the occurrence of one of the symbols presented in the representation phase (source state, highlighted by a [red] square and a [red] selection mark). In the current example, participants should have chosen all three remaining symbols (circle, triangle, and cross) given "square" as source state. See the online article for the color version of this figure.

In particular, we used a 20-min timed version of the task, which has been shown to be an adequate predictor of the untimed APM score (Hamel & Schmittmann, 2006). Participants were paid according to the number of correct responses (20 cents for each correct response, maximum 7.20 euros).

Cognitive Reflection Test (CRT). Participants answered the three questions of the CRT without any time limit. The CRT score reflected the number of correct responses in the test.

N-back task (2-back and 3-back). Participants performed a computerized version of the 2-back and 3-back task. In each of these tasks, participants were presented with a series of individual letters appearing at the center of the screen (100 letters in total) and they had to decide whether the current letter matched the one observed two (in the 2-back task) or three (in the 3-back task) trials before. Each letter was presented for 1000 ms, followed by a blank screen for 1,000 ms. At each trial, participants indicated their choice by pressing a response key for match or pressing nothing for non-match. In both tasks, participants were paid according to their proportion of correct responses (min 1 euro, max 3 euros for each task).

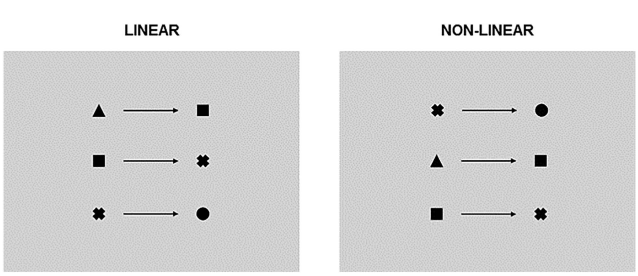

Figure 2. Types of symbol configuration in the relational-inference task. In linear trials (left panel), the spatial order of conditionals (from up to down) matched the underlying relational structure (triangle → square → cross → circle). In nonlinear trials (right panel), this was not the case: in fact, the up-down spatial order of conditionals did not match the current relational structure (triangle → square → cross → circle).

Forward and backward digit span. Participants were asked to repeat orally series of digits in the presented order (digit span forward) or in reversed order (digit span backward). They repeated increasingly long sequences of digits until they made two mistakes.

Participants and procedure. Participants were 50 students from the University of Trento, Italy (38 females, mean age 23.16, *SD* 2.80). The study was approved by the local ethics committee and all participants gave informed consent. Every participant took part in two experimental sessions on consecutive days. Participants performed the different experimental tasks in fixed order.

In the first experimental session, participants completed the relational-inference task while their eye movements were registered. After completing the relational-inference task, participants performed the Visual search control task.

In the second experimental session, participants completed the Raven Advanced Progressive Matrices Test (APM), the Cognitive Reflection Test (CRT), 2-back, and 3-back tasks and forward and backward digit span tests in fixed order. Feedback about performance and respective earnings in each task were provided at the end of the second experimental session.

Relational-inference task: Eye-tracking analysis.

Classification of transitions. To analyze eye movements, we defined six Regions of Interest (ROIs) in correspondence of the six symbols (see the Eye Movements Data Analysis section in Appendix A). We classified transitions as eye movements from one ROI to the next.

We classified a transition as Transitive Transition (henceforth, Transitive-T) if it was suitable for detecting a transitive relation within a relational set. More specifically, we focused on transitions connecting the consequent of a conditional relation to the antecedent of another conditional, because premise integration in transitive inference is generally achieved by the compression of the repeated term in a single token (Sternberg, 1980).

We also divided transitive-Ts in linear transitive-Ts and nonlinear transitive-Ts (see Figure 3).

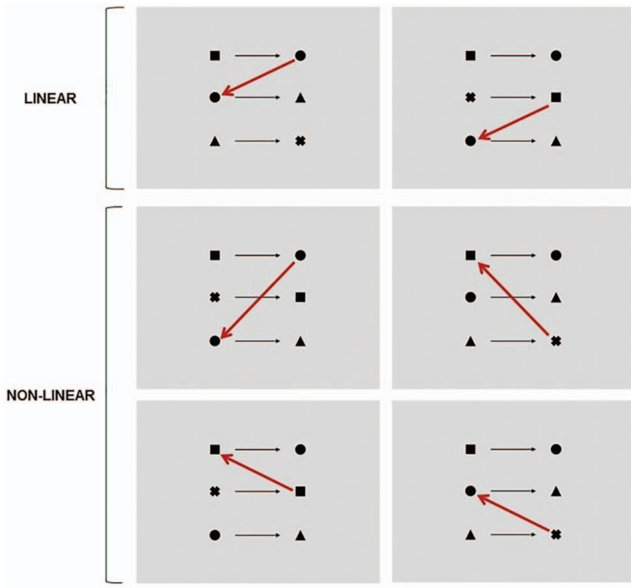

Figure 3. Depiction of the six possible transitive-Ts (arrows), grouped in linear and nonlinear transitive-Ts. See the online article for the color version of this figure.

Linear transitive-Ts were those transitions suitable for detecting transitive relations in linear relational sets (up-down transitive relations between adjacent conditionals). On the contrary, nonlinear transitive-Ts were coherent with an attempt to individuate transitive relations in nonlinear sets (any down-up transitive relations or transitive relations connecting nonadjacent conditionals).

Representation-building and representation-consolidation intervals. To individuate the type of representation process employed by each participant, we need to segregate processes purely related to the generation of representations from mechanisms associated with retention of information in working memory. In fact, within the representation phase, we expect (a) a first stage more oriented to information acquisition, meant to build a representation of the current relational structure and (b) a second stage more dedicated to the consolidation of information in working memory, in view of the response phase. These two stages should be marked by a peculiar difference in terms of cognitive load: the initial phase of information-search should require a lower memory load than the process of mnemonic consolidation of the final representation. Recent eye-tracking evidence highlighted a relation between computational load and fixation length: in particular, exploratory behavior is associated with short fixations, while higher-order processes are characterized by longer fixations (Graffeo, Polonio, & Bonini, 2015; Velichkovsky, 1999; Velichkovsky et al., 2002). Moreover, several studies on gaze data revealed that exploratory behavior emerges in an initial phase of the visual analysis, whereas integration of information intervenes in a later stage (Castelhano, Mack, & Henderson, 2009; Unema, Pannasch, Joos, & Velichkovsky, 2005). For these reasons, we expect the first stage to be characterized by shorter fixations compared with the second stage. Taking advantage of this property of gaze data, we performed several within-participant and within-trial cluster analyses using as variables of interest (a) the fixation length (ms) and (b) the time point of the fixation (ms).[2] Data sets included data-points from single trials in individual participants. We used a k-means cluster analysis using an algorithm based on L1 (Manhattan) distance to individuate two clusters in each dataset.[3] We performed 4,000 (50 participants $\times$ 80 trials) different cluster analyses on 4000 different data sets, individuating in each trial two clusters of fixation events: an early set of fixations that we associated with the representation building process and a later cluster of fixations related to representation consolidation mechanisms (see Figure 4). Henceforth, we will refer to these temporal phases as representation-building and representation-consolidation intervals. This method allowed us to individuate intervals based on actual eye data of single participant in single trials. This aspect is important because it allowed us (a) not to assume any arbitrary length of the two intervals, (b) to preserve between-subjects variability (differences in interval lengths across participants), and (c) to maintain within-subject heterogeneity (differences in interval lengths across trial categories).

Attentional indices. Once having isolated a time interval closely related to representation-building mechanisms (representation-building interval), we investigated whether we could detect distinct information-search patterns expressing sophisticated and unsophisticated representation processes. We expect sophisticated participants to explore the relational space to detect higher-order transitive relations between conditionals, whereas unsophisticated participants should not search for transitive relations and should encode binary rules without exploring the underlying higher-order relational complexity. We therefore individuated three attentional indices that could express whether agents searched for relational information in the representation-building interval.

These are the three indices of interest:

Relational search (RS). An agent who aims to search for all possible relations in the environment should perform a considerable number of transitions in a short time window. The relational search index expresses the tendency to perform a high rate of transitions in the representation-building interval.

We calculated individuals' relational search indices dividing, for each trial, the total number of transitions by the duration of the respective representation-building interval. Then we calculated the average of these trial-by-trial search indices to obtain a single individual measure of relational search. The greater the index magnitude, the higher the rate of transitions carried out by the respective participant in the representation-building interval.

Attentional bias (AB). Because the relational structure of sets can be spatially expressed in different ways (e.g., linear and nonlinear sets), searching for relations requires homogeneous distribution of attention in the entire relational space. Conversely, heterogeneous spread of attention might indicate a lack of purely exploratory behavior and suggest enhanced computation on the most-attended items, because agents tend to focus their attention

[2] We used *end of fixation* instead of *start of fixation* as temporal indicator of fixation occurrence because it facilitates the detection of the temporal switch from short to long fixations by the clustering algorithm.

[3] We chose an algorithm based on L1 distance because it has been shown to be more robust to the influence of outliers compared with higher-order distance metrics including Euclidean distance and Mahalanobis distance (Sidiropoulos & Bro, 1999; Zhong, Deng, & Jain, 2012) and to better deal with overpower of large-scale features (Loohach & Garg, 2012).

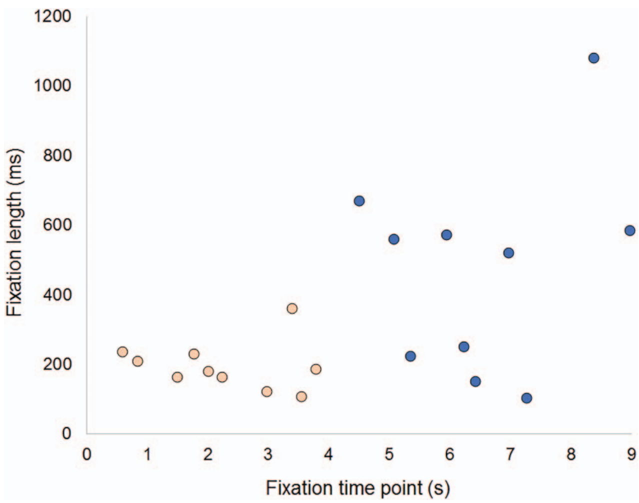

Figure 4. Example of cluster analysis on eye-tracking data from one trial of a single participant. Points represent fixations, performed in precise time points within the trial (*x* axis) and characterized by specific lengths (*y* axis). Colors of the points express the results of the cluster analysis: an early cluster of short fixations (orange [light gray] dots, representation–building interval) and a later set of longer fixations (blue [dark gray] dots, representation–consolidation interval). See the online article for the color version of this figure.

on the elements they are processing (Devetag, Di Guida, & Polonio, 2016; Polonio, Di Guida, & Coricelli, 2015; Rayner, 1998). The attentional bias index reflects the ability to distribute attention homogeneously across ROIs during the representation-building interval.

More specifically, the present index measures the magnitude of deviation from the perfect distribution of attention (1/6 of total fixation time for each of the six ROIs). The attentional bias index was generated by calculating, for each trial, the Euclidean distance from the expected homogeneous distribution of attention across the six ROIs to the actual distribution of fixations across the ROIs. We used the average of these trial-by-trial indices of attention to express individual indices of attentional bias across participants. The lower the index value, the lower the distance from perfectly homogenous distribution of attention.

Relational bias (RB). A participant who aims to build an exhaustive model of the relational environment should search for all the potential types of high-order relations in the current structure. In particular, agents should perform both types of transitive-T (linear and nonlinear) in the representation-building interval. The relational bias index expresses the ability to perform every type of transition to detect potential higher-order relations.

Because we individuated two types of linear transitive-Ts and four types of nonlinear transitive-Ts (see Figure 4), we calculated relational bias as the Euclidean distance between the actual ratio of nonlinear transitive-Ts (over transitive-Ts) and the expected proportion of nonlinear transitive-Ts (2/3 of the total number of transitive-Ts).[4] The lower the index value, the lower the distance from the expected distribution of linear and nonlinear transitive-Ts.

