UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Psicologia dello Sviluppo e dei Processi di Socializzazione

SCUOLA DI DOTTORATO DI RICERCA IN: Scienze Psicologiche

INDIRIZZO: Psicologia dello Sviluppo e dei Processi di Socializzazione

CICLO: XXIII

DEVELOPMENTAL AND INDIVIDUAL DIFFERENCES IN THE RATIO-BIAS PHENOMENON WITH AND WITHOUT TIME PRESSURE

Direttore della Scuola: Ch.mo Prof. (Clara Casco)

Coordinatore d'indirizzo: Ch.mo Prof. (Lucia Mason)

Supervisore: Ch.mo Prof. (Franca Agnoli)

Dottorando: Sarah Furlan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

Thanks to my advisor Professor Franca Agnoli. All my deep gratitude for the excellent example she provided as a person and professor. I appreciate all her contributions of time and energy to support me. It has been an honor to be her Ph.D. student

Thanks to Professor Valerie F. Reyna, for the joy and enthusiasm she has for research that were contagious and motivational for me. I am grateful for her strong support during my period at Cornell University

Thanks to Prof. Lucia Mason, head of the Developmental Doctoral Program, who gave continuous support and encouragement during the Ph.D. period

Thanks to Steve Poltrock who has taught me, both consciously and unconsciously, the relevance of being logical in science

Thanks to my friends, who helped keeping me sane especially through the past months

I am grateful to those professors who gave important suggestions to improve my research; thanks to students at University of Padua for their support in the data gathering process

Last but not least, I would like to thank my mother, my father and my family to whom this Ph.D. dissertation is dedicated

RIASSUNTO

Il *ratio bias* viene definito come la tendenza sistematica a giudicare un evento dalle basse probabilità di accadimento (per esempio una probabilità di vincita pari al 10%) come più probabile se presentato sotto forma di ampia numerosità (per esempio 10 palline vincenti su 100) piuttosto che di bassa numerosità (1 pallina su 10), nonostante le probabilità di accadimento siano le stesse (Kirkpatrick & Epstein, 1992). Il comportamento decisionale negli eventi ad alta probabilità di accadimento è stato, invece, scarsamente indagato negli adulti e mai in ottica evolutiva. Le poche ricerche a disposizione evidenziano risultati scarsamente coerenti con le ipotesi di partenza.

Secondo la *cognitive-experiential-self theory* (CEST) negli eventi positivi ad alta probabilità di accadimento gli adulti preferiscono i rapporti di probabilità espressi sotto forma di bassa numerosità (ad esempio 9 su 10) rispetto che ad alta numerosità (ad esempio 90 su 100) in quanto i primi sono percepiti come più concreti e di facile visualizzazione. La risposta corretta, coerentemente con lo sviluppo, dipende dal livello di abilità legate al sistema analitico e al ragionamento formale. La *fuzzy-trace theory* (FFT), invece, predice l'opposto, ovvero che le persone preferiscono i rapporti di probabilità espressi sotto forma di alta numerosità (ad esempio 90 su 100 rispetto a 9 su 10) perché semplificano il confronto basandosi esclusivamente sulla quantità maggiore al numeratore: 90, rispetto a 9, offre maggiori possibilità. Secondo la FTT, la risposta corretta dipende dal ragionamento formale ma anche dal concomitante sviluppo dell'intuizione la quale rappresenta l'apice dello sviluppo.

Nell'Esperimento 1 abbiamo indagato se il ratio bias cambia con l'età e diventa più evidente laddove i rapporti di probabilità da confrontare sono caratterizzati da un'elevata difficoltà computazionale. La proporzione di risposte corrette dovrebbe aumentare al crescere dell'età e del livello di istruzione. Sono stati indagati 94 studenti italiani di seconda media, 58 adolescenti italiani e 30 studenti americani della Cornell University. Ciascun partecipante ha risolto un problema a carattere matematico presentato in tre diversi trial. Ogni trial era caratterizzato dal confronto tra due rapporti numerici: il primo, che era costante per ogni trial, era caratterizzato da un rapporto di probabilità espresso sotto forma di bassa numerosità, ovvero 9 su 10. Il secondo, espresso sotto forma di alta numerosità, era diverso per ogni trial: a) 85 su 95 (minore di 9 su 10); b) 90 su 100 (identico a 9 su 10); e c) 95 su 105 (maggiore di 9 su 10). I risultati evidenziano che le risposte corrette aumentano all'aumentare dell'età. Tuttavia, i ragazzi di seconda media rispondono 4.9 volte meglio degli adolescenti nel confronto tra 9 su 10 e 95 su 105. L'analisi delle risposte *biased* mostra che, indipendentemente dalla specificità del trial considerato, i ragazzi di seconda media hanno una moderata preferenza per il rapporto ad alta numerosità. Gli adolescenti, invece, coerentemente con la CEST, mostrano un chiaro bias verso il rapporto a bassa numerosità.

Nell’Esperimento 2 abbiamo indagato se due scenari tratti dalla vita quotidiana attivano rappresentazioni contestualizzate che spingono il bias verso direzioni differenti. I partecipanti erano 157 studenti italiani di seconda media, 131 adolescenti di seconda superiore e 69 studenti americani della Cornell University. Ciascun partecipante ha risolto i tre trial descritti nell’Esperimento 1. I risultati mostrano che in uno scenario le risposte degli adolescenti vanno fortemente nella direzione dei rapporto a bassa numerosità (9 su 10) rispetto alle risposte dei ragazzi di seconda media. Nell’altro scenario il pattern è opposto: i ragazzi di seconda media hanno una preferenza maggiore per il rapporto a bassa numerosità rispetto agli adolescenti. Circa il 75% degli studenti della Cornell University rispondono correttamente e non mostrano alcuna preferenza sistematica per uno dei due rapporti di probabilità.

Nell’Esperimento 3 abbiamo fatto rispondere i partecipanti in condizione di forte pressione temporale in modo tale da comprendere l’interazione tra processamento euristico e analitico. I partecipanti erano 92 studenti italiani di seconda media, 98 adolescenti di seconda superiore e 92 studenti americani della Cornell University. A ciascun partecipante è stato assegnato uno dei tre scenari descritti negli Esperimenti 1 e 2 nei tre trial. La proporzione di risposte corrette diminuisce in tutti i gruppi di età. Inoltre, la pressione temporale, coerentemente con la FTT, favorisce intuizioni corrette basate sui rapporti di probabilità.

Errori sistematici nel confronto tra rapporti di probabilità dipendono dalle quantità numeriche presentate; inoltre, sia l’età che il contesto influiscono sui pattern di risposta. L’abilità matematica e formale sono importanti per processare correttamente l’informazione numerica indipendentemente dal contesto. Allo stesso tempo, troppo tempo per decidere favorisce la creazione parallela di euristiche di ragionamento che, indipendentemente dalle abilità formali, conducono a decisioni errate.

ABSTRACT

The ratio bias is known as the tendency to judge a low probability event as more likely when presented as a large-numbered ratio, such as 10/100, than a smaller-numbered ratio, such as 1/10 (Kirkpatrick & Epstein, 1992). Less is known about judgments when the ratio bias frame is reversed, requiring judgments of high probability events. This frame has been scarcely investigated in adults, never in children, and results, as well as predictions, are mixed.

According to cognitive-experiential-self theory (CEST), adult participants should favor a small-numbered ratio (i.e., 9-in-10) rather than a large-numbered ratio (i.e., 90-in-100) because the former is perceived as more concrete than the latter. A correct response depends on the development and strength of analytic processing. Instead, fuzzy-trace theory (FTT) predicts the opposite: people should favor a large-numbered ratio (i.e., 90-in-100) because they have a preference for simplified rather than exact numerical information and "more is better than less". According to FTT theory, analytic processing and intuition develop together and both of them can lead to a correct response.

Experiment 1 assessed whether the bias pattern changes with age and whether it is more visible when comparing ratios is difficult. The proportion of correct responses should increase with age and instruction. Seventh graders (N = 94), middle adolescents (N = 58) and adults (N = 30) completed three trials of a mathematical scenario. They compared a small-numbered ratio (which was always 9-in-10) to a large-numbered ratio that varied: a) 85-in-95 (smaller than 9-in-10); b) 90-in-100 (equal to 9-in-10); and c) 95-in-105 (larger than 9-in-10). Correct responses increased with age. Seventh graders, however, were 4.9 times more likely than middle adolescents to give the correct response in the comparison between 9-in-10 and 95-in-105. The analysis of the biased preference revealed that, independent of the ratios, seventh graders slightly prefer the large-numbered container, whereas, according to CEST, middle adolescents were biased toward the small-numbered container.

Experiment 2 assessed whether two different real-life scenarios activate world knowledge that triggers the bias in different directions. Participants were 157 seventh graders, 131 middle adolescents, and 69 adults. Each participant completed three trials with a single scenario. The ratios in these trials were the same three ratios used in Experiment 1. In one scenario the responses of middle adolescents were much more biased than those of seventh graders toward the less numerous option. In the other scenario the responses of seventh graders were more biased toward the less numerous option than those of middle adolescents. Adults responded correctly about 75% in every condition and showed little bias.

Experiment 3 investigated how a time-pressure condition influences the interaction between the heuristic and analytic process. Participants were 92 seventh graders, 98 middle adolescents, and 92 adults. Each participant completed the same three trials used in previous

experiments with a single scenario (the same three scenarios used in Experiment 1 and 2). Results show that accuracy decreases in all age groups and that, contrary to traditional dual-process theories, time pressure inhibits both analytic and heuristic processes. In addition, time pressure, according to FTT, favors gist processing.

The present findings do not appear completely congruent with the predictions of CEST or the predictions of FTT. Biases on ratios depend on the magnitudes of the probabilities, and both age and context influence the pattern of responses. Furthermore, formal and mathematical competence is needed to overcome the influence of world knowledge. Time to decide leads to wrong decisions based on heuristics.

Chapter 1

Dual-process theories in cognitive psychology

According to Piaget's theory of formal operations, the major source of errors in reasoning is lack of understanding of the formal rules of inference, which are the epistemological basis for the kind of hypothetical-deductive thinking that appears in adolescence. Children fail transitivity, probability, class-inclusion, and other tasks because they fail to understand the logic of relationships implicit in problem facts. Information-processing theories, in contrast, explain cognitive development in terms of an age-related increase in computational power and complexity, and reasoning errors are explained in terms of local breakdowns in memory (Case, 1992). As a consequence, the clear prediction is that the increasing capacity in cognitive ability, working memory, executive functions and the acquisition of formal reasoning during cognitive development should make adults, and in particular highly educated adults, capable of correctly solving any and all problems that can be identified as related to formal operations.

