The Interplay Among Knowledge, Cognitive Abilities and Thinking Styles in Probabilistic Reasoning: A Test of a Model

Francesca Chiesi and Caterina Primi

Abstract Stanovich et al. (Adv. Child Dev. Behav. 36, 251–285, 2008) outlined how people can reach a correct solution when a task besides the normative solution elicits competing response options that are intuitively compelling. First of all, people have to possess the relevant rules, procedures, and strategies derived from past learning experiences, called *mindware* (Perkins, Outsmarting IQ: the emerging science of learnable *intelligence*, Free Press, New York, 1995). Then they have to recognise the need to use and to inhibit competing responses. Starting from this assumption, Stanovich and colleagues developed a taxonomy of thinking errors that builds on the dual-process theories of cognition.

The present chapter presents a set of experiments designed to test the Stanovich and colleagues' model inside probabilistic reasoning. Since rules concerned with probabilistic reasoning (i.e. the *mindware* in Stanovich and colleagues' terms) are learned and consolidated through education, we carried on the researches with students of different grade levels. In particular, we assessed the role of the *mindware gap* (i.e. missing knowledge), taking into account individual differences in cognitive ability and thinking dispositions, and superstitious thinking as *contaminated* mindware (Study 1). Then, we conducted a set of experiments (Study 2) in order to investigate the *override failure* (i.e. the failure in inhibiting intuitive competing responses) in which participants were instructed to reason on the basis of logic or provided with example of logical vs. intuitive solutions of the same task. In this way, we aimed at stressing the need to apply the rules.

Our results provide support for the claim that the *mindware* plays an important role in probabilistic reasoning independent of age. Moreover, we found that cognitive capacity increases reasoning performance only if individuals possess the necessary knowledge about normative rules. Finally, superstitious beliefs seem to have a detrimental effect on reasoning. The overall findings offer some cues to cross the bridge from a psychological approach to an educational approach.

F. Chiesi (✉) · C. Primi
Department of Neuroscience, Psychology, Drug Research, and Child's Health (NEUROFARBA), Section of Psychology, University of Florence, Via di San Salvi 12, Padiglione 26, 50135 Florence, Italy
e-mail: francesca.chiesi@unifi.it

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_11,

According to dual-process theories, mental functioning can be characterized by two different types of process which have different functions and different strengths and weaknesses (e.g. Brainerd and Reyna 2001; Epstein 1994; Evans and Over 1996; Stanovich 1999).

To demonstrate the role of the two types of process in reasoning, consider the following example (Chiesi et al. 2011). Imagine that in order to win a prize you have to pick a red marble from one of two urns (Urn A and B). Urn A contains 20 red and 80 blue marbles, and Urn B contains 1 red and 9 blue marbles. When you respond to the task, you can simply rely on the feeling/intuition that it is preferable to pick from the urn with more red marbles (i.e. the favourable events). In this case, you are using Type 1 processes (which are sometimes called *heuristic*—see, e.g. Evans 2006; Klaczynski 2004) that they are considered to be autonomous because their execution is rapid and mandatory when the triggering stimuli are encountered, they do not require much cognitive effort, and they can operate in parallel. Type 1 processing is the default because it is cognitively economical, and people "feel" intuitively that heuristic responses are correct (Epstein 1994; Thompson 2009). Indeed, as in the example, Type 1 processing often leads to normative correct responses (e.g. Evans 2003; Stanovich and West 1999).

You could also respond to the task comparing the ratio of winning marbles in each urn (20 % vs. 10 %) which requires some time, mental effort and computations. In this case, you are using Type 2 processes, which are relatively slow and computationally expensive, available for conscious awareness, serial, and often language based. In this example, both processes cue the normatively correct answer (that is, Urn A).

On the other hand, it is possible to set up a task where Type 1 and Type 2 reasoning cue different responses. For example, if you can choose between picking a marble from an urn containing 10 red and 90 blue marbles, or from an urn containing 2 red and 8 blue marbles, the feeling/intuition that it is preferable to pick from the urn with more favourable events results in a normatively incorrect choice.

When Type 1 and Type 2 processes do not produce the same output, Type 1 process usually cues responses that are normatively incorrect and, according to dual-process theorists (e.g. Stanovich 1999) one of the most critical functions of Type 2 process in these cases is to interrupt and override Type 1 processing. However, this does not always happen. In the case of a conflict between intuitions and normative rules, even educated adults will predominantly produce heuristic responses (e.g. Klaczynski 2001).

To account for this finding, Stanovich and West (2008; see also Stanovich et al. 2008) suggested that the Type 2 override process crucially depends on whether people detect the conflict between their intuitions and their knowledge about relevant normative rules. These rules, procedures, and strategies derived from past learning experiences have been referred to as *mindware* (Perkins 1995). The concept of mindware was adopted by Stanovich and colleagues (Stanovich and West 2008) in their recent model of the role of knowledge in producing normative responses to reasoning problems (Fig. 1). If the relevant mindware can be retrieved and used, alternative responses become available to engage in the override of the intuitive

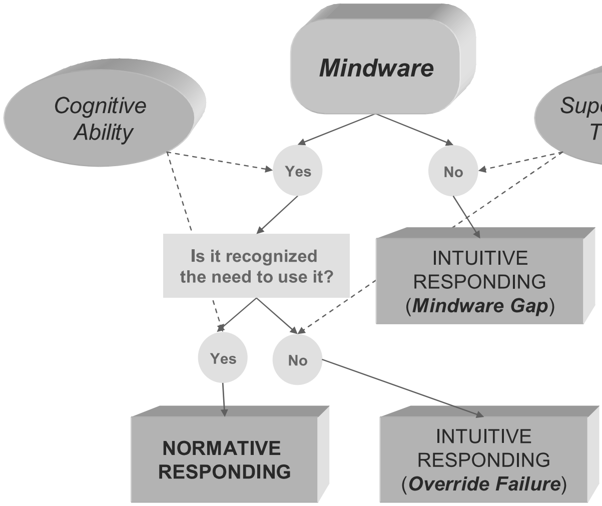


Fig. 1 Simplified representation of Stanovich et al.'s (2008) model on normative reasoning and thinking errors

compelling answers. According to this model, if people do not possess the necessary knowledge to produce a normatively correct response, their errors derive from a *mindware gap* (i.e. missing knowledge). Referring to the previous example, the intuitive answer might be the only one when the fundamental rule of proportions is missing. When relevant knowledge and procedures are not available (i.e. they are not learned or poorly compiled), we cannot have an override since to override the intuitive response a different response is needed as a substitute.