Hypotheses.

Expected gaze patterns in sophisticated and unsophisticated participants.

Sophisticated. We expect sophisticated participants to build a representation that explicitly expresses the relational structure of the relational set. They should therefore search for every possible relation characterizing a relational set, showing a high rate of transitions in their representation-building interval (high RS), exhibiting a homogenous distribution of attention across ROIs (low AB), and implementing both linear and nonlinear transitive-Ts (low RB).

Unsophisticated. In the representation-building interval, participants implementing an unsophisticated representation process should not search for higher-order relations linking conditional rules. We expect them to acquire and memorize triplets of conditionals in sequential order, without trying to manipulate and rearrange them in a model that resembles the actual relational structure of the trial. Such lack of pure exploratory behavior in favor of a tendency to memorize nonintegrated chunks of information should slow down acquisition of information, leading to a low rate of transitions in their representation-building phase (low RS). Moreover, because sequences of only two to four digits at a time can be memorized (Cowan, 2012), they should spend a significant proportion of their representation-building interval on a subset of the six elements (high AB). Finally, we predict them to perform few nonlinear transitive-Ts (high RB), because their strategy requires a simple up-down, left-right sequential and ordered scan path, as expected by Western cultural visual scan propensity (Abed, 1991; Chua, Boland, & Nisbett, 2005; Ishii, Okubo, Nicholls, & Imai, 2011).

Performance in the relational-inference task. In the relational-inference task, we expect sophisticated participants to show higher average accuracy rates than unsophisticated participants, because their comprehensive model of the relational environment should allow them to respond to the occurrence of every possible state.

The performance drop of unsophisticated participants should be particularly pronounced in nonlinear relational sets, because the mismatch between the latent relational structure and their internal representation should lead to a high error rate when applying transitive inference in the Response phase (Halford, 1984).

The role of working memory, fluid intelligence and cognitive reflection. After individuating two groups of participant expressing sophisticated and unsophisticated representation processes, we plan to compare measures of working memory, fluid intelligence, and cognitive reflection across groups. If these cognitive abilities are involved in the emergence of a specific type of representation process, we should observe differences between the two groups: in particular, we would expect higher levels of working memory, fluid intelligence, or cognitive reflection in the sophisticated group, in respect to the unsophisticated one. Moreover, it is possible that one or more of these cognitive abilities sustain processes of retention and updating of information independently of the representation process implemented. In this case, we should observe intragroup modulation of performance

[4] Number of transitive-Ts and of nonlinear transitive-Ts were computed pooling data from all trials, because single trial data in the representation-building stage contained few occurrences of transitive-Ts (especially nonlinear). Using trial-by-trial ratios, RB indices would have been noisier indicators of relational search behavior.

by these cognitive measures. This would indicate that these constructs sustain correct recall of information and efficient update of information in the response phase, when the source state is provided.

Results and Discussion

Representation-building and representation-consolidation intervals. To separate representation-building and representation-consolidation intervals, we ran 4,000 independent *k*-means cluster analyses on within-participant and within-trial fixation data using fixation length and time point of fixation as variables. On average, data sets included 22.5 data-points (fixations). The average time boundary between the two intervals was 4.37 seconds ($SD = 0.65$). Importantly, average fixation length in the representation-building interval was significantly lower than the one in the representation-consolidation interval (representation-building, $M = 261.77$ ms, $SD = 53.92$; representation-consolidation, $M = 308.87$ ms, $SD = 92.06$; Wilcoxon's matched-pairs signed-rank test, $z = -5.613$, effect size $(r) = 0.79$, $p < .001$), suggesting that they express two distinct phases of information processing.

Disclosing sophisticated and unsophisticated representation processes. To investigate the existence of two distinct representation processes, we conducted a between-subjects *k*-means cluster analysis on our three attentional indices. To estimate the optimal number of clusters in our dataset, we computed the gap statistics (Giancarlo, Scaturro, & Utro, 2008; Tibshirani, Walther, & Hastie, 2001). Results revealed that $k = 2$ was the best solution (Table A2 in Appendix A), suggesting that the heterogeneity in the three attentional indices was best explained by two types of behavior.

In Figure 5, we report results of the cluster analysis ($k = 2$, L1 as distance measures and RS, AB and RB as variables of interest). Cluster-1 ($N = 25$) was characterized by high RS, low AB, and low RB; conversely, cluster-2 ($N = 25$) showed low RS, high AB, and high RB, reflecting expected differences in the process of representation generation of sophisticated and unsophisticated agents. For this reason, we will refer to cluster-1 as the sophisticated group and to cluster-2 as the unsophisticated group. Examples of visual analyses of sophisticated and unsophisticated participants are shown in Figure 6.

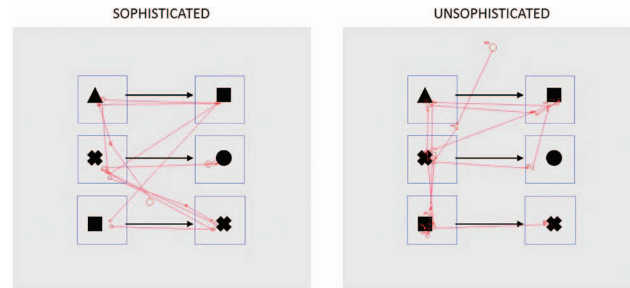

Figure 6. Examples of visual analysis of sophisticated and unsophisticated participants in the representation–building interval. The sophisticated participant (left panel) performed a high number of transitions (red arrows), distributed her fixations rather homogeneously across ROIs (red circles) and performed both linear and nonlinear transitive-Ts (as visible from the direction of arrows). The unsophisticated participant (right panel) exhibited a lower number of transitions, her attention was more focused on the top-left ROIs and did not perform any nonlinear transitive-Ts. See the online article for the color version of this figure.

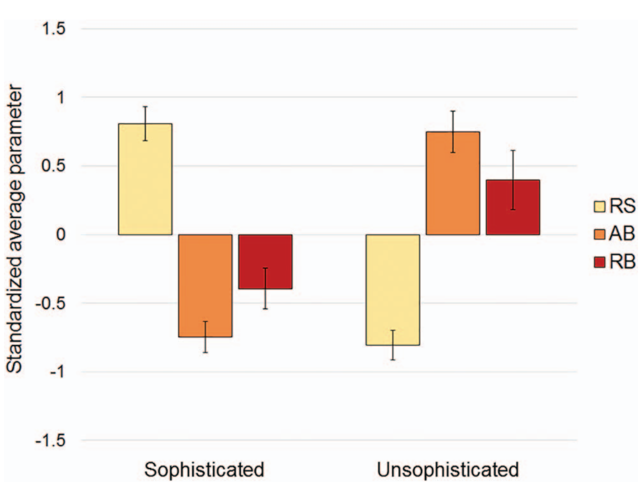

Figure 5. Bar graph of standardized indices of visual analysis in the two clusters of participants. See the online article for the color version of this figure.

A one-way multivariate ANOVA with relational search, attentional bias and relational bias as dependent variables and group as independent factor confirmed that the two groups express significantly different behaviors in terms of attentional indices, $F(3, 46) = 46.58$, $p < .001$. Individual index comparisons confirmed that all three indices were significantly different across groups (Mann-Whitney U test, RS: $z = 5.93$, $p < .001$; AB: $z = -5.52$, $p < .001$; RB: $z = 2.86$, $p = .004$. All *p* values were significantly at Bonferroni corrected threshold, $p = .017$). Interestingly, subject classification was remarkably stable along the time course of the experiment: we ran two different cluster analyses for the first and second halves of the experiment, and we found that 88% of our participants were consistent in terms of strategy.

A possible alternative explanation of the observed differences in representation strategy concerns visual processing speed: in line with this hypothesis, participants in the unsophisticated group would show the observed attentional index levels simply because of low efficiency in scanning the relational environment. We tested this hypothesis by comparing the two groups in the Visual search task: if visual scan efficiency drove the eye-movement differences in the relational-inference task, the sophisticated group would show higher performance in the visual search task. However, the two groups did not differ in any of the efficiency measures we collected (accuracy, RTs, earnings; see Table A8 in Appendix A). These results suggest that the intergroup differences observed in the relational-inference task were not related to general efficiency in visual scanning.

Then we investigated whether the lookup patterns of sophisticated and unsophisticated participants changed along the time course of the trial depending on relational set type. We considered the proportion of nonlinear transitive-Ts as measure of interest because its evolution in time should reflect the degree of understanding of the current relational structure. As shown in Figure 7, in nonlinear sets, sophisticated participants accumulated evidence about the existence of nonlinear transitive-Ts in the first part of the trial and, once they individuated them, maintained a stable ratio of

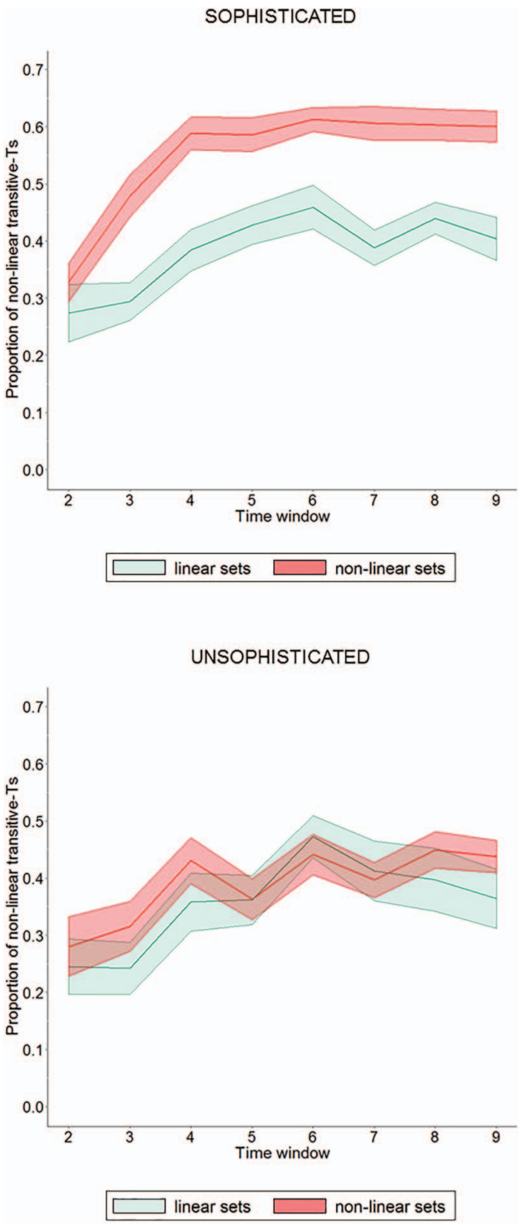

Figure 7. Time course of proportion of nonlinear transitive-Ts by trial category. We considered time windows of 1 second. The first time window (0–1 s) was discarded from the plot because of the extremely low number of transitive-Ts in this time interval (0.004% of the total number of transitive-Ts). Filled areas around lines represent standard error of the mean. See the online article for the color version of this figure.

nonlinear transitive-Ts to favor consolidation of these relations in working memory. In linear sets, sophisticated agents maintained a low proportion of nonlinear transitive-Ts, given their absence in this type of set. These results suggest that sophisticated agents were building a representation that explicitly expressed the relational structure of the environment. Conversely, we do not observe differences in the proportion of nonlinear transitive-Ts across relational sets in unsophisticated participants, suggesting that they did not grasp the relational structure of the current environment.

Performance in the relational-inference task. We ran a mixed-design ANOVA with mean accuracy as dependent variable, group (sophisticated and unsophisticated) and relational set (linear and nonlinear) as independent factors. Results show significant main effects of group, $F(1, 48) = 18.20$, $p < .001$ and category, $F(1, 48) = 27.09$, $p < .001$, and a significant interaction effect, $F(1, 48) = 17.62$, $p < .001$, indicating that the relation between performance in linear and nonlinear sets differed across groups. Figure 8 shows that sophisticated participants show higher average accuracy than unsophisticated ones, who in turn exhibited a significant decrease in performance in nonlinear relational sets. These results point out that sophisticated representation behavior is more effective than unsophisticated processing, especially when the relational structure underlying the current environment is implicit and not easily recognizable.