A large corpus of research started in the 1970s, however, established that adults often reason poorly and make systematic errors in probabilistic reasoning independently of their level of education. Specifically, Tversky and Kahneman (1974) explained that such systematic errors depend on "cognitive shortcuts", or heuristics, that allow the individual to solve a problem in a way that is consistent with the complexity of the task and the limitations of his/her memory capacity and information processing ability (Kahneman, 2003). Heuristics sometimes lead to severe and systematic errors that are costly for individuals, organizations, and society (Milkman, Chugh, & Bazerman, 2009). Swets, Dawes and Monahan (2000) report the relevance that answers to diagnostic problems have in everyday life, and they stress the fundamental utility connected to the accuracy and goodness of our decisions. For example, what is the probability of dying of cancer? Yamagishi (1997) found that University of Washington undergraduates judge the risk to die of cancer to be greater when told that it kills 1,286 people out of 10,000 (12.86%) than when told that it kills 24.14 people out of 100 (24.14%). These kinds of errors matter in everyday-life decision-making.

Evidence that human beings are rational and irrational at the same time (often consciously) led to the formulation of dual-process theories that attribute behavior to a continuous interplay between two separate reasoning processes. In this chapter we focus on some commonalities and some differences of dual-process accounts that are relevant with regard to the ratio bias phenomenon. We begin by reviewing dual-process theories and ratio bias in the literature on adult behavior because of their implications for cognitive development. As Markovits and Barrouillet (2004) pointed out, there is great fragmentation between studies that look at children's reasoning and those that look at adult reasoning. On the one hand, considerable attention has been given to the precocious competencies that children seem to have much earlier than predicted by traditional developmental theories. For example, some recent studies have shown that very young children have correct intuitions about probability (Girotto & Gonzales, 2008; Schlottmann & Christoforou, 2005). Adults' incompetence, on the other hand, has been deeply investigated with regards to those cognitive processes and individual differences that determine what Stanovich (1999) called the *normative-descriptive* gap. This gap describes the deviation of human responses from the performance deemed normative according to various models of decision making and rational judgment (e.g., the basic axioms of utility theory).

Finally, and quite paradoxically, older children, adolescents, and adults are more likely than young children to commit specific reasoning fallacies (Davidson, 1995; Jacobs & Potenza, 1991; Reyna & Ellis, 1994). These contradictory results motivated our investigation of age-related differences in the *normative-descriptive* gap. In this chapter we will present dual-process theory and its account of cognitive development. In addition, we will focus on two dual-process accounts: the Cognitive-Experiential Self Theory (CEST) and the Fuzzy-Trace Theory (FTT) because they are particularly relevant for the ongoing discussion of the ratio-bias phenomenon.

1.1 From perceptual illusions to cognitive illusions

Appearances (like visual illusions) may deviate from or *distort* reality without violating it. For example, In Figure 1, everyone perceives the square A as darker than square B and is astonished to learn that, instead, square A is exactly the same shade of grey as square B. In this visual illusion no perceptual conflict is experienced because being aware that they are the same shade of grey is not sufficient to perceive it.

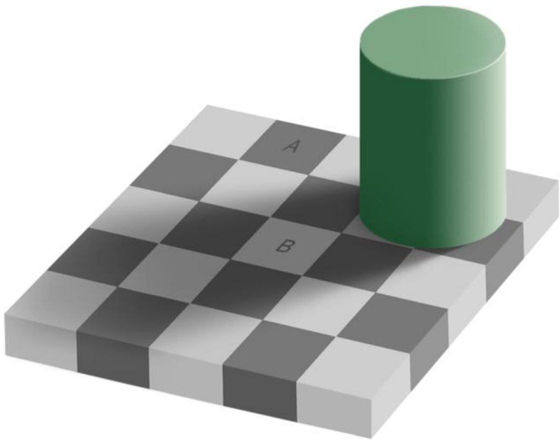

Figure 1. Adelson’s checker shadow illusion from Adelson (1995)

In other kinds of visual illusions people perceive a default image that is prepotent but they can perceive the *hidden* image by inhibiting the prepotent perception. For example, Figure 2 shows a famous perceptual illusion in which the brain switches between seeing a young girl and an old woman.

Figure 2. The young girl and the old woman illusion by W. E. Hill, who published it in 1915 in *Puck* humor

The default image that people usually see is a young woman but, with some effort, people can inhibit the default image and see the old woman. The switch between the two images is easier when you are told about the two patterns and it becomes even easier with practice. Furthermore, when faced with other kinds of perceptual illusions, people experience a conflict every time they look at them and knowing the answer is not enough to inhibit the prepotent perception. The Müller-Lyer illusion (Figure 3) is a famous example in which perception also tells us that one line is longer than the other while logic tells us that it is not.

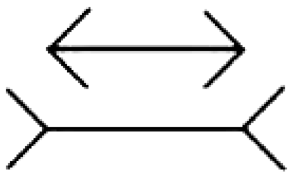

Figure 3. The Müller-Lyer illusion

Even though we can measure the lines and know they are of equal size, our perception of them does not change. We simultaneously experience contradictory information: we see that they are different even if we rationally know that the two lines have the same length. The critical point is that if being aware that the lines are of equal length was sufficient to modify perception, then we might argue that perception and knowledge constitute a single integrated system. Instead, the conflict experience is maintained continuously, implying the coexistence of a perceptual system and an abstract knowledge system. Subbotsky (1990) investigated the Müller-Lyer illusion with children of ages 4, 5, and 6. Two rulers of equal length were attached to a background that made one of the rulers appear longer than the other. When asked about the apparent lengths of the rulers, all children acknowledged that one of the rulers looked longer than the other.

The logic of studying perceptual illusions is that failures of a system are often more diagnostic of the rules the system follows than are its successes. As Gigerenzer (2008) explained, perceptual errors are good errors because sometimes our brain does not have enough information to know for certain everything about an uncertain world.

At the same time, the brain is not paralyzed by uncertainty and, as an "intelligent" machine, it uses heuristics to make bets; this is extremely adaptive. Consequently, those errors that come from this brain's adaptation to reality are good errors and a perceptual system that does not adapt to reality would be not an intelligent system (p. 68).

In a similar manner, people use heuristics to make decisions and the analogy between perceptual illusions and cognitive illusions is widespread in the literature on heuristics and biases in reasoning. For example, Sloman (1996) claimed that perceptual illusions, such as the Müller-Lyer illusion, provide evidence for a dichotomy in a domain other than reasoning (p. 11). Specifically, certain reasoning tasks satisfy what Sloman called *Criterion S* because people believe two contradictory responses. That is the case, for example, in syllogistic reasoning in which people prefer the conclusion corresponding to their own beliefs rather than the logically correct one.

According to Kahneman and Tversky (1972), people use certain heuristics, which have been categorized and labeled, that lead to wrong answers. These heuristics are procedures that are intuitive ways of solving a problem that are in conflict with algorithms or procedures involving rules. The purpose of Kahneman and Tversky's theoretical framework, known as the "heuristics and biases" approach, was to understand the cognitive processes that lead to valid or invalid judgments.

One famous example of a cognitive illusion is the so-called *Linda problem* (Tversky &Kahneman, 1983). The *Linda problem* is probably the most researched vignette in the history of psychology and it was interpreted by cognitive psychologists as an example of people's pervasive irrationality because of its paradoxical results. Linda is described as a 31-year-old woman who is single, outspoken, and very bright. In college she was a philosophy major who participated in antinuclear demonstrations and was concerned with issues of social justice.

Participants were asked to rank the likelihoods of the following possibilities using 1 for the most probable and 8 for the least probable: a) Linda is a teacher in an elementary school; b) Linda works in a bookstore and takes Yoga classes; c) Linda is active in the feminist movement; d) Linda is a psychiatric social worker; e) Linda is a member of the League of Women Voters; f) Linda is a bank teller; g) Linda is an insurance salesperson; and h) Linda is a bank teller and is active in the feminist movement.

Most people make what Tversky and Kahneman (1982) refer to as a "conjunction fallacy": over 80% of adults rated alternative h (Linda is a bank teller and is active in the feminist movement) as more probable than f (Linda is a bank teller). This is a conjunction fallacy (or error) because, according to the conjunction rule, the occurrence of two events cannot be more probable than the occurrence of either one of them, that is p (A&B) ≤ p (A), p (B). Participants judge probability according to representativeness, or similarity. Although the description given was not judged (by other participants) as very representative of women bank tellers, it was judged to be more representative of women bank tellers who are feminists. It is paradoxical that, although the solution to the Linda problem requires the application of one of the most fundamental principles of probability theory, almost everyone—including people sophisticated in statistics—gets it wrong.

The conjunction problem was debated and subsequent studies showed that, for example, adults with higher intelligence are less likely to make the conjunction fallacy in the *Linda problem* (Stanovich & West, 1998b). Fiedler (1988) and Hertwig and Gigerenzer (1999) showed that the conjunction fallacy in the *Linda problem* and similar problems largely disappeared when questions were changed from probability to frequency formats; the proportion of conjunction rule violations dropped from more than 80% to about 10% to 20% (Reeves & Lockhart, 1993). Agnoli and Krantz (1989) showed that the rate of errors in the conjunction fallacy can be substantially reduced by brief instruction in the logic of sets. Such instruction is also effective in the simpler version used with children (Agnoli, 1991). However, as Kahneman and Tversky (1996) pointed out, also the Müller-Lyer illusion disappears when the two figures are embedded in a rectangular frame, but this observation does not make the illusion less interesting (p. 586).

Many cognitive illusions have been studied during the past four decades. For example, in cognitive research on deductive reasoning and thinking, Wason and Evans (1975) showed that less than 10% of university students gave the correct logical response in the Wason selection task (Wason, 1966), a well known task used in the study of conditional reasoning.

With the discovery that adults are not optimal decision makers, research began to focus on errors in judgments, evaluations, estimations and decisions. The abundance of empirical evidence led researchers to claim that these errors were systematic rather than chance-related, causing poor decision making.

1.2 Dual-process theories: a general framework to explain cognitive illusions and reasoning fallacies

The evidence that human beings can be rational and irrational at the same time (often consciously) led to the formulation of dual-process theories that assumed two reasoning processes. The contrast between two coexisting and distinct processes of reasoning has been described either as heuristic versus analytic (Evans, 1989), associative versus rule-based (Sloman, 1996), experiential versus analytic (Epstein, 1991), System 1 versus System 2 (Stanovich, 1999), gist versus verbatim (Reyna & Brainerd, 1995) or Type 1 and Type 2 (Wason & Evans, 1975). Table 1 summarizes the different terminologies that authors proposed as two kinds of contrasting reasoning processes.

The broad framework, generally known as dual-process theories, has been applied to a wide range of studies, including learning (Reber, 1993), social psychology (Chaiken & Trope, 1999), decision making (Kahneman & Frederick, 2002), conceptual reasoning (Sloman, 1996), personality psychology (Epstein & Pacini, 1999; Kirkpatrick & Epstein, 1992), neuroscience (Frank, Cohen, & Sanfey, 2009; Houdè & Guichart, 2001; Steinberg, 2008), risk perception (Reyna, 2004), and economics (Fudenberg & Levine, 2006).

As Keren and Schul (2009) strongly criticized, the terms *system* and *process* are often confused or used as synonymous, but they are not synonymous. Dual-process accounts emphasize the idea that two different kinds of cognitive processes affect inferences and judgments. The conceptualization of two *systems* of reasoning is much stronger because it implies the presence of two distinct cognitive systems in the brain that sharply differ in their evolutionary history (Reber, 1993; Stanovich, 1999). One of the strongest critiques of the two-system approach is that, if two systems exist, then they must have the property of isolability (Keren & Schul, 2009) and this has never been demonstrated completely.