However, even if people detect the conflict between their intuitions and a normative rule, and thus relevant knowledge is available, they can still produce a normatively incorrect response. In this case, their errors result from an *override failure*: Different alternatives are produced and there is an attempt to override Type 1 processing, but this attempt fails usually because people do not have the necessary cognitive capacity to inhibit beliefs, feelings, and impressions, and at the same time, to implement the appropriate normative rules (e.g. De Neys et al. 2005; Handley et al. 2004). So, we have an override failure when people hold the rule but they do not base their answer on it.

Stanovich and colleagues (2008) outlined how people can reach a correct solution when they have necessary cognitive capacity to inhibit competing responses and to use their mindware to solve the task. A number of studies (see Stanovich and West 2000 for a review) found evidence that people with higher cognitive capacity will be more likely to produce normatively correct responses. Kahneman and Frederick (2002) pointed out that higher ability people are more likely to possess the relevant logical rules and also to recognise the applicability of these rules (i.e. they are more likely to overcome erroneous intuitions). Therefore, thinking errors are expected to decrease with increasing cognitive ability (Evans et al. 2009; Morsanyi and Handley

2008). According to Stanovich and West (2008) because it requires considerable cognitive resources carrying out slow, sequential and effortful Type 2 computations, while simultaneously inhibiting quick, low-effort, and intuitively compelling Type 1 responses.

Finally, sometimes errors arise from the use of inappropriate knowledge and strategies that people hold and drive reasoning processes far from the logical standpoint. In this case, the failure is related to *contaminated mindware* in Stanovich et al.'s model. There are various mechanisms that can lead to "contamination". Toplak et al. (2007) propose that a good candidate for contaminated mindware in the case of probabilistic reasoning could be superstitious thinking. In the previous example, the urn with more red marbles might be chosen because the respondent holds the superstitious belief that red is a lucky colour.

Starting from these premises, we present a set of experiments designed to test the Stanovich and colleagues' model inside probabilistic reasoning. The model of Stanovich and colleagues provides a theoretical framework for reconciling the educational and dual-process approaches. Indeed, studies that explore the impact of education on probabilistic reasoning (e.g. Fischbein and Schnarch 1997; Lehman et al. 1988) usually do not investigate the interactions between level of education, cognitive capacity and thinking styles. By contrast, studies inside the dual approach framework typically focus on the effect of cognitive ability, cognitive load, and thinking styles on adults' reasoning including sometimes age-related changes (e.g. Jacobs and Klaczynski 2002; Klaczynski 2009; Brainerd and Reyna 2001; Reyna and Farley 2006).

However, although Stanovich and colleagues (2008) offer a useful framework for investigating the interplay between these factors, they do not make specific predictions regarding changes based on the educational level. Thus, one important aim of the present series of experiments is to investigate knowledge, cognitive capacity and thinking styles simultaneously in population characterized by different educational levels. In particular, we investigated the *mindware gap* (i.e. the missing knowledge) taking into account individual differences in cognitive ability and *contaminated mindware* (i.e. the superstitious beliefs) (Study 1). Then, we explored the *override failure* stressing the role of mindware and the need to use it (Study 2).

1 Study 1: The Mindware Gap

Since mathematical abilities (that is, the mindware) concerned with probabilistic reasoning are learned and consolidated through education, we carried on the research with students of different grade levels. Thus, in the present study we considered children's grade level as an indicator of their knowledge regarding probability. Three experiments were conducted with primary, secondary, and high school students that were presented age-adapted probabilistic reasoning tasks. We expected that younger students, whose computational capacities involved in probabilistic reasoning are less consolidated, should perform worse than older ones. Their performance should be explained respectively by *mindware gap* and mindware availability.

In order to better ascertain the role of mindware, we controlled the impact of cognitive ability and superstitious thinking. In fact, cognitive ability should have a positive effect, whereas superstitious thinking should have a detrimental effect on probabilistic reasoning performance (see Toplak et al. 2007).

1.1 Experiment 1

1.1.1 Method

Participants The participants were 241 primary school students enrolled in primary schools that serve families from lower middle to middle socioeconomic classes in Tuscany, Italy. Children attended grade 3 ($N = 133$, 51 % boys; mean age: 8.3 yrs, $SD = 0.56$) and grade 5 ($N = 118$, 55 % boys; mean age: 10.5 yrs, $SD = 0.50$). These grade levels were chosen since some basics of probability are taught to the fourth and fifth graders following the Italian national curricular programs.[1] Then, we included in the sample students before they were taught probability issues (third graders), and students who had been taught probability issues (fifth graders). Children's parents were given information about the study and their permission was requested.

Measures *Gambler Fallacy Task* (Primi and Chiesi 2011). Following several studies that have measures gambler fallacy in children (e.g. Afantiti-Lamprianou and Williams 2003; Batanero et al. 1994) and college students (Konold 1995), we developed a specific task. A preliminary version of this task was used in a previous study run with children and college students (Chiesi and Primi 2009). It consists in a marble bag game in which different base-rates in combination with two different sequences of outcomes were used. In detail, it was composed of 3 different trials in which the proportion of Blue (B) and Green (G) marbles varied (15B & 15G; 10B & 20G; 25B & 5G). Thus, the present task allows for testing the gambler fallacy with both equally likely and not equally likely proportions. Before the task was presented, children were shown a video in which the marble bag game was played. The bag shown in the video has a see-through corner and instead of drawing a marble from the bag, the marble is pushed into that corner and then moved back. Since the bag remains always closed, visibly the number of the marbles remains always the same. After the video, each participant received a sheet where it was written the following instruction: "*15 blue and 15 green marbles have been put into the bag shown*

[1]Specifically, the curriculum include statistical surveys and their representations, some linguistic/conceptual issues related to possible, impossible, improbable events, and the development of judgement under uncertainty and estimation of odds through games of chance, inside the classical definition of probability.

in the video and one ball has been pushed in the see-through part. It was done a few times and a sequence of 5 green marbles was obtained". The question was: "*The next one is more likely to be...*". The following instruction explain that: "*The game was repeated again and a sequence of 5 blue marbles was obtained*". Then the question was: "*The next one is more likely to be...*". After this first trial, the two other trials were presented changing the proportion of blue and green marbles. For each trial the same questions of the first trial were asked. We formed a composite scores (range 0–6) summing correct answers that represent normative reasoning, i.e. the higher the score, the higher the respondent's ability to avoid the gambler fallacy.