Cognitive abilities, representation processes, and performance. We tested whether sophisticated representation behavior was accompanied by higher abilities in cognitive reflection, fluid intelligence or working memory (see Table 1). Tests of the six directional hypotheses (higher score for sophisticated participants in each cognitive test) were conducted using Bonferroni adjusted alpha levels of .008 per test (.05/6). The sophisticated group indeed showed higher CRT score than the unsophisticated group (one-tailed Mann–Whitney U test, $z = 2.508$, effect size (r) = 0.35, $p =$ .006), suggesting that cognitive reflection had an impact on the emergence of distinct representation processes. On the other hand, APM score and measures of working memory did not differ across groups (one-tailed Mann–Whitney U test: APM, $z = 0.20$, $p =$ .419; Forward digit span, $z = 1.94$, $p = .026$; Backward digit span, $z = .253$, $p = 1.00$; 2-back, $z = -0.22$, $p = .412$; 3-back task, $z =$ 0.26, $p = .397$).

To corroborate these findings, we ran a stepwise backward logistic regression (Draper & Smith, 1998; Efroymson, 1960; Hocking, 1976) with group as dependent variable and all the six cognitive measures as independent variables. A low variance inflation factor (VIF, Marquaridt, 1970) of 1.38 indicated no collinearity between variables (see Table A3 in Appendix A for between-measure correlation table). Results confirmed that the

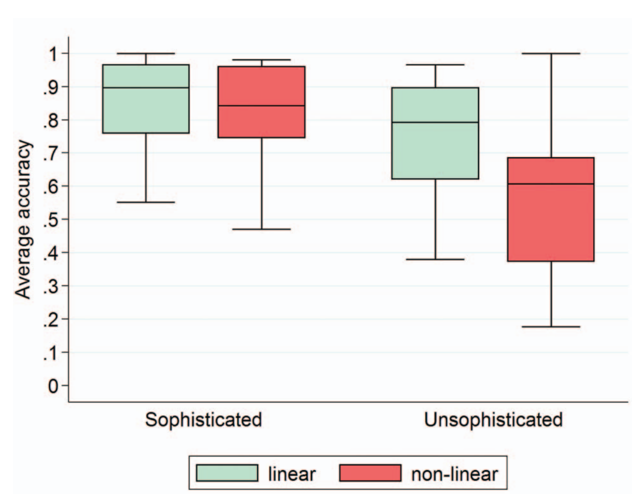

Figure 8. Boxplots of mean accuracy for the two groups in the two types of relational set. See the online article for the color version of this figure.

Table 1
Summary Statistics (Average and Standard Deviation, in Parentheses) of the Six Cognitive Tests Administered to Participants Divided by Group (Row 1 and 2) and Collapsed (Row 3) for Experiment 1

Group	Number of observations	CRT	APM	Forward digit span	Backward digit span	2-Back	3-Back
Sophisticated	25	1.84 (1.07)	21.24 (3.71)	6.64 (1.08)	5.4 (1.15)	.85 (.09)	.72 (.09)
Unsophisticated	25	1.04 (1.06)	20.88 (4.36)	6 (1.12)	5.24 (1.13)	.86 (.06)	.72 (.08)
Total	50	1.44 (1.13)	21.06 (4.01)	6.32 (1.13)	5.32 (1.13)	.85 (.08)	.72 (.09)

CRT score was the only cognitive measure significantly predicting the type of representation process used ($B = 0.78$, $p = .015$, see Table A4 in Appendix A). Furthermore, we tested whether one or more of our cognitive measures modulated within-group performance in the relational-inference task. We observed that performance in the unsophisticated group was significantly affected by the level of fluid intelligence and backward span score (stepwise backward regression, Table 2). In the sophisticated group, performance was modulated by APM score, and marginally by working memory measures such as backward span and 2-back score (see Table 2). These results suggest that fluid intelligence and working memory sustain the representation process by modulating mechanisms of retention and updating of stored information. It is not surprising that the effect of working memory is stronger in the unsophisticated group. In fact, the individuation of transitive relations in sophisticated participants could have allowed them to chunk information more efficiently in the representation phase, decreasing memory load in the response phase.

Causal mediation analysis. To understand the interplay between the type of representation process and cognitive measures in explaining task performance, we used causal mediation analysis to test whether representation behavior could serve as a mediator in explaining the effect of one or more of our cognitive measures on performance in the relational-inference task. To obtain a single and continuous measure of representation behavior that could serve as a mediator factor, we standardized and averaged our three attentional indices in a unique index (Representation Index).[5] Using the approach implemented in the "Mediation" R package (Imai, Keele, & Tingley, 2010), we first estimated a linear mediator model with Sophistication Index as dependent variable and our six cognitive measures as predictors. Only CRT score significantly predicted Sophistication Index ($B = 0.40$, $p < .001$, see Tables A5 in section A1, Appendix A). This finding is in line with the results previously reported (Table 1 in the main text, Tables A4 in section A1, Appendix A), indicating that cognitive reflection is the only measure differing across groups. The second step of the analysis consisted of estimating a linear outcome model with overall accuracy as dependent variable and sophistication index and the six cognitive measures as independent variables (Table A6, Appendix A). Sophistication index ($B = 0.56$, $p < .001$), APM score ($B = 0.29$, $p < .001$), and backward span ($B = 0.26$, $p = .015$) significantly predicted overall accuracy, whereas CRT score did not predict accuracy ($B = .11$, $p = .324$). However, running a linear regression dropping sophistication index as predictor, CRT score significantly predicted accuracy ($B = 0.33$, $p = .014$, Table A7, Appendix A), suggesting complete mediation of sophistication index on the relation between cognitive reflection and performance.

Table 2
Stepwise Backward Regression Analyses of Overall Accuracy for Sophisticated and Unsophisticated Groups

Overall accuracy	*B*	*SE*	*t*	*p*	95% CI
Sophisticated group					
APM	.32	.12	2.79	.011	[.08, .56]
Backward span	.19	.11	1.79	.087	[−.03, .41]
3-Back	.20	.10	1.94	.066	[−.01, .41]
Number of observations	25				
Unsophisticated group					
APM	.41	.13	3.15	.005	[.14, .69]
Backward span	.50	.14	3.50	.002	[.20, .80]
Number of observations	25				

Note. Only cognitive measures surviving the limit for inclusion in the model ($p < .1$) are reported. Variables excluded from the model (sophisticated group): CRT, $p = .29$; digit span forward, $p = .39$; 2-back, $p = .78$. Variables excluded from the model (unsophisticated group): CRT, $p = .29$; digit span forward, $p = .19$; 2-back, $p = .87$; 3-back, $p = .40$.

Finally, we tested the statistical significance of the indirect effect. Confidence intervals were calculated using the bias-corrected and accelerated bootstrap method (BCa; DiCiccio & Efron, 1996), a procedure specifically recommended in mediation analysis (Preacher & Hayes, 2008). As expected, the average causal mediation effect of Sophistication Index on the relation between CRT score and overall accuracy was statistically significant ($p = .02$, based on 10,000 bootstrap samples), accounting for an estimated 68% of the total effect between CRT score and overall accuracy (see Table 3).

In sum, causal mediation analysis revealed a remarkable effect of cognitive reflection on representation-building processing, which in turn highly predicted accuracy in the relational-inference task. The relationship between cognitive reflection and performance was largely due to this mediating effect.

Summary. In Experiment 1, we introduced a novel eye-tracking task to investigate the process of spontaneous generation of contingency representations. A cluster analysis on self-initiated patterns of information-search revealed the existence of two groups of participants that expressed different representation-building behaviors. Sophisticated participants searched for higher-order relational information to construct a comprehensive model of the environment that connected each state with every potential consequence of its occurrence. Conversely, unsophisticated participants encoded binary rules without searching for the higher-order

[5] We changed the sign of AB and RB indices to have a continuous index indicating sophisticated representation behavior for positive values and unsophisticated representation behavior for negative values.

Table 3
Results of Causal Mediation Analysis With Representation Index as a Mediator, CRT Score as Independent Variable, and Overall Accuracy as Dependent Variable

Effect	Estimated coefficient	95% CI	*p*
Average causal mediation effect (ACME)	.23	[.04, .38]	.02
Average direct effect (ADE)	.11	[−.13, .37]	.37
Total effect	.33	[−.02, .61]	.05
Proportion mediated	.68	[.38, 7.13]	.05

relational properties underlying them. The emergence of these two distinct processes of representation generation led to marked differences in task performance, especially in relational sets where the intrinsic structure of the relational environment was misaligned with the order of presentations of conditional rules.

Results from cognitive assessments analyses revealed that cognitive reflection is the only measure explaining the emergence of the two strategies. Conversely, fluid intelligence and working memory modulated intragroup performance levels but did not differ across groups. However, these results are only correlational and provide indirect evidence about the cognitive factors underlying sophisticated and unsophisticated behavior. For this reason, we ran Experiment 2 to investigate whether the emergence of the unsophisticated strategy was due to either the inability to implement the sophisticated strategy or to mechanisms of spontaneous strategy generation.

Experiment 2

To better understand the cognitive mechanisms underlying the emergence of sophisticated and unsophisticated representation processes, in Experiment 2 we ran two additional sessions of the relational-inference task with a new pool of 56 participants. In Session 1 (pretreatment), participants completed the task with the same modalities of Experiment 1. In Session 2 (posttreatment), participants received additional information about the existence of the two strategies and their respective average performance rates. Then they were asked to repeat the relational-inference task in the way they preferred. We compared pre- and posttreatment visual analyses to identify potential strategy switches that would indicate that unsophisticated representation behavior does not depend on cognitive ability or motivation, but rather on processes related to the spontaneous generation of sophisticated representation strategies.

Although we report Experiment 2 right after Experiment 1 for continuity in terms of research question, we acknowledge that Experiment 2 and Experiment 3 were run on the same participants and Experiment 2 was run *after* Experiment 3, to avoid any interference by the manipulation included in Experiment 2 on behavior in Experiment 3.

Method

Participants and procedure. Participants were 56 students from the University of Trento, Italy (43 females, mean age 24.16, *SD* 4.75). The study was approved by the local ethics committee and all participants gave informed consent. Every participant took part in two experimental sessions (pre- and posttreatment) on consecutive days, performing the experimental tasks in fixed order. In the pretreatment session, participants completed a shortened-version of the relational-inference task while their eye movements were registered.[6] They were reimbursed according to their proportion of correct responses (minimum 0, maximum 9 euros). Instructions and control questions were the same as in Experiment 1. At the end of Session 1, participants performed some of the cognitive tests we used in Experiment 1. In particular, we chose those tests that have been observed to impact on behavior in the relational-inference task: APM, CRT, and backward digit span. The modalities of administration of APM and backward digit span were identical to Experiment 1. Concerning the CRT, we used a recent multiitem version of the CRT (Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2016) composed of six new items. Multiitem CRTs have been recently recommended to overcome limitations coming from familiarity and range restrictions, by decreasing the probability of previous exposure to the CRT's items and floor or ceiling effects (see Bialek & Pennycook, 2018; Stieger & Reips, 2016; Toplak et al., 2014).

In the second experimental session (posttreatment), additional instructions were read to participants before repeating the relational-inference task. We informed participants about the existence of the two strategies observed in the task (sophisticated and unsophisticated) and explained in details each of them, independently of the strategy used by the participant in the pretreatment session (see online supplemental material for full transcription of the instructions administered to participants). Moreover, participants were informed about the average performance and respective gain of participants using either the sophisticated or the unsophisticated strategy, calculated on data of Experiment 1.[7] After the administration of additional information, participants were asked to perform (for the second time) the task in the way they preferred, even implementing a strategy different from the two we reported. For the second session, 51 new items were created to avoid any potential confound due to the repetition of items of Session 1. Each new item consisted in a perfect copy of the correspondent item of Session 1 in terms of relational structure of symbols, but the identity of symbols in each logical position was changed. As in Session 1, participants were paid based on their proportion of correct responses (minimum 0, maximum 9 euros).