Table 1

Labels Attached to Dual-processes in the Literature

References	**Process 1**	**Process 2**
Fodor (1983, 2001)	Input modules	Higher Cognition
Schiffrin & Schneider (1977)	Automatic	Controlled
Epstein (1994), Epstein & Pacini (1999)	Experiential	Rational
Reber (1993), Evans & Over (1996)	Implicit/tacit	Explicit
Evans (1989, 2006)	Heuristic	Analytic
Sloman (1996), Smith & DeCoster (2000)	Associative	Rule based
Hammond (1996)	Intuitive	Analytic
Stanovich (1999, 2004)	System 1 (TASS)	System 2 (Analytic)
Nisbett et al. (2001)	Holistic	Analytic
Wilson (2002)	Adaptive Conscious	Conscious
Lieberman (2003)	Reflexive	Reflective
Strack & Deustch (2004)	Impulsive	Reflective
Toates (2006)	Stimulus bound	Higher Order
Reyna & Brainerd (1995)	Gist	Verbatim

Note. Adapted from Evans (2008)

In addition, they argue that pure dichotomies are rarely sensible, as many characteristics of mental phenomena are inherently continuous. Therefore, dichotomizing implies oversimplifying. We agree with Keren and Schul (2009) regarding terminology. When presenting dual-process theories, however, we will maintain the authors' terminology, which sometimes refers to reasoning systems. With this exception, we will distinguish between heuristic processing and analytic processing without reference to systems.

According to dual-process theories, two independent processes are involved in the processing of information (Chaiken & Trope, 1999; Klaczynski, 2004; Stanovich & West, 2000; Toplak, Liu, Macpherson, Toneatto, & Stanovich, 2007), as two minds in one brain (Evans, 2003): heuristic processes are characterized as fast, automatic, and cognitively economical. Many adult judgment biases are considered to be a consequence of heuristic responses because they are the by-default responses. Heuristic and analytic processes are assumed to *compete* for control of behavior, like a brain at war with itself (Stanovich, 2004). Heuristic processes are parallel and require low effort; in addition, the idea that heuristic processing evolved earlier than analytic processing is a recurring theme in dual-process theories (Epstein & Pacini, 1999; Stanovich, 1999).

Analytic processes are characterized as conscious, slow, deliberate, analytical, and intrinsically sequential, but despite its limited capacity and slower speed of operation, analytic processing permits abstract hypothetical thinking that cannot be achieved by heuristically; analytic processes are only engaged with effort and may censor outputs of heuristic processes (e.g., Kahneman, 2003). As some researchers claim, heuristic processing may develop in parallel with the analytic processing (Klaczynski, 2000, 2001). In general, the evidence that controlled cognitive processing correlates with individual differences in general intelligence and working memory capacity while automatic processing does not is one of the stronger bases for dual-process theory (Evans, 2008). Heuristics provide default responses that may or may not be inhibited and altered by analytic reasoning.

Table 2 summarizes the main properties associated with the traditional two-process accounts; as Table 2 shows, these properties resemble the distinctions between intuition and rationality made by traditional theories of development. The difference is that, according to dual-process theories, heuristic processes persist into adulthood (explaining biases and also heuristically-produced correct performance), and work together with analytic processes. However, heuristic processes are evolutionarily old and are increasingly displaced (in performance) by more advanced information processing as development proceeds.

Referring to two reasoning systems, Evans (2008) contrasts two classes of dual-system theories that he terms parallel-competitive and default-interventionist. The former refers to the classical two-system models that assume the existence of two isolable systems operating in parallel to generate potentially conflicting responses

(Klaczynski, 2001b). Instead, default-interventionist models (Evans, 2006, Kahneman & Frederick, 2002) replace the isolability assumption with a hierarchy between the systems; a heuristic system (the rapid associative system) supplies the content for the operations of an analytic system (the conscious and controlled system).

Table 2

Attributes Associated to Dual Processes of Thinking

Heuristic process	**Analytic process**
Evolved early	Evolved late
Shared with animals	Distinctively human
Automatic	Controlled
Preconscious	Conscious
Rapid, parallel	Slow, sequential
Operates on contextualized representations	Operates on and constructs decontextualized representations
Associative	Rule-based
Belief-based, pragmatic reasoning	Abstract, logical reasoning
Implicit knowledge	Explicit knowledge
Independent of cognitive capacity	Dependent of cognitive capacity
Intuitive	Reflective
Low effort	High effort
Default process	Inhibitory

According to default interventionist models, the heuristic processing delivers contextualized representations of problems and cues intuitive responses. Evans (2006) explains that heuristic responses can control behavior with or without the intervention of analytic reasoning.

Figure 4 shows the architectural solution proposed by Evans (2006) to explain how the heuristic and analytic processes work. In this model, the assumption is that heuristic processes may bias and shape analytic thinking by the nature of the contextualized representations generated, but the two processes do not compete as parallel processes. The default responses, inferences, or decisions generated by heuristic processes may or may not be inhibited by analytic processes, and the replacement of the default heuristic response depends on different factors. For example, high cognitive ability (or working memory capacity), use of instructions requiring abstract reasoning, and time availability (Evans & Curtis-Holmes, 2005) favor analytic system intervention.

Heuristic responses are usually very quick and can be overridden by analytic processing given sufficient cognitive ability and certain mental dispositions (Stanovich & West, 2000) or given good metacognitive skills (Amsel, Close, Sadler, & Klaczynski, 2009) and inhibitory processes. Based on a default-interventionist conception of analytic processing (Evans, 2008), Frederick (2005) introduced an excellent example showing how heuristic processing works and generates intuitive answers that spring quickly to mind but that are normatively wrong.

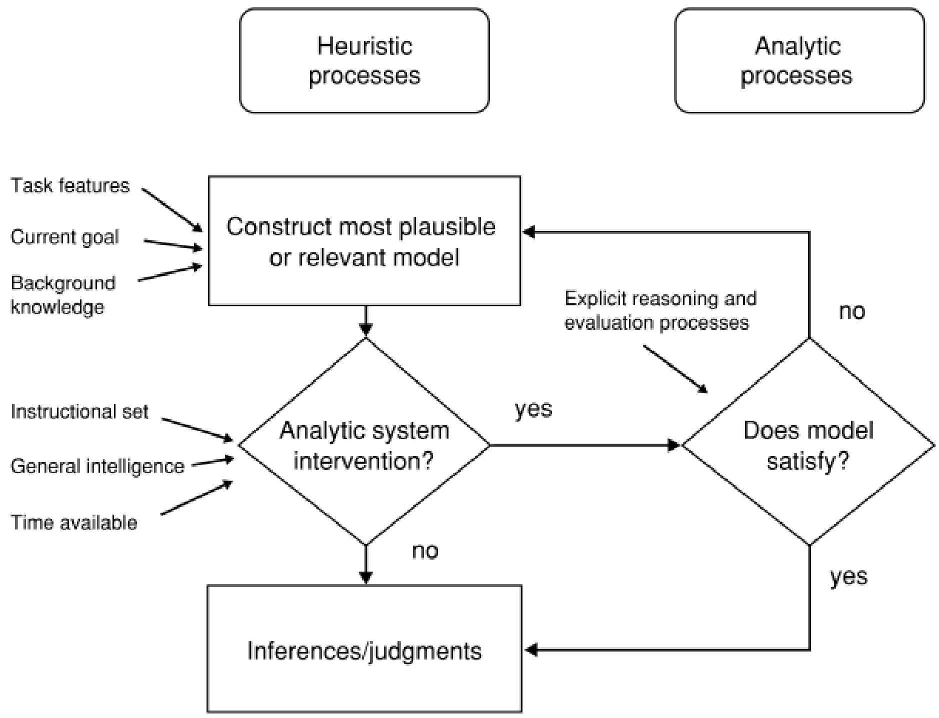

Figure 4. Model of heuristic-analytic theory (Evans, 2006)

Frederick (2005) used a simple three-item Cognitive Reflection Task (CRT) in which the correct answer requires the suppression or inhibition of an erroneous answer that springs “impulsively” to mind (p. 27). The CRT measures “cognitive reflection” which is described as the ability or disposition to resist reporting the response that first comes to mind. One example is the following:

> A bat and a ball cost $1.10. The bat costs $1.00 more than the ball.
> How much does the ball cost? _____cents

In this problem the intuitive response is “10 cents”, but this answer is wrong, because the correct answer is, instead, 5 cents. Frederick (2005) found that most people fail to solve this kind of problem. Nevertheless, when participants discover the correct answer, then this answer appears obvious. In addition, he found that the performance in CRT is positively correlated to the American College Test (.46), to the Wonderlic Personnel Test (.43) that measures a person’s general cognitive ability, and to Need for Cognition Test (.22). In addition, Oechssler, Roider, and Schmitz (2009) showed that participants with high CRT scores were more likely than participants with low CRT scores to avoid logical fallacies and they were also less overconfident.

If reflection ability is the analytic processing capacity to override the heuristic processing functioning, than we can say that the cognitive reflection ability also measures the ability to inhibit the quick response that the heuristic processing provides by default. The inhibition of inappropriate responses is considered one of the key executive functions (e.g., Baddeley, 1996; Dempster & Corkill, 1999; Shallice & Burgess, 1993).

Inhibition of heuristic processing and computations required by analytic processing are assumed to draw on limited executive working memory (WM) resources (De Neys, Schaeken, & d’Ydewalle, 2005). There is evidence that reasoning errors may occur when the capacity of working memory is overburdened and evidence of a link between working memory capacity and performance in a range of reasoning tasks and conditional reasoning (Barrouillet, 1996; Markovits, Doyon, & Simoneau, 2002). Stanovich and West (2000) suggest that individual differences in normative judgments and decisions often arise from working memory capacity limitations on computation.

Memory plays a relevant role in a recent reformulation of Evans’s theory (2009), which re-introduces terminology that was first used over 30 years ago (Wason & Evans, 1975); he proposes calling the two processes Type 1 and Type 2. The rational for a new terminology is that heuristic processing and analytic processing incorporate a multiplicity of systems that take in a wide variety of processes and this is confusing.

Evans's (2009) Type 1 and Type 2 processes share with dual-process theories many characteristics of heuristic processing and analytic processing.

Type 1 processes simply refer to any processes in the mind that can operate automatically without occupying working memory: people have habitual and automated behaviors that once required conscious Type 2 efforts but become automatic with practice and experience. Type 1 processes are defined as massively parallel. Type 2 processes are slow, sequential, and capacity limited. They require access to a single, capacity-limited central working memory resource and thus they can register in consciousness and have properties associated with executive processes and intentional, higher-order control. Type 2 processes are defined as intrinsically sequential.