Set I of the Advanced Progressive Matrices (APM-Set I, Raven 1962). To measure children's cognitive abilities the APM-Set I was administered as a short form of the Raven's Standard Progressive Matrices (SPM, Raven 1941; for a detailed analysis to test its suitability as short form see Chiesi et al. 2012b). The Set I of APM is composed of 12 matrices increasing in their difficulty level, and the items covered the range of difficulty of SPM (Raven 1962). These items are composed of a series of perceptual analytic reasoning problems, each in the form of a matrix. The problems involve both horizontal and vertical transformation: figures may increase or decrease in size, and elements may be added or subtracted, flipped, rotated, or show other progressive changes in the pattern. In each case, the lower right corner of the matrix is missing and the participant's task is to determine which of eight possible alternatives fits into the missing space such that row and column rules are satisfied. A score ranging from 0 to 12 was obtained summing the correct answers.

Superstitious Thinking Scale (Kokis et al. 2002). The scale was composed of 8 items referring to superstitious beliefs and luck. Example items are: "I have things that bring me luck" (positively scored), "I do not believe in luck" (negatively scored). The Italian version of the scale was obtained using a forward-translation method and validated through a sample of students from third to eighth grade (Chiesi et al. 2010).

Procedure The Gambler Fallacy task was presented first and then superstitious thinking and cognitive ability were measured using the above described scales. Participants completed them in a single session during school time. The session took about 30–35 minutes altogether. Tasks were collectively administered and presented in a paper and pencil version, and students had to work through them individually.

1.1.2 Results and Discussion

Correlations between the variables measuring probabilistic reasoning and individual differences in cognitive ability and superstitious thinking were computed. Probabilistic reasoning was not correlated with superstitious thinking ($r(N = 249) = -0.11, n.s.$), and it was positively correlated ($r(N = 249) = 0.41$, $p < 0.001$) with cognitive ability. Then, to examine the effect of grade levels (3 and 5) on probabilistic reasoning, a one-way ANCOVA was run in which only cognitive ability was used as covariate. The effect of grade was significant ($F(1, 248) = 6.53$, $p < 0.05$,

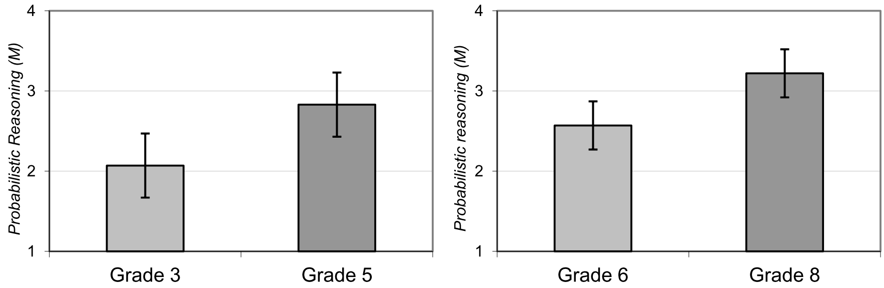


Fig. 2 Mean values representing probability reasoning for each grade group in Experiment 1 (*left*) and Experiment 2 (*right*). Standard errors are represented in the figure by the *error bars* attached to each column

$\eta_p^2 = 0.03$) once the significant effect of cognitive ability ($F(1, 248) = 31.24$, $p < 0.001$, $\eta_p^2 = 0.11$) was partialled out. As expected, grade 5 showed a better performance in probabilistic reasoning ($M = 2.83$, $SD = 1.52$) than grade 3 ($M = 2.07$, $SD = 0.87$). That is, after taking into account the effect of cognitive ability, the grade level continued to have a significant effect on children's reasoning performance (see Fig. 2).

As in previous studies (e.g. Handley et al. 2004; Kokis et al. 2002; Stanovich and West 1999), the present results confirmed the tendency of analytic processing to increase with increasing cognitive ability. Thus, our findings, in line with the argument of Kahneman and Frederick (2002), showed that children with higher cognitive ability are more likely to possess the relevant mathematical and probabilistic rules, as well as to recognise the applicability of these rules in particular situations. Additionally, our results point out that changes in probabilistic reasoning ability are related to new acquired and consolidate mindware from lower to higher grades. Finally, superstitious thinking (see Toplak et al. 2007) did not act as *contaminated mindware*. Arguably, at this age children's judgements were not affected by impression or feelings related to luck or false beliefs about random events.

1.2 Experiment 2

1.2.1 Method

Participants The experiment was conducted with a sample of secondary school students enrolled in schools that serve families from lower middle to middle socioeconomic classes in the same area of the Experiment 1. Students attended grade 6 ($N = 82$, 52 % boys; mean age: 11.8 yrs, $SD = 0.51$) and grade 8 ($N = 121$, 56 % boys; mean age: 13.7 yrs, $SD = 0.57$). These grade levels were chosen since from the 6 to 8 grade Italian students consolidate the use of fractions and proportions that represent the prerequisites for probabilistic reasoning. Students' parents were given information about the study and their permission was requested.

Table 1 Summary of the tasks employed in Experiment 2 and 3 with reference to the origin, requested rule, and related biases

	Source	Rule	Bias
Task 1	Kahneman et al. (1982)	likelihood of independent and equiprobable events	*gambler's fallacy*
Task 2	Kahneman et al. (1982)	likelihood of strings of independent and equiprobable events	*random similarity bias*
Task 3	Denes-Raj and Epstein (1994)	ratios computation and comparison:	*ratio bias*
Task 4	Kahneman and Tversky (1973)	likelihood of one event	*base-rate fallacy*
Task 5	Tversky and Kahneman (1983)	conjunction rule	*conjunction fallacy*
Task 6	Green (1982)	likelihood of one event	*equiprobability bias*

Measures and Procedure Different tasks were employed in order to measure probabilistic reasoning. Tasks were collectively administered and presented in a paper and pencil version, and children had to work through them individually. The students worked through 6 different probabilistic reasoning tasks (Chiesi et al. 2011) adapted from the heuristics and biases literature, as the gambler fallacy or the base-rate fallacy (see Table 1 for the source of each task, the normative principles required to solve them), and the related biases. There were three response options in the case of each task, and children were given either 1 (correct) or 0 (incorrect) points for each response. As in previous studies (Kokis et al. 2002; Toplak et al. 2007; West et al. 2008) the score on the six probabilistic reasoning tasks were summed to form a composite score (range 0–6).