Using this manipulation, we ensured that all the participants could have access to the sophisticated strategy in the posttreatment session. Moreover, informing them about the difference in average gain between the two strategies served as a motivation for switching strategy. We aimed to analyze differences in representation behavior across sessions, to explore whether unsophisticated par-

[6] The new version consisted of 51 trials instead of the original 80 trials. Most of the items of the shortened-version were taken from the original one, but some new items were created to maintain the same ratio between linear and nonlinear relational sets and balance the occurrence of the different symbols and source states. Participants were provided with two 1-minute breaks (one every 17 trials). All the other characteristics of the task remained unaltered.

[7] Gain magnitudes of Experiment 1 were recalibrated based on the minimum and maximum range of Experiment 2. Unsophisticated participants: 62% of correct responses, 5.58 euros on average. Sophisticated participants: 84% of correct responses, 7.56 euros on average.

ticipants were prone and able to implement the sophisticated strategy.

Eye-tracking analysis. In the pretreatment session, we replicated the analysis pipeline of Experiment 1. We first performed single-trial and single-subject cluster analyses on eye fixation data (fixation length and time point as dimensions) to distinguish between representation-building and representation-consolidation intervals. Then we considered data in the representation-building interval to isolate three attentional indices: relational search (RS), attentional bias (AB), and relational bias (RB). These three indices served as variables in a between-subjects cluster analysis to identify sophisticated and unsophisticated participants. We then compared the two groups to explore differences in performance in the relational-inference task and in cognitive assessments such as CRT, APM, and backward digit span.

In the posttreatment session, we recalculated our three attentional indices and performed a between-subjects cluster analysis on these new indices, to investigate a potential change in the proportion of agents implementing the sophisticated or the unsophisticated strategy.

Hypotheses. We believe that the emergence of sophisticated representation behavior in Experiment 1 is driven by preferential access to deliberative processes of acquisition, binding and representation of relational information (as suggested by the high average CRT score). Coherently, we do not believe unsophisticated participants to be *unable* to implement the sophisticated strategy, but rather to express a minor disposition toward spontaneously generating it. For this reason, after repeating the task and having received additional instructions about the existence of the sophisticated strategy, we expect a large proportion of the participants classified as "unsophisticated" in the pretreatment session to switch strategy in favor of a more sophisticated one in the posttreatment session.

Results and Discussion

Session 1: Pretreatment. In Session 1 we replicated results of Experiment 1. We separated representation-building and representation-consolidation intervals in the representation phase by running single-trial independent k-means cluster analyses using fixation length and time point of fixation as variables. The mean time boundary between the two intervals was 4.43 seconds ($SD = 0.65$). Fixation duration was shorter on average in the representation-building interval than in the representation-consolidation interval (representation-building, $M = 254.59$ ms, $SD = 59.22$; representation-consolidation, $M = 292.95$ ms, $SD = 64.43$; Wilcoxon's matched-pairs signed-rank test, $z = -5.490$, effect size (r) = 0.73, $p < .001$), confirming a difference in cognitive processing across the two intervals.

A cluster analysis of our three attentional indices (calculated using data from the representation-building interval) returned two groups showing the same patterns we had found in Experiment 1. Participants in cluster-1 ($N = 36$) showed high RS, low AB, and low RB, whereas cluster-2 ($N = 20$) exhibited low RS, high AB, and high RB, as expected by sophisticated and unsophisticated agents, respectively. We will refer to cluster-1 as the sophisticated group and to cluster-2 as the unsophisticated group. As expected, indices were significantly different across groups (Multivariate ANOVA, dependent variables: RS; AB; RB; independent factor: group. Effect of group: $F[3, 52] = 29.51$, $p < .001$).

A mixed-design ANOVA corroborated results of Experiment 1 in terms of relationships between group, relational set type, and performance: we found significant main effects of group, $F(1, 54) = 13.29$, $p < .001$, and relational set type, $F(1, 54) = 33.022$, $p < .001$, and interaction effect, $F(1, 54) = 15.28$, $p = .025$. Specifically, unsophisticated participants exhibited lower performance than sophisticated ones, especially in nonlinear trials (sophisticated, linear: $M = 0.84$; sophisticated, nonlinear: $M = 0.78$; unsophisticated, linear: $M = 0.68$; unsophisticated, nonlinear: $M = 0.55$). Then we tested between-groups differences in terms of cognitive reflection, fluid intelligence and working memory. Sophisticated participants showed a higher CRT score than the unsophisticated group (one-tailed Mann–Whitney U test, $z = 2.59$, effect size [r] = 0.35, $p = .005$, significant at Bonferroni corrected threshold $p = .017$ [.05/3]), confirming that cognitive reflection has an effect on sophisticated representation behavior. We found a between-groups effect of Backward digit span ($z = 2.08$, $p = .019$), but this effect did not survive Bonferroni correction (Bonferroni corrected threshold $p = .017$). APM score did not have any impact on the emergence of either sophisticated or unsophisticated behavior ($z = 1.25$, $p = .106$). The effect of cognitive reflection on representation strategy was corroborated by a stepwise backward logistic regression analysis with group as dependent variable and the three cognitive measures as independent variables, showing the CRT score was the only cognitive measure significantly predicting representation strategy (CRT, $B = 0.77$, $p = .012$. Variables excluded from the model: APM, $p = .531$; backward digit span, $p = .396$).

We also replicated results indicating that fluid intelligence and working memory modulate intragroup performance (Stepwise backward regression. Sophisticated, APM: $B = 0.45$, $p = .001$; backward digit span: $B = 0.32$, $p = .013$. unsophisticated, APM: $B = 0.55$, $p = .001$; backward digit span: $B = 0.32$, $p = .058$). Furthermore, representation strategy completely mediated the relationship between cognitive reflection and performance (Linear regression of average accuracy with CRT, APM, and backward digit span as predictors. CRT effect without representation strategy included in the model: $B = 0.24$, $p = .044$. CRT effect with representation strategy included in the model: $B = 0.17$, $p = .144$. See Tables B1 and B2 in Appendix B).

In sum, results of Session 1 of Experiment 2 replicated the ones of Experiment 1, highlighting the existence of two groups of participants differing in terms of representation behavior. The emergence of these behaviors led to higher performance in the sophisticated group and was predicted by cognitive reflection level. In contrast, fluid intelligence and working memory did not predict the representation strategy implemented, but rather modulated performance by sustaining information maintenance and manipulation mechanisms.

Session 2: Posttreatment. After additional instructions about the existence of sophisticated and unsophisticated strategies, participants performed a second instance of the relational-inference task. We performed the same analysis of Session 1 based on the new eye data, and we investigated how agents were classified after the manipulation. Interestingly, the new cluster analysis returned a large group of 49 (of 56) participants showing attentional index levels expressing sophisticated representation behavior. Only seven participants were classified as unsophisticated agents.

Comparing the classifications pre- and postmanipulation, we can observe that 35 participants were classified as sophisticated in both Session 1 and Session 2 (S-S group). Fourteen participants were classified as unsophisticated in Session 1 and as sophisticated in Session 2 (U-S group). Finally, six participants were classified as unsophisticated in both Session 1 and Session 2 (U-U group). Only one participant was classified as sophisticated in Session 1 and as unsophisticated in Session 2. We did not include this participant in subsequent analyses.

We are particularly interested in the U-S group, because it includes participants who shifted their strategy from unsophisticated to sophisticated in the posttreatment session. Comparing indices from these participants in Session 1 and Session 2, we can observe a significant difference in the direction of the sophisticated strategy for all three attentional indices (Wilcoxon's matched-pairs signed-ranks test, RS: $z = -2.98$, effect size (r) $= -0.80$, $p = .003$; AB: $z = 2.42$, effect size $= 0.65$, $p = .016$; RB: $z = 3.30$, effect size $= 0.88$, $p = .001$. All p values were significantly at the Bonferroni corrected threshold, $p = .017$). Moreover, the overall index shift was significantly higher in the U-S group than in the S-S group (one-way multivariate ANOVA with RS, AB, and RB as dependent variables and group (two levels: S-S and U-S groups) as an independent factor, $F(3, 46) = 5.93$, $p = .002$). We did not include the U-U group in any statistical analysis due to the low number of subjects ($n = 6$). Nevertheless, comparing descriptive statistics of the three attentional indices pre- and postmanipulation in this group, we can see that index levels are very similar across sessions, and maintain the typical profile of unsophisticated agents (relational search: $M[S1] = -1.76$, $M[S2] = -1.78$; attentional bias: $M[S1] = 1.07$, $M[S2] = 1.40$; relational bias: $M[S1] = 1.47$, $M[S2] = 0.90$).

These results confirm that a high percentage (70%) of unsophisticated participants switched toward the sophisticated representation strategy in the posttreatment session, suggesting that these agents are indeed capable of implementing it. Interestingly, the attentional shift in the U-S group predicted the increase in performance in the posttreatment session (linear regression with increase in accuracy as dependent variable and average index shift [average of (post − pre] differences of RS, AB, and RB indices] as the independent variable, $B = 0.71$, $p = .043$). However, despite the observed increase in performance, participants in the U-S group did not reach the average level of performance of the S-S group in the posttreatment session (U-S: $M = 0.78$; S-S: $M = 0.90$). This can be explained by the fact that participants in the S-S group had the possibility to repeat the task using and refining the same strategy, whereas U-S group implemented the sophisticated strategy for the first time in the posttreatment session. In line with this hypothesis, we can see that the average accuracy level of U-S participants in the posttreatment session (78%) was comparable with the one of S-S participants (80%) in the pretreatment session (see Table 4).

Summary. In Session 1 (pretreatment) we replicated results of Experiment 1 showing the existence of two distinct strategies in the process of generation of internal models of contingencies. Results of Session 2 (posttreatment) show that the majority of participants classified as unsophisticated in Session 1 shifted strategy toward the sophisticated one, suggesting that unsophisticated agents *can* implement the sophisticated strategy. This suggests emergence of unsophisticated behavior is not primarily related to cognitive capacity, but is linked to a preferential access to it. Furthermore, it indicates that the implementation of the unsophisticated strategy in Session 1 is not due to motivational aspects, at least for the majority of the agents in the unsophisticated group. If scarce motivation were the main driver of heterogeneity in Session 1, we would expect similar behavior in Session 2, given that incentives are identical in the two sessions.

Table 4
Average Performance by Group in Pre- and Posttreatment for Experiment 2

Group	*N*	Pretreatment	Posttreatment
S-S	35	.80 (.19)	.90 (.13)
U-S	14	.60 (.21)	.78 (.19)
U-U	6	.57 (.29)	.64 (.33)

Note. Standard deviations are shown in parentheses.

Experiment 3

In Experiment 3, we investigated whether sophisticated and unsophisticated strategies can be generalized to more ecological contexts, where verbal premises express the conditional occurrence of hypothetical events in real life scenarios (Verbal-inference task). Specifically, participants had to judge the validity of verbal arguments consisting in conditional sequences of hypothetical states (see, e.g., Byrne, 1989).

In contrast to the relational-inference task, in the verbal-inference task we did not impose any time constraint in the process of relation encoding. Moreover, participants were not required to rely on STM mechanisms to perform the task. Despite the remarkable differences between the two tasks, we wanted to test whether agents classified as sophisticated in the relational-inference task would express more sophisticated representation behavior when building the representation of real-life hypothetical states in the verbal-inference task. This would suggest the existence of general and context-independent strategies in the process of encoding and representation of contingencies.