According to Evans (2009) working memory cannot be matched with one of the systems and we cannot say that working memory is *a* system that does reading, reasoning, planning or explicit learning (p. 38). Furthermore, Systems 1 and 2 cannot distinguish between autonomous processes that control behavior directly without need for any kind of controlled attention and preattentive processes that supply content into working memory. Finally, a distinction is necessary regarding the two systems' working and interaction. According to parallel dual-process theories (or *parallel-competitive*), the two systems provide alternative routes to behavior control, they may or may not conflict, and finally one or the other kind of process takes control of behavior. Evans (2009) claimed provocatively that the parallel model describes a horse race between a very fast horse (Type 1) and a much slower horse (Type 2) where not only the fast horse has to wait for the slow horse to arrive, but the slow horse also gets to decide who has won (p. 48). In sequential dual-process theories (or *default-interventionist*) a fast Type 1 process precedes and shapes subsequent conscious and effortful (Type 2) reasoning; sequential theories concern the interaction between working memory and its many support systems (p. 47).

Evans (2009) explained that both kinds of theory are necessary and he proposed the hybrid model, which incorporates both parallel and sequential processes. He added a third kind of process called Type 3 processes that include resource allocation, conflict resolution, and ultimate control of behavior. These processes, contrary to System 2, cannot be conscious. Figure 5 shows Evans's hybrid model. In this model one system operates entirely with Type 1 processes and the other includes a mixture of Type 1 and Type 2 processes. The defining feature of Type 1 processing is its autonomy and being related to emotions.

As Figure 5 shows, the hybrid model incorporates both the parallel and sequential modes. The parallel mode is represented by concurrent operation of associative, procedural processing on the left side of the figure and rule-based processing on the right side. The sequential mode is represented by perceptual and linguistic processing interacting with working memory on the right side, which provides opportunities for evaluating the adequacy and utility of the response (De Neys & Glumicic, 2008; Evans, 2007). Moreover, the figure explicitly shows the preconscious Type 3 processes that resolve conflicts in the parallel model. Evans (2009) claimed that no longer need the fast horse necessarily wait for the slow horse, as Type 3 processes may decide to pass on the fast response immediately without recruiting or waiting for System 2 processing to complete (p. 49). Finally, Type 2 processes and Type 3 processes interact bi-directionally, which has many implications for conflict resolution.

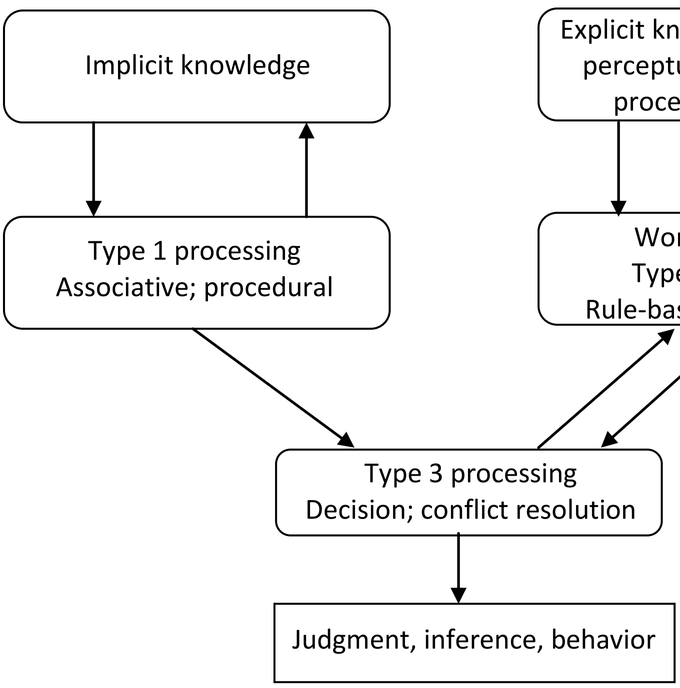

Figure 5. The hybrid model (Evans, 2009)

1.3 Alternative dual-process accounts

Alternative dual-process theorists have pointed out the danger in assuming that normative correctness of responding always implies intervention of analytic processing. Some authors (Gigerenzer & Goldstein, 1999) suggest that people often resort to simple "fast and frugal" algorithms to make inferences because of constraints of limited time and computational capacity. Gigerenzer (2008) explains why heuristics work, and he defines heuristics as frugal because they ignore part of the available information (p. 20). Contrary to common misunderstandings concerning heuristics (i.e., that heuristics are always second-best strategies, or that logic or probability is always the best way to solve problems), Gigerenzer claims that these algorithms may perform adequately.

Gigerenzer and Goldstein (1996) showed that fast and frugal heuristics match or outperform more rational models that integrate various items of information. For example, Gigerenzer and Goldstein (1996) proposed simple heuristics such as *Take The Best* (TTB), which is a memory-based heuristic for inference. Given the choice between two options, TTB looks up the most valid binary cue and infers which of two alternatives has the higher value. First, TTB discriminates between options and searches between cues in order of validity, ignoring all potential additional information. TTB stops the search as soon as a cue discriminates and chooses the alternative the cue favors. On average, TTB uses only a small fraction of the available cue information, and it is therefore "fast and frugal" (Gigerenzer & Goldstein, 1996, p. 650). For example, TTB predicts as accurately or more accurately than multiple regression neural networks.

Another adaptive heuristic is to look for the most successful person and imitate his or her behavior (Boyd & Richerson, 2005). The benefit is cultural evolution. In addition, in many situations, people do not judge the likelihood of relevant events and instead base their decisions on intuitive rules or rationales that appear to fit the circumstances (Rottenstreich & Kivetz, 2006). Consequently, heuristics work because "They exploit evolved capacities that come for free, and thus they can provide solutions to problems that are different from strategies of logic and probability" (Gigerenzer, 2008, p. 27). Heuristics are tools that have been customized to solve diverse problems and heuristics can be extremely useful, as is often the case with experts (Reyna & Farley, 2006).

In the next section we will present two dual-process accounts in detail: Cognitive-Experiential Self Theory (CEST) and Fuzzy-Trace Theory (FTT), because they are crucial for this dissertation.

1.3.1 The cognitive-experiential self-theory (CEST)

According to traditional dual-process theories, judgment biases are the result of heuristic processes (i.e., Kahneman & Tversky, 1982), which refer to the use of cognitive shortcuts to arrive at decisions. Seymour Epstein and his colleagues originally used Kahneman, Tversky and associates' studies of heuristic processing to test the validity of so-called cognitive-experiential self-theory (CEST) and, specifically, study the principles of operation of the experiential system.

Cognitive-experiential self-theory (Epstein, 1973, 1990, 2003) is a global theory of personality that integrates cognitive views on information processing (Shiloh, Salton, & Sharabi, 2002); with psychodynamic theories and personality in different fields (Kemmelmeier, 2010; Klaczynski, Fauth, & Swanger, 1998; Marks, Hine, Blore, & Phillips, 2008; Novak & Hoffman, 2008). This theory is grounded on the evidence that there are two basic modes of processing information in everyday life, and people are intuitively aware of two different ways of knowing: an intellectual and an insightful one (Epstein, 1994) that are often in conflict. As a theory of personality, CEST (Epstein, 1990) assumes that everyone, automatically, constructs an implicit model of the world, or "theory of reality" that includes a self-theory, a world-theory, and connecting propositions. This implicit theory of reality consists of a hierarchical organization of schemas. Toward the apex of the conceptual structure are highly general, abstract schemas; toward the end are narrow, situation-specific schemas (Epstein, 1990). People adaptively learn from their concrete experiences and emotions have a central role in thinking.

As a cognitive theory, CEST undoubtedly shares with some traditional dual-process accounts the two-system view and the assumption that humans operate by two fundamental information processing cognitive systems: a rational system and an experiential system. The rational system operates consciously through a person's understanding of logical rules of inference, and it is analytic, intentional, verbal and relatively affect-free. CEST has little new to say about the development of the rational system. It is obvious, Epstein argues (2003), that people's ability to think rationally improves with age and maturity. However, even if there is widespread agreement among dual-process accounts on the existence of a conscious, deliberative, analytic and rational system, less is clear about the heuristic/experiential system.

CEST provides an insightful perspective on the experiential system. Contrary to Tversky and Kahneman's approach (1973) that introduced the concept of heuristics as separate and convenient strategies or cognitive shortcut, that people use naturally in making decisions in conditions of uncertainty, CEST emphasizes the experiential system as a dynamic and adaptive system that operates at different levels of complexity (Epstein, 2003). Consequently, an experiential system capable of integrating representations and related to personality differs greatly from a view of heuristics selected from a mental toolbox of reasoning procedures (Pennington, 1990, p. 32).

According to CEST, heuristics should not be eliminated but rather they are considered a complex way in which the experiential system operates. Classical conditioning is an example of the operation of the experiential system at its simplest level. The experiential system learns from experience and encodes experience in the form of concrete exemplars and narratives, particularly those events that were experienced as highly emotionally arousing; however, it encodes also in a more general, abstract way. Moreover, the experiential system responds to overall context of situations rather than isolated, abstracted elements; it is emotionally driven, automatic, nonverbal, holistic, intimately associated with the experience of affect and minimally demanding of cognitive resources. Table 3 summarizes some of the main characteristics of the two systems (the experiential system operates according to a set of inferential rules and is less able to comprehend abstract than concrete representations).

Rational and experiential systems are independent, operate in parallel and are interactive. Contrary to dual-process accounts that share, with some differences, the view of the rational system as superior to the experiential system, CEST assumes that all behaviors are the product of the joint operation of the two systems and both of them are equally important. Epstein (2003) explains that on the one hand the rational system seems superior because it is unique to human species and it is capable of a high level of abstraction and complexity.

Table 3

Comparison Between the Experiential and the Rational Systems (An adaptation from Epstein, Lipson, Holstein, & Huh, 1992)

Experiential System	**Rational System**
Holistic	Analytic
Automatic, effortless	Intentional, effortful
Emotional/affective: pleasure-pain oriented	Logical: reason oriented
Associative connections	Logical, cause-effect connections
Behavior mediated by past experiences	Behavior mediated by conscious evaluation of events
Encodes reality in concrete images, metaphors, and narratives	Encodes reality in abstract symbols, words and numbers
More rapid processing, oriented toward immediate action	Slower processing: oriented toward delayed action
Slower and more resistant to change: change with repetitive or intense experience	Change more rapidly and easily; changes with strength of argument and new evidence
More crudely differentiated; context-specific processing; broad generalization gradient; categorical and stereotypical thinking	More highly differentiated, context-general principles
More crudely integrated: organized in emotional complexes	More highly integrated
Experienced passively and preconscious	Experienced actively and consciously
Self-evidently valid; “experiencing is believing”	Requires justification via logic and evidence
More outcome oriented	More process oriented

Nevertheless, the rational system can understand the operation of the experiential system, whereas the reverse is not true (Epstein, 2003, p. 161). For example, the rational system rationalizes the operations of the experiential system; thus, even when people believe that their thinking is completely rational it is often biased by their experiential processing.

On the other hand, the experiential system has demonstrated its adaptive value over millions of year of evolution and is associated with the ability to establish interpersonal relationships and to establish empathy. Personality is a central point in CEST because, for example, people tend to make holistic judgments about personality characteristics and these judgments are adaptively appropriate rather than incorrect according to normative rules. Epstein and colleagues (Donovan & Epstein, 1997; Epstein, Denes-Raj, & Pacini, 1995) disagree with the interpretation of the Linda problem as an example of people's pervasive irrationality. They explain the conjunction error as the result of the unnatural context of the problem; the Linda problem appears to require matching behaviors to personality according to natural experience. Instead, the correct response requires judging the problem according to statistical rules that are unnatural. Consequently, the Linda problem is difficult because it is outside a natural context, not because of a weak knowledge of the conjunction rule or a failure to think of it.