The scales to measure superstitious thinking and cognitive ability as well as the procedure are described in Experiment 1. The administration time for this experiment was about 35–40 minutes.

1.2.2 Results and Discussion

Probabilistic reasoning was negatively correlated with superstitious thinking ($r(N = 193) = -0.27$, $p < 0.001$) and positively correlated with cognitive ability ($r(N = 193) = 0.34$, $p < 0.001$). Then, to examine the effect of grade levels (6 and 8) on probabilistic reasoning, a one-way ANCOVA was run in which both cognitive ability and superstitious thinking were used as covariates. The effect of grade was significant ($F(1, 192) = 7.81$, $p < 0.01$, $\eta_p^2 = 0.04$) once the significant effects of cognitive ability ($F(1, 192) = 12.17$, $p < 0.001$, $\eta_p^2 = 0.06$) and superstitious thinking ($F(1, 192) = 6.93$, $p < 0.01$, $\eta_p^2 = 0.03$) were partialled out. As we expected, higher grade children were more competent ($M = 3.22$, $SD = 1.20$) than lower grade children ($M = 2.57$, $SD = 1.22$) in solving probabilistic reasoning tasks (Fig. 2).

The present results showed, in line with previous studies (e.g. Handley et al. 2004; Kokis et al. 2002; Stanovich and West 1999) and Experiment 1, that children with higher cognitive ability performed better and, differently from Experiment 1, superstitious thinking deteriorated probabilistic reasoning performance, playing a role as *contaminated mindware*. However, the main effect of grade level indicated that some rules need to be taught and exerted.

In sum, Experiment 1 and Experiment 2 demonstrated an effect of education on probabilistic reasoning once controlled the effect of individual differences in cognitive ability and superstitious thinking.

1.3 Experiment 3

In the previous experiments, in order to investigate *mindware gap*, we used the grade level as an indirect measure of the acquisition and consolidation of knowledge related to probability. In this experiment, to establish if people lack or hold the mindware, we measured directly the relevant knowledge needed for dealing with the probability tasks they were asked to solve.

1.3.1 Method

Participants The experiment was conducted with a sample of high school students ($N = 372$, 68 % boys; mean age: 16.3 yrs, $SD = 0.89$) enrolled in schools that serve families from lower middle to middle socioeconomic classes in the same area of the previous experiments. We employed a sample of high school students in order to work with students who had encountered issues related to probability throughout primary to high school years. Parents of minors and students aged higher than 18 years were given information about the study and required a consent form.

Measures and Procedure To measure probabilistic reasoning we employed the *Gambler Fallacy Task* described in Experiment 1 and the tasks described in Experiment 2. We obtained a composite score summing the correct answers (range 0–12). Before performing the tasks, participants were presented four questions measuring knowledge of basic mathematical principles involved in probabilistic reasoning (i.e. the ability to reason correctly with proportions and percentages). An example of item was *"Smokers are 35 % of the population. There are 200 passengers on a train. How many of them will be smokers?"*. Students were given one point for each correct answer, thus a total score (ranged from 0 to 4) was obtained. This score was intended to measure the mindware.

Superstitious thinking was measured using the scale described in Experiment 1. To measure cognitive ability we employed the Advanced Progressive Matrices—Short Form (APM-SF; Arthur and Day 1994; for a detailed analysis of its suitability as short form see Chiesi et al. 2012a). The APM-SF consists of 12 items selected

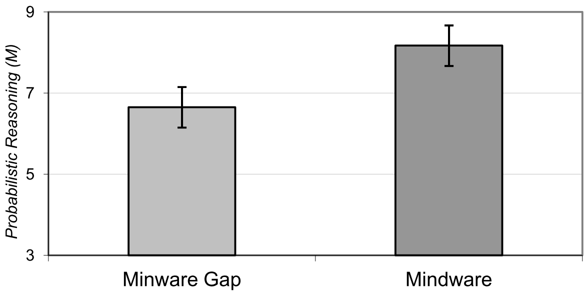


Fig. 3 Mean values representing probability reasoning for each group of high school students differentiated on mindware. Standard errors are represented in the figure by the *error bars* attached to each column

from the 36 items of the APM-Set II (Raven 1962). Participants have to choose the correct response out of eight possible options. A score ranging from 0 to 12 was obtained summing the correct answers.

For the procedure see Experiment 1. The administration session took about 40–45 minutes altogether.

1.3.2 Results and Discussion

Probabilistic reasoning was negatively correlated with superstitious thinking ($r(N = 363) = -0.20$, $p < 0.001$) and positively correlated with cognitive ability ($r(N = 363) = 0.32$, $p < 0.001$). Referring to the mindware scores, we created two groups: mindware gap (scores ranging from 0 to 3) and mindware (score equal to 4). To examine the effect of mindware gap and mindware on probabilistic reasoning, a one-way ANCOVA was run in which cognitive ability and superstitious thinking were used as covariates. The effect of mindware was still significant ($F(1, 354) = 19.16$, $p < 0.001$, $\eta_p^2 = 0.05$) once the significant effects of cognitive ability ($F(1, 354) = 28.71$, $p < 0.001$, $\eta_p^2 = 0.08$) and superstitious thinking ($F(1, 354) = 7.01$, $p < 0.01$, $\eta_p^2 = 0.02$) were partialled out. As we expected, the mindware group obtained higher scores ($M = 8.17$, $SD = 2.05$) than the mindware gap group ($M = 6.65$, $SD = 2.42$) in solving probabilistic reasoning tasks (Fig. 3).