Method

Verbal-inference task. Participants of Experiment 2 ($n = 56$) performed an additional verbal-inference task while their eye movements were monitored. The verbal-inference task was performed in a different experimental session preceding both Session 1 and 2 of Experiment 2. As in the previous experiments, participants were paid based on their proportion of correct responses (minimum 0, maximum 9 euros). The task consisted of 66 conditional sequences divided in three blocks. Each trial was composed of a sequence of two hypothetical conditional premises, followed by an assertion revealing the actual occurrence (or nonoccurrence) of one of the previous states and a conclusion to be evaluated as valid or not valid. The two conditional premises were connected by a shared proposition, whose characteristics could return either *transitive* or *nontransitive* sequences. In transitive sequences, the shared proposition contained two identical terms; in nontransitive

Table 5
Examples of Items in the Verbal Conditional Sequence Task for Experiment 3

Linear	Nonlinear
Transitive	
If she goes out for dinner, she will eat sushi.	If she eats sushi, she will be happy.
If she eats sushi, she will be happy.	If she goes out for dinner, she will eat sushi.
She went out for dinner.	She went out for dinner.
She will be happy.	She will be happy.
Nontransitive	
If she works, she will go home late.	If she doesn't go home late, she will go out.
If she doesn't go home late, she will go out.	If she works, she will go home late.
She worked.	She worked.
She will go out.	She will go out.

sequences, one of the terms of the shared proposition consisted in the negation of the other (see Table 5). As in the relational-inference task, in both transitive and nontransitive sequences, the presentation of the two statements could follow the temporal order of events (linear sets) or be misaligned with it (nonlinear sets).

Possible inferences consisted in modus ponens (MP), modus tollens (MT), affirmation of the consequent (AC), and denial of the antecedent (DA). In some transitive sequences, participants were required to perform two inferences of the same type to judge the validity of the argument (e.g., MP ∴ MP). Nonetheless, in some of the nontransitive sequences, participants had to make two different inferences to express the validity of the argument (MP&DA; MT&AC).[8] Therefore, sequences differed among each other along four dimensions: linearity (linear or nonlinear), transitivity (transitive or nontransitive), number of inferences to perform (one or two) and type of inference (MP, MT, AC, and DA).

Feedback about performance in the verbal-inference task was provided at the end of the entire experiment (Session 2 of Experiment 2).

Eye-tracking analysis. To investigate whether the sophisticated and the unsophisticated strategy would also emerge in the construction of internal models of real life hypothetical states, we explored visual patterns of information acquisition in different temporal intervals of the Verbal-inference task. First, we defined an interval in which participants encoded and integrated the conditional statements (i.e., constructing an internal model of the premises) before knowing anything about the actual occurrence of states (as in the representation phase of the relational-inference task). This temporal interval, which will be referred to as integration interval, reflected mechanisms of encoding and integration of the premises without including any inferential process dependent on the actual occurrence of states. To this aim, we defined six rectangular ROIs (647×167 pixels) around the six propositions of each argument (see Figure 9). In each trial, we defined as belonging to the Integration interval every fixation data falling in one of the premise ROIs (R1–R4) before participants looked at the assertion (R5) or the conclusion (R6).

Using data from this interval, we investigated whether sophisticated participants tended to focus more on the integration of the two conditional statements to form an exhaustive and explicit model of the relational structures underlying the premises. Specifically, we analyzed transitions connecting the two states of the shared proposition following the temporal order of events (i.e., independently of their spatial order). These transitions could indeed indicate an attempt at integrating the two conditional statements in a unitary and ordered model of the premises. We will refer to these transitions as Integrative transitions (henceforth, integrative-Ts). Integrative-Ts could be either linear or nonlinear, depending on the current type of relational set (linear or nonlinear). Because linear sets could not contain any nonlinear transitive relation and therefore participants did not need to perform nonlinear integrative-Ts, our eye-tracking analysis focused on nonlinear sets.

Then we considered as *judgment interval* every fixation following the first attendance of either the assertion or the conclusion, until the response was made. The judgment interval reflected the inferential processing sustaining the judgment of the validity of the argument given the information about the actual occurrence (or nonoccurrence) of one of the states and the conclusion to be evaluated. In this interval, we investigated allocation of attention and cognitive resources to specific propositional elements in sophisticated and unsophisticated agents. In particular, we focused on those hypothetical states of the premises whose relationship had to be judged: the state (of the premises) whose occurrence has been revealed in the assertion and the state (of the premises) corresponding to the conclusion to be evaluated as valid or invalid. In the judgment interval, we will refer to these two ROIs as judgment states (see Figure 10).

We believe judgment states to be the key pieces of information in the reasoning process in the judgment interval, because the validity of the argument had to be derived from the evaluation of the hypothesized relationship between the judgment states. Therefore, in the judgment interval, we extracted attentional patterns that could indicate deeper information processing on these propositional elements. Specifically, we tested (a) distribution of attention between the judgment states and the other ROIs and (b) differences in depth of information processing between the judgment states and the other ROIs. The former parameter has been operationalized by calculating the proportion of time spent in the judgment

[8] MP&DA and MT&AC trials were treated independently of the order of the two inferences. Therefore, in MP&DA trials both MP ∴ DA and DA ∴ MP are included, whereas MT&AC trials consist in either MT ∴ AC or AC ∴ MT sequences. We also included some fillers with obvious solutions to balance valid and invalid responses in participants. Fillers were solved with very high accuracy (97%) and were not included in subsequent analyses.

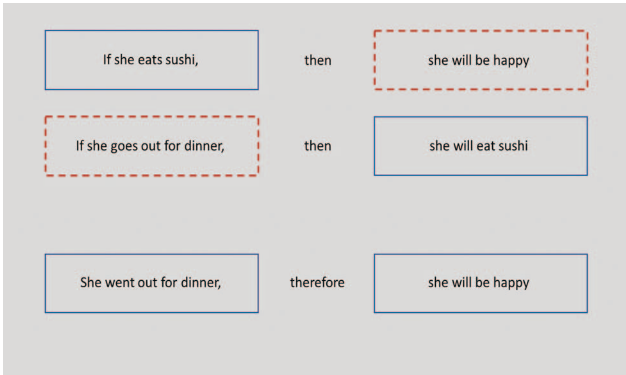

Figure 9. Example of trial with ROIs (square boxes around propositions, not shown to the participants) used for eye-tracking analysis. R1-R4 constitute the premises of the argument, and fixation falling in these ROIs before any fixation occur in R5–R6 are included in the Integration interval. In this example, dotted (red) ROIs represent the shared proposition. See the online article for the color version of this figure.

states compared with the other four ROIs in the judgment interval. The latter index has been calculated as the increase in fixation duration (increase in allocation of cognitive resources, see Graffeo et al., 2015; Velichkovsky, 1999; Velichkovsky et al., 2002) in the judgment states in respect to the other four ROIs.

Hypotheses. We expect participants classified as sophisticated in the relational-inference task (Experiment 2, pretreatment session) to devote greater attention to the generation of an exhaustive and explicit representation of the hypothetical chain of events when compared with unsophisticated participants. In the integration interval, before obtaining any information about the occurrence of states, we expect them to show a higher rate of nonlinear integrative-Ts in nonlinear relational sets than unsophisticated participants, who in turn should move to the assertion with a less comprehensive representation of the relationship underlying the premises.

In the judgment interval, sophisticated participants should allocate more cognitive resources to the states whose relationship has to be evaluated (i.e., judgment states), and devote less attention to other contextual pieces of information, because they should have already built an explicit representation of the underlying relational structure. This would translate into a higher proportion of time spent on the judgment states, as well as an increase in fixation duration in these two ROIs. On the contrary, we believe unsophisticated participants' relational representation not to explicitly express the relationship between judgment states. Therefore, once they have encoded the information expressed by the assertion, they should sequentially attend all the pieces of information in the set to concatenate conditional and transitive inferences. Consequently, we predict unsophisticated participants to allocate resources more homogeneously between judgment states and other ROIs in respect to sophisticated ones.

We also predict "sophisticated" attentional indices to modulate the ability to judge the validity of conditional arguments, because they reflect a deeper understanding of the underlying relational structure.

Results and Discussion

Behavioral results. First, we tested whether linearity (linear or nonlinear), transitivity (transitive or nontransitive), and number of inferences (1 or 2) modulated performance in the verbal-inference task. A mixed-effect logistic model (subject as random effect on all regressors) did not show any effect of linearity ($B = 0.01$, $p = .939$), transitivity ($B = -0.12$, $p = .102$), or number of inferences ($B = 0.04$, $p = .567$). Given these results, we will treat performance only in terms of type of inference. Table 6 reports average performance for each type of inference (MP, AC, DA, MT, MP&DA, MT&AC).

Representation behavior in the verbal-inference task. In the integration interval, we tested whether sophisticated agents (as classified in the relational-inference task) exhibited a higher tendency to integrate premises in a unitary model of the relational environment. We indeed observed that sophisticated agents showed a higher ratio of nonlinear integrative-Ts in the integration interval of nonlinear sets when compared with unsophisticated

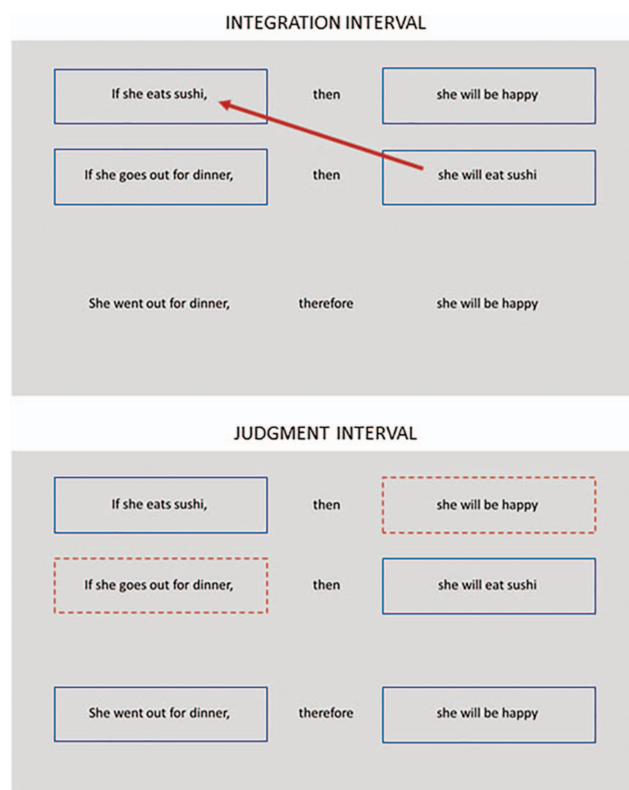

Figure 10. Eye-tracking analysis in the integration interval and in the judgment interval. In the Integration interval (upper panel), before participants have acquired information about the occurrence of states and inference to evaluate, we focused on nonlinear integrative-Ts (red [dark gray] arrow) in nonlinear relational sets, reflecting the attempt to individuate a nonlinear transitive relation between the states of the premises (square ROIs). In the judgment interval (lower panel), once participants have looked at the assertion or the conclusion, we focused on distribution of attention and depth of information processing in the judgment states (dotted [red] ROIs) in comparison to the other four ROIs (solid [blue] ROIs). See the online article for the color version of this figure.

Table 6
Average Accuracy by Type of Inference

MP	AC	DA	MT	MP&DA	MT&AC
.97	.38	.37	.70	.45	.45

Note. MP, AC, DA, and MT inferences include transitive (1 or 2 inferences) and nontransitive (1 inference) sequences. MP&DA and MT&AC consist of only nontransitive sequences (2 inferences). All six categories include linear and nonlinear sets.

ones (one-tailed Mann–Whitney *U* test: $z = 1.76$, effect size [r] = 0.79, $r = .24$, $p = .039$), suggesting that they focused more on the integration of conditional statements in a relationally explicit model *before* moving to the assertion. To describe this effect, in Figure 11 we plotted the temporal evolution of the proportion of nonlinear integrative-Ts in the Integration interval of nonlinear trials for sophisticated and unsophisticated participants. Sophisticated and unsophisticated agents showed similar proportions of nonlinear integrative-Ts in the first seconds of information accumulation, due to an initial reading of the premises. However, after a few seconds of accumulation of evidence about the relational structure of the environment, sophisticated agents significantly increased their rate of nonlinear integrative-Ts. In sum, sophisticated participants detected the nonlinearity in the relational structure and focused on the integration of the two conditional statements to build a comprehensive model of the hypothetical scenario.