The experiential and rational systems have a reciprocal influence and interact with each other, but it sometimes happens that one threatens the other. In fact, each system has its advantages and disadvantages, and the advantages of one can offset the limitations of the other. In a series of studies (Epstein, Lipson, Holstein, & Huh, 1992), Epstein and colleagues showed that not only it is possible to induce subjects to respond analytically or experientially, based on a change of instructions, but also that subjects whose experiential system had been activated were subsequently less able to respond analytically. The interdependence of the two systems is also demonstrated by the fact that the experiential system may occasionally use language, usually dominated by the rational system; whereas the latter may make use of metaphor and imagination, usually used by the experiential system. The two systems thus can interact, but it also happens that one dominates the other, as in the case in which the rational system detects and blocks an inadequate response produced by the experiential system, or when the experiential system guides the behavior through sensations (Epstein, 1990).

Epstein's research on conjunction fallacy (Epstein, Donovan, & Denes-Raj, 1999), if only effect[1] (Epstein, Lipson, Holstein, & Huh, 1992), and ratio bias (a phenomenon in which a low probability event is subjectively judged more likely when its probability is presented as a ratio of larger numbers; Denes-Raj & Epstein, 1994; Kirkpatrick & Epstein, 1992; Epstein & Pacini, 1999), confirmed the existence of two interactive processing systems, in which the experiential one influences rational thinking, leading people to a non-normative response.

In addition to these two systems, Epstein assumes the existence of a third unconscious system that corresponds to the unconscious mind postulated by Freud (1953, 1959). Freud (1953) regarded rational, conscious thinking as only the tip of an iceberg. He held that the foundation of all mental activity consisted of the submerged part, the unconscious, that operated by a primary unaware process, through wish fulfilment, displacement, condensation, symbolic representation, and association. It was considered essentially a maladaptive system, capable, perhaps, of generating dreams and psychotic aberrations but not up to the task. Therefore, operating under the direction of the primary process alone, individuals would starve to death (Freud, 1959).

Epstein (1994) left the psychoanalytic tradition and began to formulate a new view of the unconscious. This was a *cognitive unconscious*, a fundamentally adaptive system that automatically, effortlessly, and intuitively organizes experience and directs behaviour. Unlike the thinking of Freud (1953), who assumed that all information would be conscious in the absence of repression, the new concept held that most information processing occurs involuntarily and without effort, outside of awareness, intuitively organizing experience and rapidly directing behavior. This natural mode of operation is far more efficient than conscious, deliberative thinking because it does not require that information elaboration spend important resources (Epstein, 1994). According to CEST, the unconscious system, as postulated by psychoanalysis, is not an important contributor to organizing experience and directing behavior in everyday life, because this is the function of the experiential-intuitive system.

[1]The *if only effect* (Epstein, Lipson, Holstein, & Huh 1992) is a particular phenomenon in which people consider one outcome more irritating than another, although both lead to the same consequences and do not involve differences in the actor's direct responsibility.

Finally, CEST assumes that the two systems interact simultaneously as well as sequentially. Sequential interaction is demonstrated in studies of the conjunction problem, in which presenting concrete, natural problems facilitate the solution of abstract problems. Simultaneous interaction, instead, generates a compromise between the two systems and people, for example, experience a conflict between what they know to be the optimal response and what they prefer on the basis of experience ("My reasons tells me to go in one direction, but my feelings tell me to go in another direction"). It is rare for behavior to fall completely at either end of the continuum between the joint products of the two systems (Epstein & Pacini, 2001).

1.3.2 The Fuzzy-Trace Theory (FTT)

In 1990, two papers appeared (Brainerd & Reyna, 1990a; Reyna & Brainerd, 1990) that introduced fuzzy-trace theory (FTT), which comes from the psycholinguistic and memory fields. Often mentioned as a dual-process account (e.g., Morsanyi & Handley, 2008; Osman, 2004), FTT shares some commonalities but also differs in other crucial aspects that we will describe in detail. A relevant aspect is that FTT comes from the psycholinguistic and memory fields. Specifically, the theoretical assumption that there are two processing modes is based on evidence that there are two different memory systems that encode information and influence, for example, probability judgments.

Fuzzy-trace theory (FTT) is a theory of memory and cognition that has motivated many studies of judgment and decision-making, paying attention to their development from childhood to adolescence to adulthood. FTT has relevant similarities with dual-process accounts but it differs from them in some crucial and distinctive points (Brainerd & Reyna, 1990a, 2004).

Integration between memory and reasoning: the role of interference

FTT is a theory of cognition that focuses on the interface between memory and higher reasoning and on developmental changes in these domains. Reyna (1995) claims that many cognitive phenomena have interference as a common denominator.

Interference explains some reasoning errors found in classic phenomena of cognitive development. According to traditional accounts of memory (Bjorklund, 1989; Perner & Mansbridge, 1983), measures of memory and reasoning are dependent: memory enables reasoning and reasoning shapes memory. Traditional views of memory, such as constructivism (Piaget & Inhelder, 1973) and working-memory capacity (Miller, 1956; Baddeley 1976, 1986), predict strong associations between the accuracy of reasoning and the accuracy of memory for problem information.

The idea that improvement in reasoning performance necessitated good memory capacities continued to be a main motivating assumption behind subsequent approaches to judgment and decision making, including both heuristics-and-biases. According to these views, humans used heuristics strategies as mental shortcuts because of information processing limitations (Tversky & Kahneman, 1974). Making optimal judgments and decisions required conserving limited mental resources.

In 1985 Brainerd and Kingma tested the proposition that probability judgments are dependent on memory for exact frequencies, with surprising results. They found that judgments and memory for frequencies were independent. They suggested that reasoning performance might be based on memories for the gist of presented problem information (as opposed to details such as exact frequencies). The reasoning-remembering independence effect was found in standard developmental paradigms such as conservation, transitive inference, probability judgment and class inclusion (Brainerd 1985, Brainerd & Kingma, 1984, 1985; Brainerd & Reyna, 1988).

Reasoning-remembering independence is also supported by data showing that manipulations that affect memory do not affect reasoning, and vice versa. For instance, in transitive inference, increasing the number of premises increases memory load, but has no effect on problem solving (Chapman & Lindenberger, 1988; Reyna & Brainerd, 1990). The conception that humans are assumed to store dissociated representations of different aspects of their experience has achieved particular success in accounting for false-memory effects and situations in which subjects recognize or recall events that did not happen but that are congruent with the gist of their experience.

Building on research in psycholinguistics, the reasoning-remembering independence motivated FTT (Reyna & Brainerd, 1991, 1992) as a new account for cognitive development that incorporates the assumption of multiple representational systems developing at different rates. Gist representations are fuzzy (less precise than verbatim representations) traces of experience in memory, hence the name fuzzy-trace theory; gist representations capture the overall sense of information and refer to the meaning that an individual extracts from information. Gist representations reflect a subject's knowledge, understanding, culture, and developmental level. Because these traces are representations of semantic or other relational information, they can support subjective impressions that targets are a generic type of information previously presented, creating feelings of nonspecific resemblance.

Instead, verbatim memory incorporates episodically coded surface details of stimuli (though these representations are symbolic, rather than literal copies of reality). Verbatim representations are detailed and quantitative: they could be compared to a sort of literal memory (the exact words), whereas the gist could be compared to a memory representation of content (meaning global). Because verbatim traces are representations of actual surface content, supporting the subjective impressions of re-experiencing a target's physical occurrence; they support feeling of conscious remembrance. Unfortunately, as time passes, verbatim traces become inaccessible more rapidly than gist traces. People are generally capable of both forms of thinking, but the gist process prevails when they make judgments. This means, for example, that even if people are capable of understanding ratio concepts, like probabilities and prevalence rates, their judgments will be governed by the bottom-line meaning (the gist trace) and not by the numbers (Reyna, 2008). Subsequent research (Brainerd & Reyna, 1993; Reyna & Kiernan, 1994) demonstrates that three possible relationships between memory and reasoning can be produced by manipulating reliance on these two types of representations: independence, positive dependence, and negative dependence (or interference). The relationship that is produced depends on subjects' ages and characteristics of the task ("task calibration").

Gist and verbatim traces can be viewed as two extremes of a continuum, which implies different degrees of precision of the encoding. The existence of this continuum permits adoption of several representations depending on the specific request of a task. There is, however, an inclination of the mind to shift from verbatim to gist across development. This change might be due either to a greater generalized ability to inhibit responsiveness, or to an increasing reliance on more resilient gist, that permits reasoning to become more resistant to interference (Brainerd & Reyna, 1993, 1995, 2001). The distinction between verbatim and gist representations is important because, in contrast to the assumptions of information-processing theories, verbatim representations are less central to optimal decision making than are gist representations (Klaczynski, 2001b, p. 293).

According to FTT, people rely primarily on gist. Many different studies suggest that reasoning gravitates toward processing the gist of experience, operating on the simplest and least representations, rather than focusing on detailed content (Brainerd & Reyna, 1993, 1995, 2001). Relying first on gist processing has some advantages: even if the content of gist memories is less precise than the content of verbatim memories, they are more widely accessible (gist memories are more persistent over time than verbatim memories). In addition, gist memories are more suitable to different forms of reasoning (gist memories are applicable to a broader range of problem solving contexts) and easier to process (Brainerd & Reyna, 1993, 1995, 2001).

Because two kinds of representations are posited, fuzzy-trace theory is an example of a dual-process model of memory. Fuzzy-trace theory explains findings of reasoning-remembering independence because responses to memory tests often require the details found in verbatim representations, whereas responses to reasoning tests often require only gist representations. Thus, reasoning accuracy is independent of memory accuracy because gist representations are independent of verbatim representations. Evidence of two independent memory representations emerges also from neuroscience. In a study by Aizpurua and Koutstaal (2010) older and younger adults were tested on a picture recognition task that required them to make episodic memory decisions at an item-specific (verbatim) versus category-based (gist-based) level on randomly intermixed trials. Their findings suggest that aging attenuates the capacity to use episodic memory adaptively and flexibly at different levels of specificity. Reyna and Mills (2007) summarize neurocognitive development from childhood to adult age, showing that verbatim-gist distinctions are essential in characterizing memory across the lifespan.

Other characteristics differentiate fuzzy-trace theories from traditional theories of reasoning. First, while the latter were modelled on logic or computation and reasoning was ordered in a series of steps (serial/linear process), FTT claims that reasoning processes unfold in parallel, operating on the barest sense of ideas (the gist) in a fuzzy or qualitative manner (Brainerd & Reyna, 1995). The relation between gist and verbatim traces is defined as parallel rather than serial dependency. The processing results are stored in parallel: one mechanism deposits episodic traces of the exact surface forms of targets and the other deposits gist traces of the same targets by using them as retrieval cues to locate relevant concepts in long-term memory (Brainerd & Reyna, 1993, 1995, 2001). Brainerd (Brainerd & Reyna 1992; Reyna & Brainerd, 1995) claimed that accurate reasoning develops under conditions that demonstrate the interaction between the gist and verbatim representations.