In sum, Experiment 3 confirmed that individual differences in cognitive ability and superstitious thinking affected reasoning but that acquired rules (mindware) represent a necessary tool to deal with probability.

2 Study 2: The Override Failure

The findings of the previous study provide evidence for the relevance of the mindware in probabilistic reasoning. Nonetheless, from these results it is not possible to ascertain if some students were wrong because of override failures, that is, if some students hold the relevant mindware but they do not base their judgements on it. In fact, there is some evidence that even those who give incorrect responses experience a conflict between intuitions (related to Type 1 process) and logic (related to Type 2 process) (e.g. De Neys et al. 2008). That is, those who end up giving an incorrect response might spend some time evaluating the different response options, equally available, but eventually they chose the intuitively compelling one. Thus, an incorrect response does not necessarily imply that the normative option is missed but that it is just ignored, i.e. people may possess more knowledge about rules of probabilities than their answers show. The following three experiments aimed at better exploring this point.

One possible way of distinguishing between reasoning errors that arise from a lack of relevant knowledge and those that are the result of participants' not investing enough effort into implementing the rule properly is to use different instructional conditions. Dual-process theories predict that increasing cognitive effort (that is, increasing the amount of Type 2 processing) should lead to an increase in normative responding. In a study conducted with college students, Ferreira et al. (2006) found that instructions to be intuitive vs. rational oriented the tendency for using heuristic vs. rule-based reasoning when participants solved base-rate, conjunction fallacy, and ratio bias tasks. Similarly, Klaczynski (2001) reported that *framing* instructions to reason "like a perfectly logical person" boosted the performance of adolescents and adults on ratio bias problems.

The aim of the Experiment 4 and Experiment 5 was to ascertain if older secondary school children (i.e. those who possess more relevant knowledge) and college students (i.e. those who had acquired and exerted the relevant knowledge) were able to generate two responses, and to operate an *instruction oriented* choice asking them to answer in an intuitive or logical manner. Moreover, we aimed to verify if cognitive ability mediated the impact of instructions, and we explored these aspects controlling the effect of superstitious thinking. Finally, in Experiment 6 we assessed the effect of a training in prompting the override of the intuitive answer.

2.1 Experiment 4

2.1.1 Method

Participants The experiment was conducted with 126 secondary school students in grade 8 (48 % boys; mean age: 13.7 yrs, $SD = 0.59$) enrolled in schools that serve families from lower middle to middle socioeconomic classes in the same area of the previous experiments. We employed a sample of older high school students in order to work with students who were supposed to hold the relevant knowledge. All students' parents were given information about the study and their permission was requested.

Measures and Procedure We administered the same scales employed in Experiment 2 and 3 but, in order to obtain a more itemed measure of probabilistic reasoning, we added four tasks (Table 2) to the previous ones. The score on the ten probabilistic reasoning tasks were summed to form a composite score (range 0–10).

Table 2 Summary of the tasks added to those employed in Study 1 with reference to the origin, requested rule, and related biases

	Source	Rule	Bias
Task 7	Tversky and Kahneman (1974)	departures from population are more likely in small samples	*sample size neglect*
Task 8	Stanovich and West (2003)	likelihood of independent events	*random similarity bias*
Task 9	Kirkpatrick and Epstein (1992)	ratios computation and comparison	*ratio bias*
Task 10	Konold (1989)	likelihood of compound events	*equiprobability bias*

The procedure was the same as in Experiment 2 except that participants were given one of two different instructions (based on Klaczynski 2001; and Ferreira et al. 2006). In the intuitive condition, participants were told: *"Please answer the questions on the basis of your intuition and personal sensitivity."* In the rational condition, participants were told *"Please answer the questions taking the perspective of a perfectly logical and rational person."* As in Ferreira et al. (2006), a between-subjects design was used where participants were randomly assigned to the intuitive or to the rational condition. Sixty-three students were given rational instructions, and 63 students were given intuitive instructions.

2.1.2 Results and Discussion

In order to investigate the role of cognitive ability on the capacity to cope with the instruction, two groups were created by using the median (9) of the Raven's Matrices score as a cut-off: students below the median formed the low to medium cognitive ability group ($N = 54$, $M = 6.65$, $SD = 1.54$), and students at or above the median formed the high cognitive ability group ($N = 71$, $M = 10.19$, $SD = 0.96$).

To examine the effect of instructions (intuitive vs. rational) and cognitive ability (low vs. high) on probabilistic reasoning, a 2 × 2 ANCOVA was run in which superstitious thinking was used as a covariate. The effect of superstitious thinking was significant ($F(1, 124) = 3.96$, $p < 0.05$, $\eta_p^2 = 0.03$). Once this significant effect was partialled out, the main effect of instructions was not significant as well as the main effect of cognitive ability. The interaction between cognitive ability and instruction was significant ($F(1, 124) = 5.71$, $p < 0.05$, $\eta_p^2 = 0.05$). This was because higher ability children benefited from the instruction to reason logically (rational: $M = 5.17$, $SD = 1.36$, intuitive: $M = 4.43$, $SD = 1.32$) whereas in lower ability children there was no difference between the two instruction conditions (Rational: $M = 4.57$, $SD = 1.73$, Intuitive: $M = 4.13$, $SD = 1.20$). That is, whether students were asked to be rational or intuitive, they performed in the same way (Fig. 4).

The main aim of the present experiment was to manipulate the mental effort that participants invest in solving the tasks. The impact of instructions was moderated by cognitive ability (i.e. we did not find a main effect of instructions). Specifically, only

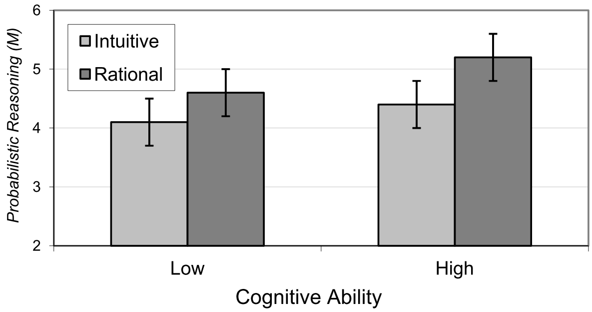


Fig. 4 Mean values representing probability reasoning for each cognitive ability groups (secondary school students). In each group, respondents were differentiated on the received instruction (intuitive or rational). Standard errors are represented in the figure by the *error bars* attached to each column

students with high cognitive capacity benefited from investing effort into thinking "like a perfectly logical person." The most straightforward explanation for this is that higher ability people are more likely to possess the relevant rules, and for them it is easier to recognise when these rules have to be implemented.