Afterwards, we compared attentional indices in the judgment interval across groups. Results show that sophisticated agents spent more time on the judgment states (one-tailed Mann–Whitney *U* test: $z = 1.91$, effect size [r] = 0.26, $p = .027$) and showed a higher increase in fixation duration in the Judgments states (one-tailed Mann–Whitney *U* test: $z = 2.05$, effect size [r] = 0.27, $p = .021$) than unsophisticated ones. Interestingly, the attentional index in the Integration interval predicted the level of indices in the judgment interval (see Table 7), suggesting that the tendency to integrate premises in a unitary relational model during integration was associated with an enhanced attentional focus on key pieces of information during the validity judgment.

Representation behavior and performance in the verbal-inference task. Although the proportion of nonlinear integrative-Ts in the Integration interval predicted the level of the attentional indices in the judgment interval, it did not have a direct impact on performance (Table C1, Appendix C). We therefore tested the hypothesis that patterns of information acquisition in the judgment interval predicted performance in the task. Because proportion of time spent on the judgment states and increase in fixation duration in these ROIs were highly correlated (Spearman's rank correlation, $r = .64$, $p < .001$), we ran a stepwise backward regression with the two indices as independent variables and mean overall accuracy in the verbal-inference task as the dependent variable to select the best predictor among the two. Results show that increase in fixation duration was excluded from the model ($p = .343$), whereas the proportion of time spent in the judgment states had an impact on performance ($B = 0.43$, $p = .001$). Therefore, we used the latter variable as an indicator of behavior in the judgment interval to explore its effect on performance. We ran a multivariate regression with the six inference categories as dependent variables and proportion of time spent in the judgment states as the independent variable, and we found that the attentional index predicted higher performance in AC, DA, MT&AC, and MP&DA inference categories (see Table 8).

Finally, we tested whether cognitive measures such as cognitive reflection, working memory and fluid intelligence modulated performance in the verbal-inference task. We ran a stepwise backward regression with mean overall accuracy in the verbal-inference task as the dependent variable and APM score, CRT score and backward digit span as independent factors. Results indicated that working memory, as reflected by the backward span, predicted performance in the task ($B = 0.38$, $p = .005$, see also Table C2, Appendix C, for individual inference type analysis), whereas cognitive reflection and fluid intelligence levels were unrelated to mean accuracy (APM, $p = .86$; CRT, $p = .16$). This result is consistent with several studies showing correlations between working memory capacity and reasoning, for instance in the evaluation of syllogistic arguments (Capon, Handley, & Dennis, 2003; Copeland & Radvansky, 2004; Gilhooly, Logie, Wetherick, & Wynn, 1993, Gilhooly, Logie, & Wynn, 1999; Kyllonen & Christal, 1990). Nonetheless, the association between working memory abilities and validity judgments in syllogistic arguments is in line with several theories of syllogistic reasoning (Fisher, 1981; Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Sternberg & Turner, 1981).

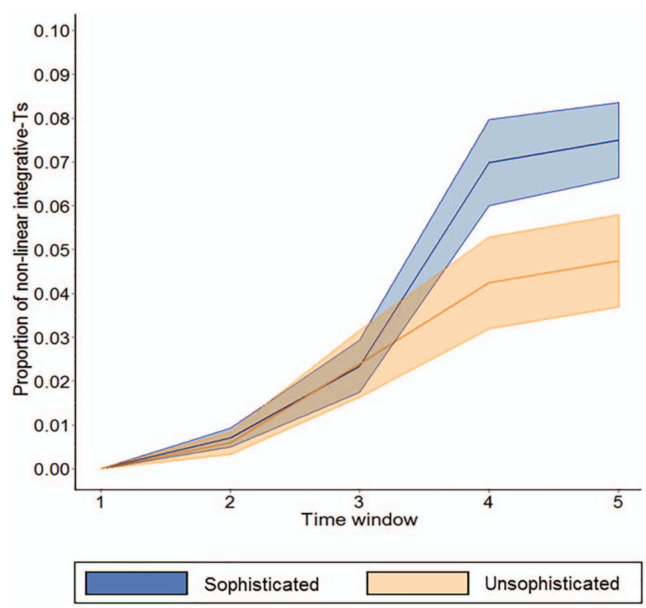

Figure 11. Time course of proportion of nonlinear transitive-Ts (over the total number of between-ROI transitions) by group in the Integration interval of nonlinear trials. Fixation distribution was normalized across trial time by assigning fixations to five homogeneous intervals based on total number of fixations in the Integration interval of that specific trial. In this way, each trial was characterized by five temporal intervals containing equivalent numbers of fixations. Trial-by-trial proportions of transitions were averaged for each participant and then individual time courses were averaged across participants. Filled areas around lines represent between-subjects standard error of the mean. See the online article for the color version of this figure.

Table 7
Multivariate Regression With Attentional Indices in the Judgment Interval as Dependent Variables and Proportion of Nonlinear Integrative-Ts as Independent Variable

Attentional indices: Judgment interval	*B*	*SE*	*t*	*p*	95% CI
Prop. time on judgment states					
Proportion of nonlinear integrative-T	.38	.13	3.02	.004	[.13, .63]
Increase fix. duration in judgment states					
Proportion of nonlinear integrative-Ts	.30	.13	2.29	.026	[.04, .56]
Number of observations	56				

Summary. Results of Experiment 3 indicate that heterogeneity of performance in the verbal-inference task is linked to the amount of cognitive resources allocated to the judgment states in the judgment interval, which is in turn predicted by the tendency to integrate premises in a unitary and explicit representation of the hypothetical scenario in the Integration interval. All these indices are associated with the emergence of either sophisticated or unsophisticated behavior in the relational-inference task, suggesting the existence of general, context-independent strategies in building relational representations of contingencies.

General Discussion

In three eye-tracking experiments, we investigated individual differences in the generation of internal representations of interrelated contingencies. In Experiment 1 and 2 we introduced a novel relational-inference task with symbolic content, whereas in Experiment 3 participants had to judge the validity of arguments in verbal conditional sequences expressing real life hypothetical situations. Taken together, results of the three experiments revealed the existence of two strategies for building relational models of contingencies. Sophisticated participants spontaneously tended to construct a sequential ordered model of interrelated events, generating a mental representation that explicitly expressed the relational structure of the environment. Conversely, unsophisticated agents encoded binary conditional relations among states without grasping the underlying relational complexity.

Several insights from the three experiments unravel the cognitive nature of this heterogeneity. Results from analyses of cognitive measures across our two groups in the relational-inference task suggest that cognitive abilities such as fluid intelligence and working memory do not have a crucial role in the process of representation strategy generation. These results are in line with recent studies investigating the emergence of different strategies in categorical learning (Goldwater, Don, Krusche, & Livesey, 2018; Little & McDaniel, 2015). These studies underlined the existence of agents either memorizing simple feature-based rules or encoding higher order relations between elements. In both studies, fluid intelligence did not predict learning strategy, even though it modulated learning rates. Moreover, evidence about the existence of a relationship between learning strategy generation and working memory capacity is inconsistent across studies (see Little & McDaniel, 2015; McDaniel et al., 2014).

Importantly, unlike fluid intelligence and working memory, cognitive reflection robustly predicted the type of representation process applied. Cognitive reflection traditionally expresses the tendency to implement either deliberative or reflexive processes (Frederick, 2005; Travers, Rolison, & Feeney, 2016). Moreover, it has been recently associated with accuracy in processes of information search (Cokely et al., 2009; Cokely & Kelley, 2009) and representation of task-relevant information (Mata et al., 2014; Sirota et al., 2014). In line with these findings, high cognitive reflection levels may reflect a preferential access to more deliberative representation processes (Osman, 2004), which leads to the generation of more sophisticated strategies in task resolution. Moreover, it may suggest that the emergence of either sophisticated or unsophisticated behavior in representation-building pro-

Table 8
Multivariate Regression With Accuracy in Each Type of Inference as Dependent Variables and Proportion of Time Spent on the Judgment States as Independent Variable

Mean accuracy	*B*	*SE*	*t*	*p*	95% CI
MP					
Prop. time on judgment states	.20	.13	1.48	.144	[−.07, .47]
AC					
Prop. time on judgment states	.44	.12	3.61	.001	[.20, .69]
DA					
Prop. time on judgment states	.35	.13	2.72	.009	[.09, .60]
MT					
Prop. time on judgment states	−.25	.13	−1.89	.064	[−.51, .15]
MP and DA					
Prop. time on judgment states	.40	.12	3.18	.002	[.15, .65]
MT and AC					
Prop. time on judgment states	.34	.13	2.62	.011	[.08, .59]
Number of observations	56				

cesses represents a more malleable thinking disposition, rather than an unmodifiable cognitive ability (for discussions on these issues, see Campitelli & Labollita, 2010; Toplak & Stanovich, 2002).

This interpretation is supported by the results of Experiment 2, which show that the majority of participants classified as unsophisticated in the pretreatment session switched toward sophisticated behavior in a repetition of the task (posttreatment session), after having received additional information about the existence of sophisticated and unsophisticated strategies and their respective efficacy rates in the task. These findings confirmed that most of our participants were cognitively able to build ordered representations of sequential events, but only reflective agents had a spontaneous and direct access to sophisticated representation processing when receiving relational information about conditional occurrence of hypothetical states. However, feedback, additional instructions, or simple practice can trigger analytical and deliberative processing that overcomes initial intuitive strategies (Ball, 2013), in line with two-stage reasoning process theories (e.g., Evans, 1984, 2006).

Nevertheless, Experiment 3 revealed that heterogeneity in representation behavior emerges spontaneously when agents reason about real life conditional sequences of events (verbal-inference task). In particular, participants classified as sophisticated in the relational-inference task (Experiment 2, pretreatment) showed a higher tendency to integrate between-state relations in an exhaustive model of contingencies before searching for information about the actual occurrence events in the verbal-inference task. On the contrary, unsophisticated agents were more prone to encode minimal units of relational information and start the inferential process without having built a model explicitly expressing direct and indirect consequences of states. This result is extremely important because, in the verbal-inference task, the encoding of hypothetical states was not constrained by time or STM limitations, indicating the existence of a spontaneous tendency to integrate relational information about contingencies in a coherent and exhaustive model of the relational space. This tendency also predicted behavior during the validity judgment, once information about the occurrence of a state and the conclusion to be evaluated had been attended. Specifically, participants who had already integrated premises in a comprehensive model (i.e., sophisticated participants) selectively allocated cognitive resources on the hypothetical states whose relationship had to be evaluated (assertion and conclusion). This is consistent with reasoning with mental models (Johnson-Laird, 1983; Johnson-Laird, 2010), which predicts the generation of counterexamples to the hypothesized relationship between the states whose relationship has to be evaluated as valid or invalid. On the contrary, unsophisticated participants allocated resources more homogenously across ROIs after attending the assertion and the conclusion, suggesting that they had a less comprehensive representation of the underlying relational structure when starting inferential processing. This difference in resources allocation explained part of the heterogeneity in performance in the verbal-inference task, showing preliminary evidence about the role of attention and representation processes in reasoning with conditional sequences.

We believe that the results of this study provide novel insights about the way agents encode and represent relational information about contingencies. Because these processes are crucial in several areas of investigations, including learning, decision-making, and reasoning, we hope that our results would fuel further research into the role of representation-building functions in explaining the heterogeneity underlying higher cognition.

References

Abed, F. (1991). Cultural influences on visual scanning patterns. *Journal of Cross-Cultural Psychology, 22,* 525–534. http://dx.doi.org/10.1177/0022022191224006

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49,* 5–28. http://dx.doi.org/10.1080/713755608

Ball, L. J. (2013). Microgenetic evidence for the beneficial effects of feedback and practice on belief bias. *Journal of Cognitive Psychology, 25,* 183–191. http://dx.doi.org/10.1080/20445911.2013.765856

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11,* 280–289. http://dx.doi.org/10.1016/j.tics.2007.05.005

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory & Cognition, 4,* 265–284. http://dx.doi.org/10.1016/j.jarmac.2014.09.003

Bialek, M., & Pennycook, G. (2018). The Cognitive Reflection Test is robust to multiple exposures. *Behavior Research Methods, 50,* 1953–1959.