Correct reasoning involves more than just recognizing the appropriate gist in problem information. It also involves inhibiting interference from irrelevant details, editing out irrelevant gists, knowing the relevant reasoning principle (e.g. proportionality), retrieving that principle in context, and correctly implementing that principle (rule, principle, or mathematical operation). Each of these components has been shown to make independent contributions to successful reasoning (Brainerd, 1983; Brainerd & Reyna, 1990b; 1993; Reyna, 1991; Wilkinson, 1982).

Second, contrary to traditional theories that place intuition at the nadir of development, FTT theorists place it at the apex, because of developmental evidence about the typical sequence of errors as reasoners gain expertise in reasoning. Finally, whereas some dual-process theorists elevate emotion above reason, according to FTT emotions are important but they are not an unerring signal of what is adaptive (Reyna, 2004).

Research that support the existence of the two types of traces and the development of reasoning toward intuitive thinking

According to FTT, reasoning errors are less likely to spring from ignorance of logical principles, but rather, they tend to occur because both adults and children fail to extract relevant gist from verbatim inputs, retrieve appropriate gist-processing operations, or apply processing operations coherently to stored representations. For example, in the standard class-inclusion paradigm, children are presented with two sets, say seven cows and three horses, and are asked about the relative numerosity of pairs of sets (Brainerd, Reyna, & Poole, 2000). When asked whether there are more cows or more animals, most children erroneously respond that there are more cows, until the surprisingly advanced age of about 10. Adults routinely make the error on slightly more complex versions of the problem (Reyna, 1991). Why are such problems so difficult?

Piaget (1952) proposed that an inability to reason about sets caused these deficits. However, several studies have demonstrated that, outside the context of this task, children could not only reason about sets, but they could apply the cardinality principle, that more inclusive sets were also more numerous (Brainerd & Reyna, 1990b). Reyna and Brainerd (1990), in fact, claimed that class-inclusion errors occur when people encode the relevant relationships and know the cardinality principle, but fail to retrieve the principle or have difficulty applying it. In other words, they mistakenly report the horizontal relationship between subsets, rather than the vertical one between set and subset. Although counting the number in each subset offers a seemingly foolproof method of verification, people have trouble solving the problem quantitatively (Reyna, 1991), because the relative numerosity of the subsets is salient in the display. In other words, processing the relationship with more inclusive sets quantitatively presents a problem of mental bookkeeping: it is difficult to keep track of the cows who count as cows but also as animals.

Concerning development, younger children are more likely to have problems accessing the cardinality principle, whereas older children and adults may know and access the principle, but they have difficulty processing it.

According to Reyna (1991), the *Linda Problem* is analogous to Piagetian class-inclusion problems (Inhelder & Piaget, 1964). Wolfe and Reyna (2009) suggested that *p* (feminist bank teller) is estimated fallaciously as greater than *p* (bank teller) because the normatively relevant denominator "bank tellers" is ignored. Attention is drawn to the numerator, the "feminists" who are bank tellers, ignoring the larger class of bank tellers that includes feminists plus non-feminists (the appropriate denominator). This conjunction fallacy is a class-inclusion error because all feminist bank tellers are bank tellers and, thus, they cannot be more frequent or more probable than the members of the more inclusive class.

The errors in the class-inclusion reasoning are also similar to those mistakes that people make in the *two-urn-choice-task*, which was first applied by Piaget and Inhelder (1975) to investigate the development of probabilistic thinking. Mistakes happen because this task is a particular example of class inclusion in which there are overlapping inclusion relationships, complicated processing, and horizontal relationships that usurp the role of vertical relationships. Reasoners focus on target members of a class and lose track of the larger universe of possibilities. Neglect of the denominator is the basic assumption to explain people's errors: in tasks characterized by overlapping classes, people focus on the number of times a target event has happened without thinking about the overall number of opportunities for it to happen (Brainerd & Reyna, 2001; Reyna 2004). According to FTT's accounts, these phenomenon are evident both in adults' and children's reasoning and there is no sudden leap from incompetence to competence (Brainerd & Reyna, 2001).

The framing effect is an important *cognitive bias* in which presenting the same option in different formats alters people's decisions: linguistically different descriptions of equivalent options lead to inconsistent choices. Framing effects occur when different framed but equivalent options lead to different choices, demonstrating preference reversals even though options are mathematically equivalent. A typical example is based on Kahneman and Tversky's problem (Tversky & Kahneman, 1981). The authors gave participants two alternative solutions for 600 people affected by a hypothetical deadly disease:

Option A saves 200 people's lives

Option B has a 33% chance of saving all 600 people and a 66% possibility of saving no one

These decisions have the same expected value of 200 lives saved, but option B is risky. 72% of participants chose option A, whereas only 28% of participants chose option B (Tversky & Kahneman, 1981). In general, most people prefer the more certain option in the gain frame and the risky option when the scenario is framed as a loss.

FTT explains framing in terms of verbatim (quantitative, calculative, processing) and gist (categorically driven, qualitative) processing (Reyna & Brainerd, 1991). When presented with statistical information, children are more likely than adults to reason quantitatively, within the limits of their computational knowledge. This is because as development proceeds reasoning generally relies less on exact memory for informational inputs and more on memory for qualitative gist. Adults more than children tend to base their decisions on the gist of information. For framing problems, when something is to be gained, most people prefer something rather than nothing, and so they choose the more certain option; instead, for loss scenarios, nothing is better than something, and so people choose the risky option (Reyna & Ellis, 1994). The classical view, that cognitive development progresses away from intuition and toward quantitative thinking, has been challenged further by demonstrations that adults often engage in intuitive and qualitative reasoning, because the verbatim process is gradually relinquished in favour of gist based processing (Reyna, 1996).

A large body of research (Brainerd & Reyna, 1990a, 1990b, 1993, 1995, 2001; Kühberger & Tanner, 2009; Reyna, 2004; Reyna & Adam, 2003; Reyna & Ellis, 1994; Reyna, Lloyd, & Brainerd, 2001; Reyna & Farley, 2006) shows the encoding of dual verbatim and gist representations of information and the reliance on the latter whenever possible. This research supports the claim that intuitive gist-based reasoning increasingly supplants analytical verbatim-based reasoning as children gain experience and as novices become experts.

1.4 Dual-process theories in cognitive development

Most traditional theories regard cognitive development as a unidirectional progression either from intuitive thinking to logical and formal reasoning (e.g., Piaget) or from an initially inefficient state to a state of greater efficiency (e.g., information processing). As a consequence, the clear prediction is that the increasing capacity in cognitive ability, working memory, executive functions and the acquisition of formal reasoning during cognitive development should make adults, and in particular highly educated adults, able to reason correctly in any and all problems that can be identified as formal operational.

We have seen that adults, instead, are often poor reasoners, rely on inappropriate decision-making shortcuts, and make suboptimal decisions across a wide range of situations. This view of adult reasoning clearly differs substantially from traditional theories of cognitive development that generally posit some kind of replacement: for example, a more adequate and normatively justified reasoning replaces less adequate and non-normative reasoning (Amsel et al., 2008). Dual-process theories challenged traditional accounts of cognitive development because they explain development as the co-development of both analytically based competencies and experientially based heuristics and beliefs. However, as Agnoli (1991) argues, research provides evidence that informal heuristics strongly influence adult reasoning, but we know little about how these heuristics develop, or whether they influence children's reasoning. After 20 years, the question is still open, research in the development of heuristic and analytic responding is limited (Morsanyi & Handley, 2008), and results are mixed.

1.4.1 Heuristic process and analytic process: different developmental trajectories

The developmental trajectory of heuristic and analytic processes remains an open issue because it is unclear how the two processes co-develop, how they interact, and how the nature of these interactions changes with age (Jacobs & Klaczynski, 2002). Specifically, there is general consensus that analytic processes and reasoning improve with cognitive development and formal education (Kuhn & Pearsall, 2000; Janveau-

Brennan, & Markovits, 1999). Because analytic processes rely on general intelligence and working memory, they can be expected to develop with age until late adolescence (Daniel & Klaczynski, 2006: Kokis, Macpherson, Toplak, West, & Stanovich, 2002). Contrary to heuristic processing, analytic processing operates on “decontextualized” representations. To decontextualize task representations is essential if analytic competencies are to be engaged consistently and used effectively (Stanovich, 1999; Stanovich & West, 1997).

According to dual-process theories, the tendency for analytic processing to override heuristic processing is expected to increase with development (Kokis, et al., 2002), and it is also expected to be positively associated with differences in computational capacity among individuals of the same age. This makes sense with regard to a number of results regarding adult behavior (e.g., Kahneman & Tversky, 1979; Smith & Levin, 1996; Stanovich & West, 1998b, 1999), but research with children has not always showed the expected developmental pattern. Counterintuitive developmental age trends are often neglected by adult research *because of the implicit assumption that, if adults perform poorly on reasoning and decision-making tasks, then children’s performance must be even worse* (Klaczynski, 2009, p. 265).

On the other hand, it is less clear when informal heuristics are learned and how they develop. Children may lack heuristics at an early age but acquire them gradually as a consequence of experience despite formal education. For example, young children might show a decontextualized (correct, indeed) response by default because of their cognitive immaturity rather than because of a decontextualization process that operates on contextualized representations that children do not already have (Morsanyi & Handley, 2008). This is coherent with the view of two reasoning processes that are independent, develop in parallel and account for the key role of analytic processing in controlling and replacing heuristic responses given enough cognitive capacity.

On the contrary, children might acquire such heuristics very early, apply them very broadly, and gradually learn to restrict their use as a consequence of formal education or some dispositions toward an analytic thinking style. For example, Fisk, Bury and Holden (2006) found that even 4-year olds commit the conjunction fallacy.

Possibly heuristic responses in some tasks remain constant, whereas normative responses increase. In this case, the heuristic system would be relatively independent of general intelligence and age (Reber, 1993; Stanovich, 1999)

Klaczynski (2001b) investigated the trajectory of heuristic and analytic responses in early adolescents (12-year olds), middle adolescents (16-year olds), and adults. Six decision-making problems involving probability judgment, counterfactual thinking and sunk cost (when current decisions are influenced by inconsequential past decisions; Arkes & Ayton, 1999; Arkes & Blumer, 1985) decisions were shown to each participant, and the final question about their preferences was framed either as a *usual question* or as a *logic question*. The purpose of the logic question was to elicit a shift from participants' *usual* processing to analytic processing (p. 296). Klaczynski (2001b) found that normatively correct responses increase and heuristic responses remain stable with age but predominant across ages. Finally, the change in perspective from *usual* to *logical* was effective in all age-groups. Klaczynski (2001b) interprets such results as evidence that the prevalence of heuristic responding is not, itself, an indicator of its adaptive value and he makes a clear distinction between heuristic processes themselves and heuristic products of the heuristic processes. He argued that many heuristic products have a cultural basis, and they can be the result of overgeneralizations of automatic responses. However, since Klaczynski (2001b) did not include children in his sample, the trajectory is limited to adolescents.