2.2 *Experiment 5*

2.2.1 Method

Participants The experiment was conducted with 60 college students (49 % men; mean age: 24.5 yrs, $SD = 3.70$) enrolled in different degree programs (Psychology, Educational Sciences, Biology, and Engineering) at the University of Florence. They were given information about the study and required to fill a consent form. All participants were volunteers and they did not receive any reward for their participation in this study.

Measures and Procedure The same ten probabilistic reasoning tasks used in Experiment 4 were administered. To measure cognitive ability we employed the Advanced Progressive Matrices—Short Form (APM-SF; Arthur and Day 1994) described in Experiment 3. Finally, students were administered the same scale used for children and adolescents to measure superstitious thinking (Chiesi et al. 2010) as done in previous experiments (Morsanyi et al. 2009).

The procedure was the same as in Experiment 4. Thirty students were given rational instructions, and 30 students were given intuitive instructions.

2.2.2 Results and Discussion

To investigate the role of cognitive ability on the instruction, two groups were created by using the median (9) of the Raven's Matrices score as a cut-off. Students below the median formed the low to medium cognitive ability group ($N = 29$, $M = 5.17$, $SD = 1.87$), and students at or above the median formed the high cognitive ability group ($N = 31$, $M = 10.16$, $SD = 1.39$).

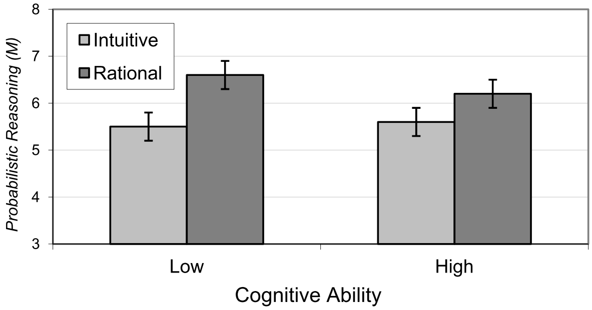


Fig. 5 Mean values representing probability reasoning for each cognitive ability group (college students). In each group, respondents were differentiated on the received instruction (intuitive or rational). Standard errors are represented in the figure by the *error bars* attached to each column

To examine the effect of instructions (intuitive vs. rational) and cognitive ability (low vs. high) on probabilistic reasoning, a 2 × 2 ANCOVA was run in which superstitious thinking was used as a covariate. The results indicated that the effect of superstitious thinking was not significant. There was also no effect of cognitive ability, whereas the main effect of instructions was significant ($F(1,58) = 4.64$, $p < 0.05$, $\eta_p^2 = 0.09$, rational: $M = 6.33$, $SD = 1.27$, intuitive: $M = 5.58$, $SD = 1.23$). No interaction between cognitive ability and instructions was found. That is, college students showed better performance when instructed to be rational (Fig. 5).

Based on the results of Experiment 4, we expected that, given their more consolidated knowledge of probability rules, college students would generally be able to follow instructions, that is, it easier for them to recognise when to apply the relevant normative rules. The results supported these predictions. Whereas, differently from Experiment 4, college students' performance depended mainly on instruction conditions, and there was no effect of superstitious thinking and cognitive ability. Indeed, college students were able to give more normative responses when instructed to reason logically, regardless their cognitive ability. Presumably, because they all had the necessary ability to follow the instructions.

In sum, Experiment 4 and Experiment 5 demonstrated that some students were wrong because of override failures. Thus, some students hold the relevant mindware but they do not base their judgements on it. In fact, when instructed, they were able to use their mindware and make a correct *instruction oriented* choice.

2.3 Experiment 6

In making override failures, people hold the relevant mindware but they are not able to override the compelling intuitive response since some judgement seem to be right regardless logical and rule-based considerations. Using different instructional conditions is one possible way of distinguishing between reasoning errors that arise from a lack of relevant knowledge and those that are the result of participants' not implementing the rule properly. Nonetheless, participants' potential interpretation of the problems and the experimental instructions might represent a limit of this procedure. Thus, instead of giving the instruction to reason logically, in the present experiment we aimed to assess the effect of a training in which the logical vs. intuitive solutions of the same task were presented and compared. In this way, we expected to prompt the override of the intuitive answer when dealing with random events. Given the effect of superstition, we also tried to reduce its impact on reasoning explaining the irrationality of some belief about random events.

2.3.1 Method

Participants Participants were 83 high school students (65 % boys, mean age: 16.1; $SD = 0.56$) enrolled in schools that serve families from lower middle to middle socioeconomic classes in the same area of the previous experiments. As stated above, we employed a sample of high school students in order to work with students who had encountered issues related to probability throughout primary to high school years. Parents of minors and students aged higher than 18 years were given information about the study and required to fill a consent form.

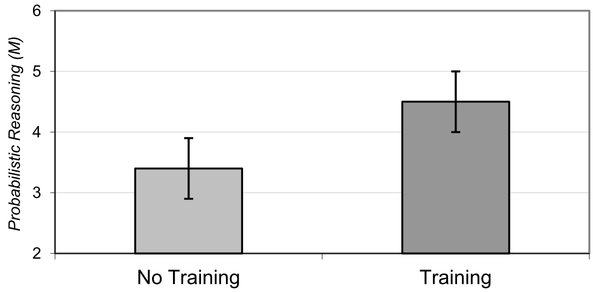
Measures and Procedure As in Experiment 3, participants were presented the four questions measuring knowledge of basic mathematical principles involved in probabilistic reasoning (mindware), and the superstitious thinking and cognitive ability scales. Participants completed this battery of questionnaires in a single session during school time. The session took about 25 minutes altogether.

Then, participants were randomly divided in two groups: the training group ($N = 36$) and the control group ($N = 48$). The training consisted in two units (one hour each). In the first unit, students made experiments with random generators (i.e. throwing dice, sorting a card from a deck) and they were invited to compare intuitions and rule-based considerations. In the second unit, they were showed the irrationality of the superstitious beliefs about random events providing evidences that the outcome of chance events cannot be influenced or controlled. The control group followed a lesson about risk and health behaviours in adolescence. After, in both groups, probabilistic reasoning was measured employing the *Gambler Fallacy Task* described in Experiment 1.