Boote, S. K., & Boote, D. N. (2018). ABC problem in elementary mathematics education: Arithmetic before comprehension. *Journal of Mathematics Teacher Education, 21,* 99–122.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16,* 106–113. http://dx.doi.org/10.1016/j.tics.2011.12.010

Byrne, R. M. (1989). Everyday reasoning with conditional sequences. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 41,* 141–166. http://dx.doi.org/10.1080/14640748908402357

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making, 5,* 182–191.

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning, 9,* 203–244. http://dx.doi.org/10.1080/13546781343000222

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3), 6.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22. http://dx.doi.org/10.1037/h0046743

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 12629–12633. http://dx.doi.org/10.1073/pnas.0506162102

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making, 4,* 20.

Cokely, E. T., Parpart, P., & Schooler, L. J. (2009). On the link between cognitive control and heuristic processes. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive Science society* (pp. 2926–2931). Austin, TX: Cognitive Science Society. https://escholarship.org/uc/item/6qj0k72h

Copeland, D., & Radvansky, G. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57,* 1437–1457. http://dx.doi.org/10.1080/02724980343000846

Cowan, N. (2012). *Working memory capacity*. New York, NY: Psychology Press. http://dx.doi.org/10.4324/9780203342398

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69,* 1204–1215. http://dx.doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8,* 1704–1711. http://dx.doi.org/10.1038/nn1560

Devetag, G., Di Guida, S., & Polonio, L. (2016). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics, 19,* 177–201. http://dx.doi.org/10.1007/s10683-015-9432-5

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11,* 189–228. http://dx.doi.org/10.1214/ss/1032280214

Don, H. J., Goldwater, M. B., Otto, A. R., & Livesey, E. J. (2016). Rule abstraction, model-based choice, and cognitive reflection. *Psychonomic Bulletin & Review, 23,* 1615–1623. http://dx.doi.org/10.3758/s13423-016-1012-y

Doumas, L. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115,* 1–43. http://dx.doi.org/10.1037/0033-295X.115.1.1

Draper, N., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York, NY: Wiley. http://dx.doi.org/10.1002/9781118625590

Efroymson, M. A. (1960). Multiple regression analysis. In A. Ralston & H. S. Wilf (Eds.), *Mathematical methods for digital computers* (pp. 191–203). New York, NY: Wiley.

Evans, J. S. B. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75,* 451–468. http://dx.doi.org/10.1111/j.2044-8295.1984.tb01915.x

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13,* 378–395. http://dx.doi.org/10.3758/BF03193858

Fisher, D. L. (1981). A three-factor model of syllogistic reasoning: The study of isolable stages. *Memory & Cognition, 9,* 496–514. http://dx.doi.org/10.3758/BF03202344

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19,* 25–42. http://dx.doi.org/10.1257/089533005775196732

Giancarlo, R., Scaturro, D., & Utro, F. (2008). Computational cluster validation for microarray data analysis: Experimental assessment of Clest, Consensus Clustering, Figure of Merit, Gap Statistics and Model Explorer. *BMC Bioinformatics, 9,* 462. http://dx.doi.org/10.1186/1471-2105-9-462

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science, 317,* 1351–1354. http://dx.doi.org/10.1126/science.1144161

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 21,* 115–124. http://dx.doi.org/10.3758/BF03211170

Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology, 11,* 473–498. http://dx.doi.org/10.1080/095414499382264

Goldwater, M. B., Don, H. J., Krusche, M. J. F., & Livesey, E. J. (2018). Relational discovery in category learning. *Journal of Experimental Psychology: General, 147,* 1–35. http://dx.doi.org/10.1037/xge0000387

Graffeo, M., Polonio, L., & Bonini, N. (2015). Individual differences in competent consumer choice: The role of cognitive reflection and numeracy skills. *Frontiers in Psychology, 6,* 844. http://dx.doi.org/10.3389/fpsyg.2015.00844

Halford, G. S. (1984). Can young children integrate premises in transitivity and serial order tasks? *Cognitive Psychology, 16,* 65–93. http://dx.doi.org/10.1016/0010-0285(84)90004-5

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences, 14,* 497–505. http://dx.doi.org/10.1016/j.tics.2010.08.005

Hamel, R., & Schmittmann, V. D. (2006). The 20-minute version as a predictor of the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement, 66,* 1039–1046. http://dx.doi.org/10.1177/0013164406288169

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics, 32,* 1–49. http://dx.doi.org/10.2307/2529336

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters, 110,* 97–100. http://dx.doi.org/10.1016/j.econlet.2010.11.015

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15,* 309–334. http://dx.doi.org/10.1037/a0020761

Ishii, Y., Okubo, M., Nicholls, M. E. R., & Imai, H. (2011). Lateral biases and reading direction: A dissociation between aesthetic preference and line bisection. *Brain and Cognition, 75,* 242–247. http://dx.doi.org/10.1016/j.bandc.2010.12.005

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, UK: Cambridge University Press.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 18243–18250. http://dx.doi.org/10.1073/pnas.1012933107

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 615–622. http://dx.doi.org/10.1037/0278-7393.33.3.615

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352–358. http://dx.doi.org/10.1037/h0043688

Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory & Cognition, 38,* 667–676. http://dx.doi.org/10.3758/MC.38.6.667

Koenigs, M., Barbey, A. K., Postle, B. R., & Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *The Journal of Neuroscience, 29,* 14980–14986. http://dx.doi.org/10.1523/JNEUROSCI.3706-09.2009

Konovalov, A., & Krajbich, I. (2016). Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. *Nature Communications, 7,* 12438. http://dx.doi.org/10.1038/ncomms12438

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389–433. http://dx.doi.org/10.1016/S0160-2896(05)80012-1

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making, 25,* 361–381. http://dx.doi.org/10.1002/bdm.752

Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory & Cognition, 43,* 283–297. http://dx.doi.org/10.3758/s13421-014-0475-1

Loohach, R., & Garg, K. (2012). Effect of distance functions on simple k-means clustering problem. *International Journal of Computers and Applications, 49,* 7–9. http://dx.doi.org/10.5120/7629-0698

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods, 128*(1–2), 85–93.

Marquaridt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics, 12,* 591–612. http://dx.doi.org/10.1080/00401706.1970.10488699

Mata, A., Schubert, A. L., & Ferreira, M. B. (2014). The role of language comprehension in reasoning: How "good-enough" representations induce biases. *Cognition, 133,* 457–463. http://dx.doi.org/10.1016/j.cognition.2014.07.011

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematics problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29–53). Mahwah, NJ: Erlbaum.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General, 143,* 668–693. http://dx.doi.org/10.1037/a0032963

Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: Standardization and normative data from an Italian adult population. *Neurological Sciences, 34,* 749–754. http://dx.doi.org/10.1007/s10072-012-1130-x

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 45–100). Amsterdam, the Netherlands: Elsevier Academic Press. http://dx.doi.org/10.1016/S0079-7421(09)51002-X

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization, 72,* 147–152. http://dx.doi.org/10.1016/j.jebo.2009.04.018

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11,* 988–1010. http://dx.doi.org/10.3758/BF03196730

Polonio, L., Di Guida, S., & Coricelli, G. (2015). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior, 94,* 80–96. http://dx.doi.org/10.1016/j.geb.2015.09.003

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879–891. http://dx.doi.org/10.3758/BRM.40.3.879

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making, 29,* 453–469. http://dx.doi.org/10.1002/bdm.1883

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual section 4: Advanced progressive matrices*. Oxford, UK: Oxford Psychologists Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.372

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599. http://dx.doi.org/10.1126/science.275.5306.1593

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11,* 771–799. http://dx.doi.org/10.1177/1745691616650647

Sidiropoulos, N. D., & Bro, R. (1999). Mathematical Programming Algorithms for Regression-based Non-linear Filtering in RN. *IEEE Transactions on Signal Processing, 47,* 771–782. http://dx.doi.org/10.1109/78.747782

Sirota, M., Juanchich, M., & Hagmayer, Y. (2014). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review, 21,* 198–204. http://dx.doi.org/10.3758/s13423-013-0464-6

Sternberg, R. J. (1980). Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General, 109,* 119–159. http://dx.doi.org/10.1037/0096-3445.109.2.119

Sternberg, R. J., & Turner, M. E. (1981). Components of syllogistic reasoning. *Acta Psychologica, 47,* 245–265. http://dx.doi.org/10.1016/0001-6918(81)90011-1

Stieger, S., & Reips, U. D. (2016). A limitation of the Cognitive Reflection Test: Familiarity. *PeerJ, 4,* e2395. http://dx.doi.org/10.7717/peerj.2395

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences, 30,* 299–313. http://dx.doi.org/10.1017/S0140525X07001975

Szaszi, B., Szollosi, A., Palfi, B., & Aczel, B. (2017). The cognitive reflection test revisited: Exploring the ways individuals solve the test. *Thinking & Reasoning, 23,* 207–234. http://dx.doi.org/10.1080/13546783.2017.1292954

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a dataset via the gap statistics. *Journal of the Royal Statistical Society Series B. Methodological, 63,* 411–423. http://dx.doi.org/10.1111/1467-9868.00293

Toplak, M. E., & Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 94,* 197–209. http://dx.doi.org/10.1037/0022-0663.94.1.197

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39,* 1275–1289. http://dx.doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20,* 147–168. http://dx.doi.org/10.1080/13546783.2013.844729

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition, 150,* 109–118. http://dx.doi.org/10.1016/j.cognition.2016.01.015

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition, 12,* 473–494. http://dx.doi.org/10.1080/13506280444000409

Velichkovsky, B. M. (1999). From levels of processing to stratification of cognition: Converging evidence from three domains of research. In B. H. Challis & B. M. Velichkovsky (Eds.), *Stratification in cognition and consciousness* (pp. 203–235). Amsterdam, the Netherlands: John Benjamins Publishing Company. http://dx.doi.org/10.1075/aicr.15.13vel

Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhofer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour, 5,* 145–156. http://dx.doi.org/10.1016/S1369-8478(02)00013-X

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). New York, NY: Psychological Corporation.

Zhong, Y., Deng, Y., & Jain, A. (2012). Keystroke dynamics for user authentication. *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Providence, RI (pp. 117–123). http://dx.doi.org/10.1109/CVPRW.2012.6239225

(*Appendices follow*)

Appendix A

Experiment 1

Additional Method

Relational-inference task.

Eye-tracking procedure. In the relational-inference task, participants were seated in a chair with a soft head restraint to ensure a viewing distance of 55 cm from a monitor with 1920 × 1980 resolution. Presentation of the stimuli was performed using a custom-made program implemented using Matlab Psychtoolbox. Eye movements were monitored and recorded using a tower mounted Eyelink, 2000 system (SR. Research Ontario Canada) with a sampling rate of 2000 Hz. A fixation was defined as an interval in which gaze was focused within 1 degree of visual angle for at least 100 ms (Manor & Gordon, 2003). Calibration of the eye-tracking was repeated at the beginning of each block (four times in total). The calibration phase was repeated until the difference between the positions of the points on the screen and the corresponding eye locations was less than 1°. We used a 13-point custom calibration: points were placed at the center of each of the six symbols, at the center of the arrows expressing conditional relations, and in place of the four possible positions of the fixation point.

After the calibration phase, a validation phase was executed to make sure that the calibration had been accurate. The position of each point in the validation phase was identical to the one in the calibration phase. Recalibrations and revalidation were performed if these had been unsuccessful. Moreover, before the beginning of each trial, a drift correction procedure was introduced to force participants to look at the current location of the fixation point. More precisely, stimuli were presented after the fixation point was fixated for 300 milliseconds. The first fixation on each trial was discarded from analysis because its length and spatial location could be biased by the previous fixation point. Stimuli were placed at optimal distance between each other to precisely distinguish goal-directed saccades and fixations.