Amir and Williams (1999) investigated the role of culture in the development of probabilistic concepts with six graders of different ethnicity (English, Asian, or African) to understand how children's culture (beliefs, language, and experience) and the informal knowledge (informal knowledge, primary intuitions, and heuristics) that children acquire in daily life from their culture may interfere with their learning of probability. They found connections between culture, beliefs, and probability thinking; however, these differences were mediated by the use of everyday language. The relevance of language is strongly supported by CEST (Epstein, 1990), which assumes that infants, before acquiring language, can only respond in the experiential mode. The rational system develops in relation to the acquisition of language. Language permits a higher level of abstract reasoning and a distancing from immediate experience. Not only is the older child able to respond with a wider array of rational reactions as the result of maturation, the acquisition of language, and of training, but the older child also learns about the priority that is expected to be given to rational processes under most circumstances.

If heuristic processing depends on the activation of shortcuts that are acquired though experience, then people's repertoire of heuristics should become increasingly diverse and more easily activated with age. Klaczynski and Cottrell (2004) argue that this does not imply that adults will necessary use heuristics more than children, but instead, when experiential processing is predominant, adults' judgments and decisions will reflect more variability in the types of heuristics they use (p. 150). Some studies, however, show that heuristic processing increases with age, and often children perform well on tasks that seem to perplex adults.

Jacobs and Potenza (1991) provide the first evidence that some heuristics increase with age. They presented different scenarios modelled on Kahneman and Tversky's (1972) judgment tasks to first graders, third graders, and sixth graders, and they charted developmental trends in the use of base rates and the representativeness heuristic (which happens when people judge the probability of an event by finding a 'comparable known' event and assuming that the probabilities will be similar). College students formed a comparison group. The scenarios varied by domain (social versus object judgment) and by information provided (base rates alone versus base rates and individuating information). An example of an object judgment scenario with base rates alone is the following:

> Mike's dresser drawer contains three pairs of white socks and six pairs of coloured socks. One morning he is late for school so he reaches in and grabs a pair of socks without looking. Which kind of socks do you think he got out of the drawer?

Instead, an example of social scenario with base rates and individuating information is:

> In Juanita's class 10 girls are trying out to be cheerleaders and 20 are trying out for the band. Juanita is very popular and very pretty. She is always telling jokes and loves to be around people. Do you think Juanita is trying out to be a cheerleader or for the band?

The findings indicate that in those problems that do not involve social content, the use of base-rate information increases by grade when individuating information is available. A similar increase in the use of base rates was found for social-domain tasks when the base rate was presented without individuating information in the social domain. Paradoxically, the age trend reverses when both base rate and individuating information are provided in the social scenario. Older subjects, including adults, generally commit the representativeness fallacy and ignore frequencies when the qualitative information matches the stereotype of the less frequent category. This means that adults are capable of using base rates, but also that base rates are considered less informative for social judgments.

Jacobs and Potenza’s (1991) study has been criticized because social scenarios imply a knowledge of the social stereotype, such as that popular girls are drawn more to cheerleading than to band. Because knowledge of this stereotype is expected to increase with age and experience, younger children perform better than older children and even adults in the social condition; younger children generally lack knowledge of the stereotype and may not be reasoning more normatively according to base rates (Stanovich, Toplack, & West, 2008). The same problem was detected in Davidson’s (1995) study that showed the counterintuive finding that susceptibility to the conjunction fallacy increases during the elementary-school years. Davidson (1995) gave second, fourth, and sixth graders scenarios that were all in the social domain and included descriptions of either the elderly or children in stereotypical or non-stereotypical situations. An example of a scenario is the following:

> Mrs. Hill is not in the best health and she has to wear glasses to see. Her hair is gray and she has wrinkles. She walks kind of hunched over.

Then, the children were asked to judge how likely Mrs. Hill was to have various occupations, such as Mrs. Hill is “an old person who has grandchildren,” and “an old person who has grandchildren and is a waitress at a local restaurant.” In Davidson’s study, second graders gave more class inclusion responses than sixth graders (65% vs. 43%), and this may be attributed to the content of the scenario and the fact that children are known to hold negative views of the elderly.

Furthermore, Markovits and Dumas (1999) report age increases (among first, third and fifth graders) in biases on transitive inference problems (i.e., A = B = C, therefore A = C) in situations with particular social content (inferences about nonfriendship), even though competencies improve over this same age range.

Some investigators (Macpherson, 2001) argue that the counterintuitive findings that children perform better than adults in these three studies (Davidson, 1995; Jacobs & Potenza, 1991; Markovits & Dumas, 1999) might be due to artifacts in the experimental design. According to this view, developmental trajectories toward more biased reasoning depend on a confounding in the interpretation of the results rather than either an age-related reduction in analytic processing or an increase in the normative/descriptive gap. However, according to some authors, it is reasonable that this gap increases with age; for example, Klaczynski (2001b) hypothesized that, if heuristic processing is an adaptive tool and cognitively advanced responding increases with age, the normative/descriptive gap must increase with age (p. 291).

This hypothesis may be correct of course, but it does not account for other research in which non-normative responding increases with age, and this cannot be explained by increasing capacity. For example, Reyna and Ellis (1994) investigated the development of framing effects that occur when preferences vary across superficial variations in the description of the same options (Tversky & Kahneman, 1986). For instance, people react to an 80% survival rate differently than a 20% mortality rate, and they are more risk-seeking when the outcome is a possible gain than when it is a possible loss. Reyna and Ellis (1994) found that preschoolers do not exhibit the framing effect and seem to be “more rational” than second graders and fifth graders (and even adults).

In addition, research indicates that conditional reasoning fallacies increase with age (Klaczynski & Narasimham, 1998); preadolescents avoid the sunk cost fallacy but adults do not (Arkes & Ayton, 1999). Finally, college students fail Piagetian concrete operational tasks such as conservation of mass (Winer, Craig, & Weinbaum, 1992). For example, they think that they weigh more when sitting down than when standing up.

1.4.2 Cognitive development according to FTT

The direction of development in reasoning is usually taken to be from intuition to computation. According to Piaget (Inhelder & Piaget, 1958; Piaget & Inhelder, 1975), this implies on the one hand, that children make reasoning errors until the ultimate stage of formal operations is reached (more or less 12 years old). On the other hand, adults are able to reason and make decisions in a rational way, maximizing gains and minimizing losses. Contrary to Piagetian conception of thinking as the rigid application of logic rules to premise-like input, FTT views thinking as fundamentally intuitive. Not only does the conception of thinking differ in FTT, so do ideas about development. The fuzzy-processing preference is ascribed to mature reasoners, and represents a flexible and adaptive approach to reasoning that, overall, has the effect of reducing errors. This is the opposite of the traditional view of development, which progresses away from intuition towards greater logic or computation.

Reyna and colleagues (Reyna, 1996; Reyna & Ellis, 1994; Reyna & Farley, 2006) affirm that children's competence develops much earlier than traditional theories predict and that empirical evidence undermines the view of adults as rational decision makers. Children as young as five or six can make accurate probability judgments (Brainerd, 1981; Acredolo et al., 1989). The probability concept is present early in development, although implementing that competence steadily improves with age. The implication is that when those same children become adults, they perform at least as well, if not better than, young children.

FTT includes some major assumptions about cognitive development (Brainerd & Reyna, 1990) of the processes involved in going from initial informational inputs to ultimate response outputs. At all ages, people encode information according to the *reduction to essence* rule of fuzzy-trace theory (Brainerd & Reyna, 1998b). As items of information are encoded, they are mined for their senses, patterns and gists, and incoming data are rendered for essence. Often gist can be derived at various levels of abstraction from verbatim information, in which events are distinguishable according to a *hierarchy of gist* that corresponds to degrees of exactness roughly analogous to scales of measurement (Reyna & Brainerd, 1995b). The fine-grained quantitative distinctions represented in verbatim memories are analogous to ratio-level information, and the categorical distinctions of lower level gist are like crude nominal information.

FTT assumes that at all ages subjects prefer to reason intuitively, relying on fuzzy traces. Fuzzy traces have several advantages compared to verbatim and detailed traces: for example, fuzzy traces are more available than verbatim traces in storage. Fuzzy traces enjoy advantages over verbatim traces in other areas such as processing complexity, parallel processing, and effort. Different probability judgments are generated from different levels of gist that differ in complexity. For example, in the framing paradigm (i.e., to save 200 of 600 people or accept a 1/3 chance of saving all 600), probability judgments may be generated from three levels of gist. At the more elaborate level, subjects can consider the numerical and exact information (verbatim traces). To save 200 of 600 people is the same as accepting a 1/3 chance of saving all 600 because probabilities are the same. However, according to FTT, this level operates when arithmetical calculations, including difficult ratio computations, are required to generate probability judgments. Instead, at a lower level of complexity, subjects encode the information as a comparison between "fewer live than die or all may live, or, even more simply, as "to save some people versus to save some or save none". As a consequence, probability judgments are generated from such representations as "saving some lives (200) is better than saving none".

When precise information is required, such as memory for details or judgments of exact quantities, verbatim memory representations must be tapped to accomplish the task (Reyna & Brainerd, 1995a). Otherwise, gist representations are relied on as the default mode of processing. What develops with age and experience, according to research, is the dissociation between gist and verbatim representations depending on the task (called *task calibration*, which posits that the level of precision required in a response constrains the level of representation recruited; Reyna & Kiernan, 1994, 1995).

Adults show a *fuzzy-processing preference* that represents a system-wide adaptation to the limits of information processing, a means of avoiding systematic errors caused by poor verbatim memory. In this developmental interaction between the two parallel processes, reasoning errors are seen as low-level bookkeeping mistakes that are made even late in development among advanced reasoners rather than a lack of logic or conceptual competence (Piaget & Inhelder, 1975).

Brainerd and Reyna (2001) claim that children learn to rely on gist representations that are operated over by the primary system for accurate and efficient reasoning and that this persists into adulthood. FTT (Brainerd & Reyna, 2001; Reyna, Lloyd, & Brainerd, 2003) predicts that reasoning develops from computational (verbatim) to intuitive (gist-based) thinking, and this is the only developmental theory that places intuition at the apex rather than at the nadir of development. The functional dissociation between gist and verbatim memory, as well as the fuzzy-processing preference, also increase with age (Brainerd & Reyna, 1993; Reyna 1991). FTT departs sharply from traditional approaches in assigning a central role to gist in advanced reasoning. In this view, a *fuzzy-processing preference* represents a system-wide adaptation to the limits of information processing, a means of avoiding systematic errors caused by poor verbatim memory. Development, in this view, consists of increasing resistance to interference, and greater reliance on gist magnitude relationships. Across a wide range of perceptual and inferential tasks, older age groups are more likely to focus on global patterns in judgment and decision making, whereas younger groups are more likely to focus on superficial details (e.g., Brainerd & Reyna, 1993; Carey & Diamond, 1977; Liben & Posnansky, 1977; Perner & Mansbridge, 1983). Output interference increases the amount of task-irrelevant information that must be gated out from working memory. Developmentally, output interference affects children more than young adults (Brainerd & Reyna, 1998).