2.3.2 Results and Discussion

As a preliminary step, we had verified that there were not differences between training and control group as regards cognitive ability ($t(81) = 1.82$, $n.s.$), superstitious thinking ($t(81) = 1.16$, $n.s.$) and mindware ($t(81) = 0.90$, $n.s.$). After the experimental manipulation, correlations between the variables in the study—probabilistic reasoning and individual differences in cognitive ability and superstitious thinking—were computed separately for each group. In the training group, probabilistic reasoning was not correlated with superstitious thinking ($r(N = 36) = 0.10$, $n.s.$), whereas it was positively correlated with cognitive ability ($r(N = 36) = 0.47$, $p < 0.01$). In the control group, probabilistic reasoning was

Fig. 6 Mean values representing probability reasoning for the group who followed the training and the control group (No training). Standard errors are represented in the figure by the *error bars* attached to each column



negatively correlated with superstitious thinking ($r(N = 47) = -0.29$, $p < 0.05$) and it was positively correlated with cognitive ability ($r(N = 47) = 0.30$, $p < 0.05$).

To examine the effect of training on probabilistic reasoning, a one-way ANCOVA was run in which superstitious thinking and cognitive ability were used as covariates. The effect of training was significant ($F(1, 78) = 4.31$, $p < 0.05$, $\eta_p^2 = 0.05$) once the significant effect of cognitive ability ($F(1, 78) = 11.76$, $p < 0.001$, $\eta_p^2 = 0.14$) was partialled out (the effect of the other covariate, i.e. superstitious thinking, was not significant). As we expected, the training group obtained higher scores ($M = 4.50$, $SD = 1.99$) than the control group ($M = 3.37$, $SD = 1.81$) in solving the probabilistic reasoning task (Fig. 6).

As expected, activities in which the logical and the intuitive approach to randomness were experienced prompted the override of the intuitive answer. Looking at correlations in each group, this result can be also referred to a possible effect of training in preventing the detrimental effect of superstitious beliefs on probabilistic reasoning.

3 General Discussion

The model of Stanovich and colleagues provides a theoretical framework for reconciling the educational and dual-process approaches. In a series of experiments, we examined the interactions between level of education, cognitive capacity and superstitious thinking in determining reasoning performance and the related errors. As detailed below, the results of the experiments presented in Study 1 (referring to the *mindware gap*) and Study 2 (referring to the *override failure*) are consistent with this model.

Referring to Study 1, in Experiment 1 and Experiment 2, we investigated the effect of grade levels on probabilistic reasoning ability, while controlling for the effects of cognitive capacity and superstitious thinking. Using different age-adapted tasks and scales, we found that grade levels accounted for a significant proportion of variance in probabilistic reasoning once the significant effects of cognitive ability and superstitious thinking were removed. The effect of grade levels might be explained referring to the increases of probability knowledge with education, and supports the assumption that relevant mindware plays an important role in probabilistic reasoning. This finding was confirmed in Experiment 3 in which the relevant mindware was directly measured. Additionally, our findings are in line with the claim that cognitive capacity will be good a predictor of normative reasoning only if the relevant knowledge to solve a task is acquired (see Stanovich and West 2008).

Referring to Study 2, in Experiments 4 and 5, we manipulated experimentally the effort that respondents invested in solving the problems by providing them with instructions to reason rationally/intuitively. Instructions to reason rationally have been found to increase normative performance as in Ferreira et al. (2006) and Klaczynski (2001), especially in the case of higher ability participants, as in Morsanyi et al. (2009) and Chiesi et al. (2011). In the terms of Stanovich and colleagues' (2008) model, increasing the mental effort that participants invest into reasoning will reduce the *override failures*, i.e. people succeed in overriding Type 1 processing and at the same time are able to implement the appropriate normative rules. In Experiment 6, we confirmed that performance was boosted when cues to resist tempting heuristic responses were given, and we confirmed that cognitive capacity will be a good predictor of normative reasoning only when the relevant mindware is available.

Overall the current studies, according to Stanovich et al. (2008), suggest that the correct solution can be reached holding the relevant mindware and recognising the need to use it and that individual differences in thinking styles and cognitive ability can be accounted for explaining some thinking errors. Additionally, they lend support to the claim of developmental dual-process theorists (e.g. Brainerd and Reyna 2001; Klaczynski 2009) that cognitive capacity *per se* is insufficient to explain changes in reasoning performance, because normative responding crucially depends on participants' relevant knowledge.

Although dual-process theories are very popular in many different areas of psychological science, including social, developmental and cognitive psychology, these theories are not without controversy (see, e.g. Keren and Schul 2009; Osman and Stavy 2006), as well as the use of the classic problems in the heuristics and biases tradition that has been criticized on the basis that they create an unnatural conflict between pragmatic/interpretative processes (see, e.g. Hertwig and Gigerenzer 1999). Nonetheless, we think that conceptualising probabilistic reasoning as an interplay between intuitive and rule-based processes offers some cues to cross the bridge from a psychological approach to an educational approach. From a psychological standpoint, the relevance of mindware in probabilistic reasoning stresses the role of education since probability rules might be very hard to derive from the experience. Thus, what learned at school about normative probabilistic reasoning assume a relevance in everyday life in which the ability to make decisions on the basis of probabilistic information is extremely important and the inability to make optimal choices can be extremely costly. From an educational perspective, psychology helps in understanding why probability is a hard subject to learn and teach (e.g. Kapadia and Borovcnik 1991; Shaughnessy 1992). Thus, teachers have to know that students are *naturally* inclined to rely on intuitions, and that normative rules are often at odds with these intuitions that appear to be right regardless rule-based considerations.

References

Afantiti-Lamprianou, T., & Williams, J. (2003). A scale for assessing probabilistic thinking and the representativeness tendency. *Research in Mathematics Education*, *5*, 172–196.

Arthur, W., & Day, D. (1994). Development of a short form for the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement*, *54*, 395–403.

Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, *25*, 527–547.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: dual processes in memory, reasoning, and cognitive neuroscience. *Advances in Child Development and Behavior*, *28*, 49–100.

Chiesi, F., & Primi, C. (2009). Recency effects in primary-age children and college students using a gaming situation. *International Electronic Journal of Mathematics Education*, *4*, 3. Special issue on "Research and development in probability education". www.iejme.com.

Chiesi, F., Donati, M., Papi, C., & Primi, C. (2010). Misurare il pensiero superstizioso nei bambini: validità e attendibilità della Superstitious Thinking Scale (Measuring superstitious thinking in children: validity and reliability of the superstitious thinking scale). *Età Evolutiva*, *97*, 5–14.

Chiesi, F., Primi, C., & Morsanyi, K. (2011). Developmental changes in probabilistic reasoning: the role of cognitive capacity, instructions, thinking styles and relevant knowledge. *Thinking & Reasoning*, *17*, 315–350.

Chiesi, F., Ciancaleoni, M., Galli, S., Morsanyi, K., & Primi, C. (2012a). Item response theory analysis and differential item functioning across age, gender and country of a short form of the advanced progressive matrices. *Learning and Individual Differences*, *22*, 390–396. doi:10.1016/j.lindif.2011.12.007.

Chiesi, F., Ciancaleoni, M., Galli, S., & Primi, C. (2012b). Using the advanced progressive matrices (set I) to assess fluid ability in a short time frame: an item response theory-based analysis. *Psychological Assessment*, *24*, 892–900. doi:10.1037/a0027830.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: when people behave against their better judgment. *Journal of Personality and Social Psychology*, *66*, 819–829.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*, 349–381.

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: when our brains detect that we are biased. *Psychological Science*, *19*, 483–489.

Epstein, S. (1994). An integration of the cognitive and psychodynamic unconscious. *The American Psychologist*, *49*, 709–724.

Evans, J. St. B. T. (2006). The heuristic–analytic theory of reasoning: extension and evaluation. *Psychological Bulletin*, *13*(3), 378–395.

Evans, J. St. B. T. (2003). In two minds: dual process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454–459.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology Press.

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2009). The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal of Experimental Psychology*, *63*, 892–909.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*, 797–813.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*(1), 96–105.

Green, D. R. (1982). *Probability concepts in 11–16 year old pupils* (2nd ed.). Loughborough: Centre for Advancement of Mathematical Education in Technology, College of Technology.

Handley, S., Capon, A., Beveridge, M., Dennis, I., & Evans, J. St. B. T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, *10*, 175–195.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revised: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.

Jacobs, J. E., & Klaczynski, P. A. (2002). The development of decision making during childhood and adolescence. *Current Directions in Psychological Science*, *4*, 145–149.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: the psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Bulletin*, *80*, 237–251.

Kapadia, R., & Borovcnik, M. (1991). *Chance encounters: probability in education*. Dordrecht: Kluwer Academic.

Keren, G., & Schul, Y. (2009). Two is not always better than one: a critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533–550.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *63*, 534–544.

Klaczynski, P. A. (2001). Framing effects on adolescent task representations, analytic and heuristic processing, and decision making: implications for the normative/descriptive gap. *Journal of Applied Developmental Psychology*, *22*, 289–309.

Klaczynski, P. A. (2004). A dual-process model of adolescent development: implications for decision making, reasoning, and identity. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 31, pp. 73–123). San Diego: Academic Press.

Klaczynski, P. A. (2009). Cognitive and social cognitive development: dual-process research and theory. In J. B. St. T. Evans & K. Frankish (Eds.), *In two minds: dual processing and beyond* (pp. 265–292). Oxford: Oxford University Press.

Kokis, J. V., MacPherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26–52.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*, 59–98.

Konold, C. (1995). Issues in Assessing Conceptual Understanding in Probability and Statistics. *Journal of Statistics Education*, *3*(1). Retrieved from http://www.amstat.org/publications/jse/v3n1/konold.html, May 25, 2009.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: formal discipline and thinking about everyday-life events. *The American Psychologist*, *43*, 431–442.

Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology*, *99*, 18–36.

Morsanyi, K., Primi, C., Chiesi, F., & Handley, S. (2009). The effects and side-effects of statistics education. Psychology students' (mis-)conceptions of probability. *Contemporary Educational Psychology*, *34*, 210–220.

Osman, M., & Stavy, R. (2006). Intuitive rules: from formative to developed reasoning. *Psychonomic Bulletin & Review*, *13*, 935–953.

Perkins, D. N. (1995). *Outsmarting IQ: the emerging science of learnable intelligence*. New York: Free Press.

Primi, C., & Chiesi, F. (2011). The role of relevant knowledge and cognitive ability in gambler fallacy. In *The 7th congress of the European Society for Research in Mathematics Education*, Rzeszów, Poland. http://www.cerme7.univ.rzeszow.pl/WG/5/CERME_Primi-Chiesi.pdf.

Raven, J. C. (1941). Standardization of progressive matrices. *British Journal of Medical Psychology*, *19*, 137–150.

Raven, J. C. (1962). *Advanced progressive matrices*. London: Lewis & Co. Ltd.

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision-making: implications for theory, practice, and public policy. *Psychological Science in the Public Interest*, *7*, 1–44.

Shaughnessy, J. M. (1992). Research in probability and statistics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: MacMillan.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah: Erlbaum.

Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding acceptance principle. *Cognitive Psychology*, *38*, 349–385.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–726.

Stanovich, K. E., & West, R. F. (2003). Evolutionary versus instrumental goals: how evolutionary psychology misconceives human rationality. In D. E. Over (Ed.), *Evolution and the psychology of thinking* (pp. 171–230). Hove: Psychology Press.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*, 672–695.

Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: a taxonomy of heuristics and biases. *Advances in Child Development and Behavior*, *36*, 251–285.

Thompson, V. A. (2009). Dual process theories: a metacognitive perspective. In J. Evans & K. Frankish (Eds.), *Two minds: dual processes and beyond*. London: Oxford University Press.

Toplak, M., Liu, E., Macpherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: a dual-process taxonomy. *Journal of Behavioral Decision Making*, *20*, 103–124.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Tversky, A., & Kahneman, D. (1983). Extentional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Bulletin*, *90*, 293–315.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*, 930–941.