Eye movements data analysis. To analyze eye movements of participants, we defined six regions of interest (ROIs) centered in each of the six symbols. ROIs had a squared shape with a size of 200 pixels. We discarded every fixation that was not located inside any ROIs. Although a large part of the screen was not included in any ROI, the vast majority of fixations (92.1%) fell inside the ROIs.

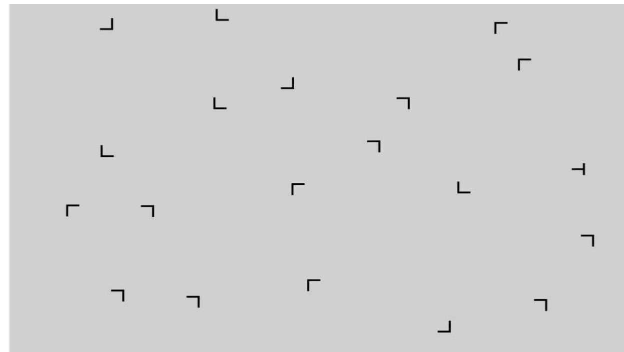

Figure A1. Example of the visual search task (target [t] present).

Visual search control task: Experimental design. In this task, participants had to detect as fast as possible a target among several distractors. The target element was a letter T and was actually present in the array in half of the total 120 trials. Distractors (letter L) as well as target letter were randomly located in the full screen space (Figure A1); the number of stimuli in each trial could be either 16, 20, or 24. In each trial, participants were asked to judge whether the Target letter was present or not, pressing the respective keys on the keyboard (P = present; Q = absent). They were instructed to be as accurate and fast as possible and the task was made incentive-compatible by paying participants based on both accuracy and RTs. In particular, participants received 0.07 euros for each correct trial, from which we subtracted 0.01 euro for each second used to respond. For example, if a participant gave a correct response in 2.37 seconds, she obtained 0.0463 euros in that trial. In case of an incorrect response, the participant received 0 euros. The final outcome of each participant was the sum of the trial-by-trial earnings. Participants were provided with a break (up to 2 min) every 40 trials (two breaks in total).

(*Appendices continue*)

Additional Results

Relational-inference task.

Table A1
Mixed Effect Logistic Model With Trial Accuracy as Dependent Variable, Trial Source State as Independent Variable, and Participant as Random Effect

Overall accuracy	*B*	*SE*	*z*	*p*	95% CI
Source state	.044	.035	1.25	.212	[−.025, .113]
Number of observations	4000				
Number of independent observations	50				

Note. We did not find any effect of source state on accuracy.

Table A2
Gap Statistics for Different Number of Clusters (k: 1–5) Based on 10,000 Monte Carlo Bootstrap Samples

Number of clusters (*k*)	1	2	3	4	5
Gap statistics	.224	.258	.174	.158	.120

Note. The value of gap that best explained data is 2.

Table A3
Correlation Table of Our Six Cognitive Measures

Cognitive measure	CRT	APM	Forward digit-span	Backward digit-span	2-Back	3-Back
CRT	1.00					
APM	.34	1.00				
Forward digit-span	.39	.22	1.00			
Backward digit-span	.40	.11	.49	1.00		
2-Back	.21	.30	.25	.20	1.00	
3-Back	.12	.20	.13	.07	.58	1.00

Table A4
Stepwise Backward Regression Analysis of Group (Sophisticated or Unsophisticated)

Group	*B*	*SE*	*z*	*p*	95% CI
CRT	.78	.32	2.44	.015	[.15, 1.41]
Number of observations	50				

Note. Only cognitive measures surviving the limit for inclusion in the model ($p < .1$) are reported. Measures excluded from the model: APM, $p = .56$; digit span forward, $p = .22$; digit span backward, $p = .24$; 2-back, $p = .16$; 3-back, $p = .59$.

Causal mediation analysis.

Table A5
Linear Model of Representation Index With Our Six Cognitive Measures as Independent Variables

Representation index	*B*	*SE*	*t*	*p*	95% CI
CRT	.40	.15	2.70	.010	[.10, .70]
APM	.03	.14	.20	.839	[−.25, .31]
Forward digit-span	.24	.15	1.60	.116	[−.62, .54]
Backward digit-span	.05	.15	.33	.746	[−.26, .36]
2-back	−.25	.15	−1.66	.105	[−.55, .05]
3-back	.05	.15	.31	.759	[−.25, .34]
Number of observations	50				

Note. This regression will serve as mediator model for causal mediation analysis.

Table A6
Linear Model of Overall Accuracy With Representation Index and Our Six Cognitive Measures as Independent Variable

Overall accuracy	*B*	*SE*	*t*	*p*	95% CI
Representation index	.56	.10	5.47	<.001	[.36, .77]
CRT	.11	.11	1.00	.324	[−.11, .33]
APM	.29	.08	3.02	.004	[.09, .48]
Forward digit-span	−.17	.10	−1.65	.106	[−.38, .04]
Backward digit-span	.26	.10	2.55	.015	[.05, .47]
2-back	.08	.10	.76	.452	[−.13, .29]
3-back	.15	.10	1.47	.149	[−.05, .35]
Number of observations	50				

Note. This regression will serve as outcome model for causal mediation analysis.

Table A7
Linear Model of Overall Accuracy With Our Six Cognitive Measures as Independent Variables

Overall accuracy	*B*	*SE*	*t*	*p*	95% CI
CRT	.33	.13	2.57	.014	[.07, .59]
APM	.30	.12	2.47	.018	[.06, .55]
Forward digit-span	−.04	.13	−.28	.780	[−.30, .23]
Backward digit-span	.29	.13	2.18	.035	[.02, .56]
2-back	−.06	.13	−.46	.645	[−.32, .20]
3-back	.17	.13	1.34	.188	[−.09, .43]
Number of observations	50				

Note. The presence of a significant effect of CRT, absent when controlling for representation index (Table A5), indicates complete mediation of representation index on the relation between CRT and overall accuracy.

(*Appendices continue*)

Table A8
Summary Statistics (Mean and Standard Deviation) of Measures of Performance in the Visual Search Task

Group	Number of observations	Accuracy	RT	RT (correct yes)	Earnings (€)
Sophisticated	25	.91 (.06)	2.02 (0.48)	1.43 (0.26)	5.44 (0.39)
Unsophisticated	25	.90 (.07)	2.00 (0.45)	1.42 (0.24)	5.39 (0.38)
Total	50	.91 (.07)	2.01 (0.46)	1.43 (0.25)	5.42 (0.38)

Note. None of these measures was significantly different across groups.

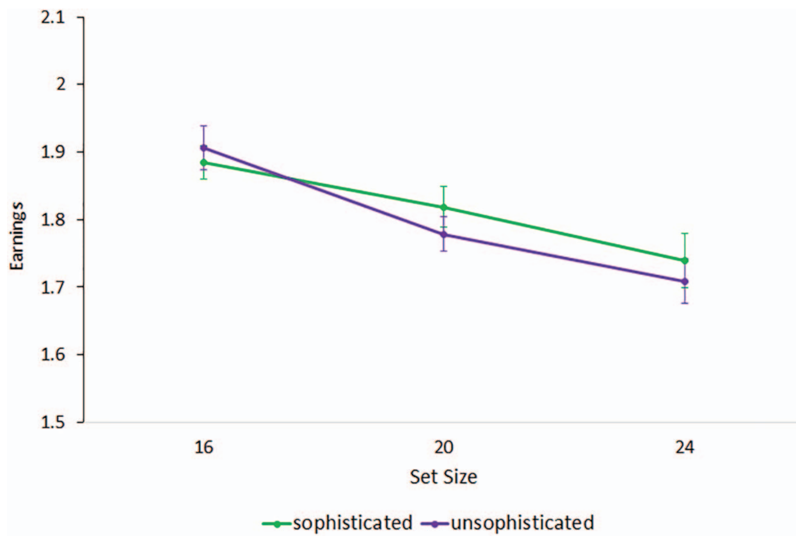

Figure A2. Average earnings of sophisticated and unsophisticated groups by set size. See the online article for the color version of this figure.

Visual search control task. We collected several measures of performance: average accuracy, average RTs, and total earnings (Table A8). We tested between-groups differences performing a two-tailed Mann–Whitney *U* test for each measure of interest. Results did not show any differences in performance across groups (accuracy, $p = .83$; RTs, $p = .88$; earnings, $p = .53$).

To investigate whether task difficulty influenced visual scan efficiency in our two groups, we looked at the magnitude of earnings across set sizes in sophisticated and unsophisticated groups. As shown in Figure A2, both groups decreased their earnings as the difficulty of the task increased. We calculated individual indices of difficulty sensitivity by subtracting earnings in trials with set size = 24 to earnings in trials with set size = 16. No difference in terms of difficulty sensitivity was found across groups (two-tailed Mann–Whitney test, $p = .41$).

(*Appendices continue*)

Appendix B

Experiment 2

Additional Results

Table B1
Linear Model of Overall Accuracy With CRT, APM, and Backward Digit Span as Predictors

Overall accuracy	*B*	*SE*	*t*	*p*	95% CI
CRT	.24	.12	2.07	.044	[.01, .48]
APM	.39	.11	3.48	.001	[.17, .62]
Backward digit-span	.31	.10	3.01	.004	[.10, .52]
Number of observations	55				

Note. All the measures, including CRT, predict performance in the task. One observation missing in the backward digit span (measure not collected).

Table B2
Linear Model of Overall Accuracy With Representation Strategy (Group) and Our Three Cognitive Measures as Independent Variables

Overall accuracy	*B*	*SE*	*t*	*p*	95% CI
Group	−.46	.17	−2.66	.011	[−.80, −.11]
CRT	.17	.11	1.48	.144	[−.06, .40]
APM	.41	.11	3.88	<.001	[.20, .63]
Backward digit-span	.28	.10	2.81	.007	[.08, .48]
Number of observations	55				

Note. When representation strategy is included in the model, CRT score is no more significant, indicating full mediation of representation strategy on the relationship between cognitive reflection and performance. One observation missing in the backward digit span (measure not collected).

Appendix C

Experiment 3

Additional Results

Table C1
Multivariate Regression With Accuracy in Each Type of Inference as Dependent Variables and Proportion of Nonlinear Integrative-Ts in Nonlinear Trials as Independent Variable

Mean accuracy	*B*	*SE*	*t*	*p*	95% CI
MP					
Prop. nonlinear integrative-Ts	.12	.14	.88	.381	[−.15, .39]
AC					
Prop. nonlinear integrative-Ts	.18	.13	1.32	.191	[−.09, .45]
DA					
Prop. nonlinear integrative-Ts	.05	.14	.40	.689	[−.22, .33]
MT					
Prop. nonlinear integrative-Ts	−.01	.14	−.09	.932	[−.28, .26]
MP and DA					
Prop. nonlinear integrative-Ts	.04	.14	.27	.788	[−.24, .31]
MT and AC					
Prop. nonlinear integrative-Ts	.05	.14	.39	.695	[−.22, .33]
Number of observations	56				

Table C2
Multivariate Regression With Accuracy in Each Inference Type as Dependent Variables and Backward Span as Independent Variable

Mean accuracy	*B*	*SE*	*t*	*p*	95% CI
MP					
Backward digit-span	.34	.13	2.65	.011	[.08, .60]
AC					
Backward digit-span	.26	.13	1.98	.053	[−.00, .53]
DA					
Backward digit-span	.35	.13	2.66	.010	[.09, .61]
MT					
Backward digit-span	−.12	.14	−.86	.395	[−.05, .49]
MP&DA					
Backward digit-span	.22	.13	1.66	.102	[−.25, .34]
MT&AC					
Backward digit-span	.24	.13	1.81	.076	[−.03, .51]
Number of observations	55				

Note. One subject excluded from analysis (backward digit span score not collected).

Received January 23, 2018
Revision received March 25, 2019
Accepted March 26, 2019 ■