This view of intuition supports other developmental studies of children’s learning and of adults’ acquisition of expertise, which show a progression from detail-oriented and computational processes to fuzzy and intuitive processing (Davidson, 1991; Davidson, Suppes & Siegel, 1957; Jacobs & Potenza, 1991). Fuzzy-trace theory is a dual-process approach that assumes both early analytical competence and a developmental increase in intuitive reasoning (as a result of greater experience and knowledge). This means that analytical competence is present early and that, even if children are capable of both intuitive and (rudimentary) computational processing, they rely more than adults on precise and verbatim details. Intuitive processing also has predictable pitfalls such as *denominator neglect*, the tendency to assume equal denominators when comparing ratios, in order to apply magnitude estimation (which if denominators were actually equal, would deliver consistently correct responses). Such denominator neglect was observed by Offenbach, Gruen and Caskey (1984) and by Callahan (1989) and similar examples from everyday cognition are discussed by Reyna and Brainerd, 1993. In FTT quantitative and intuitive processing are indeed separate and independent, especially for adults.

Brainerd (2004) describes the relationship between verbatim and gist processes as following U-shaped and inverted U-shaped patterns of development. Increases in a particular treatment variable (e.g., forgetting or varying the exposure duration of stimuli in subliminal semantic activation tasks) will lead to increases in the influence of the primary system (i.e., gist memory), which decreases performance on a behavioral measure (e.g., recalling a word list), resulting in U-shaped performance curves. Increases in the treatment variable produce larger increases in the influence of the secondary system (i.e., verbatim memory) than in that of the primary system. Increases in the treatment create equilibrium, represented as the plateau on the developmental curve. Eventually, increases in the treatment, in turn, increase the influence of the secondary system, which then dominates the primary system. The same principle works in reverse for inverted U-shaped curves.

Chapter 2

The ratio-bias phenomenon

The term *ratio* is strongly related to the concept of probability and it is represented by a pair of positive numbers used to compare two sets. Most adults and children know that the probability of any event is expressed as a ratio of the number of potential outcomes that may be considered successful (the numerator) over the number of all possible outcomes, which is equal to the sum of successful and unsuccessful outcomes (the denominator). Moreover, decimals, percentages, proportions and fractions are all considered ratio concepts in the probabilistic judgment literature, and difficulties with ratio concepts are widespread (Reyna & Brainerd, 2008). Considering the ratio-bias phenomenon and the way in which it has been investigated and explained, fractions are the most frequently used ratio concept. Fractions denote rational numbers as the ratio of two integers, and they represent continuous magnitudes through the ratio of two discrete magnitudes.

In the contemporary scientific literature on adults' judgment and decision making Kirkpatrick and Epstein (1992) were the first to describe the ratio bias as a systematic pattern of biased responses. They defined the ratio bias as the tendency to judge a low probability event as more likely when presented as a large-numbered ratio, such as 10/100, than as a smaller numbered but equivalent ratio, such as 1/10 (Kirkpatrick & Epstein, 1992). As we will explain in detail, the ratio bias is observed in an experimental paradigm called the *two-urn-choice task* that Piaget and Inhelder (1975) employed to investigate the development of probabilistic thinking. The contemporary literature on adults' judgment and decision making rarely notes that the ratio bias phenomenon has been deeply investigated in the developmental field where it is viewed as the tendency to focus on relative magnitudes rather than on the correct proportion of elements presented. , Some of the fallacies regarding fractions and proportions that have recently been studied with adults are the same as those discovered in children long before the 1990s. Developmental studies of probability judgments help to make sense of some surprising results regarding adults' conceptions of probability (Reyna & Brainerd, 1994).

According to dual-process theories, the ratio-bias phenomenon is a prominent example of weakness in adults' reasoning (Pacini & Epstein, 1999b). The ratio bias occurs because of a tendency to rely on misleading information (i.e., absolute numerosity) rather than probabilistic evidence. The ratio bias is just one example of how people make judgments based on misconstruals of probability data, but our choice to investigate such phenomenon is motivated by several reasons.

First, the ratio-bias originated in an experimental paradigm called as *two-urn-choice task* and was first reported by Piaget and Inhelder (1975). Consequently, investigating the developmental trend of the ratio bias is a bridge between the past results of traditional developmental theories and recent theoretical and methodological conceptualizations of probabilistic reasoning. Second, although the ratio bias is often studied in decision making research as a systematic manifestation of irrationality (e.g., Dale, Rudski, Schwarz, & Smith, 2007), the reasoning processes that give rise to the ratio bias are unclear. Both of these points are explored further in this chapter.

2.1 The origin of the ratio bias: developmental studies of the two-urn-choice task

2.1.1 The Piagetian period (Piaget & Inhelder, 1975)

Historically, the concepts of ratios and proportions have been widely studied, with initial research undertaken by Piaget and Inhelder in the field of probability (Piaget & Inhelder, 1975). Their account of the development of probability judgment dominated theorizing for many years. According to the Piagetian account, conceptions of probability develop from a preoperational stage (4 to 7 years) in which chance and non-chance events cannot be distinguished and children do not make comparisons based on quantified relationships, through a subsequent concrete operational stage (8 to 11 years) in which such events can be distinguished, to an ultimate stage of formal operations (beyond 11 years) in which the mathematics of probability is understood and fractions can be calculated. Thus, development progresses from an initial awareness that events can be causally determined; that awareness sets the stage for the analysis of events that are not determined during middle childhood; and the logical quantification of probabilities becomes perfected in formal operations.

However, as some researchers historically argued (Offenbach, Gruen, & Caskey, 1984), the difference between *proportionality* and *probability* should be taken into account. Generally, proportional reasoning is assumed to involve the understanding of ratios, or the comparisons of two ratios; instead, the term *probabilistic judgment* is used to refer to a higher-level judgment that presumes that an understanding of proportions is necessary, but not sufficient, for an understanding of probability. Piaget and Inhelder (1975), for example, presume that proportionality and probability schemes are closely related and that an integration of proportionality and chance schemes is necessary for a complete understanding of probability. Moreover, children are incapable of proportional reasoning until about 11 years of age; proportional reasoning involves understanding the *relation between relations*, and is a hallmark of formal operations.

One the research procedure used to investigate the probabilistic judgment is the *one-container-choice task*. A child is first presented with one mixture of elements (e.g. 8 yellow, 4 red, 2 green, and 1 blue) on a table. Then, the child is asked to predict the object membership of an element or of a couple of elements drawn at random from a bag (Figure 6).

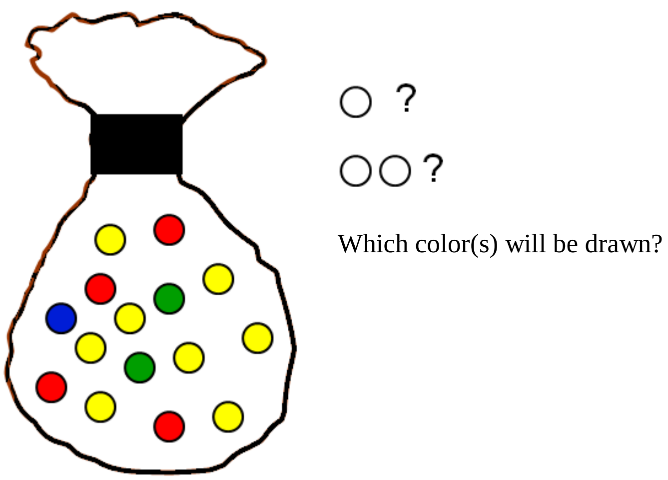

Figure 6. One-sample probability judgment task

Piaget and his colleagues observed that older children managed frequency information better than younger children. As Reyna and Brainerd (1994) explain in their review of the origins of probability judgments, children's difficulties were considered to be logical rather than computational. Preoperational children failed to understand frequency information because of their inability to differentiate probabilistic from nonprobabilistic causes. Children based their preference on idiosyncratic judgments such as favorite color or according to the absolute number of winning elements when they were asked to say in which urn the probability of the designated winning color was greater. For example, Piaget and Inhelder (1975) observed that one child concluded that he had a better chance of drawing a target item when it was one of six rather than one of three "because there are more " (p. 135).

Preoperational children's judgments are set in a "world of perceptive and subjective intuitions" (p.136). Concrete operational children, instead, are able to exploit frequency information but fail to integrate outcomes into complex relations. Children in the concrete operational stage can, however, differentiate between certainty and uncertainty and are beginning to quantify probabilities in restricted situations. Finally, formal operational children are supposed to perfectly understand complexity and the multiplicative relationships between rational quantities ($a/b = c/d$), which is fundamental to manage ratios and proportions. As Jones and Thornton (2005) explain, Piaget and Inhelder's claim about the need for formal operations in dealing with probability was a powerful deterrent limiting the study of probability to high school and college mathematics for more than three decades (Jones & Thornton, 2005, p. 69).

According to the Piagetian traditional theory, cognitive development is a unidirectional progression either from intuitive thinking to logical, scientific reasoning or from an initially inefficient state to a state of greater efficiency. As Jacobs and Klaczynski (2002) pointed out, recent perspectives of cognitive development (Kokis, et al., 2003; Stanovich & West, 2000) follow the same core assumption that development is a progression from states of limited understanding and complexity to more advanced understanding, computational complexity, and abstraction.

Although Piaget and Inhelder provided interesting speculations about the cognitive roots of probability judgments, their conclusion that such judgments emerge late in development has been the focus of subsequent research that revised Piaget's methods and paid more attention to the link between experimental manipulations and specific hypotheses. For example, Piaget's methods were too qualitative to address his research questions adequately, and children's isolated nonverbal responses were interpreted based on how children talked about them. In Piaget's studies, no hypotheses were made or data collected that could be statistically tested. A peculiar feature of Piagetian research, one that was noted by Gopnik (1996), is that Piaget's goal was not to explain why children change, but to determine what those changes could tell us about the origins of knowledge (p. 223). To deeply understand what Reyna and Brainerd (1994) called the "details" of the Piagetian program and how information processing abilities develop, later researchers started to systematically vary the nature of the subject's response (e.g., verbal versus non verbal), the task and the problem.

2.1.2 From logic to competence: information-processing rules and strategy categories

During the *neo-Piagetian* period, researchers generally shared Piaget's view of development but addressed the problem of Piaget's characterization of a child as being in a particular stage, because not all children act in accordance with their specific stage. The particular materials, task, and instructions appear to influence children's performance.

One of Piaget's assumptions is that cognitive development is a slow process (Brainerd, 1979), and that is a possible reason why Piaget used his *clinical method*, although it is also possible that Piaget considered this method the best one to investigate his children and their cognitive processes. Braine (1959) argued that these procedures resulted in high levels of false negative errors and, consequently, simplified tests may produce positive evidence of the focal concept in younger children.

This idea was first investigated by Yost, Siegel and Andrews (1962) through a decision-making method that was developed for assessing understanding of probabilities. They used a paired-comparison procedure similar to the two-urn-choice-task to compare Piaget's methods to more direct tests of underlying competence. Table 4 shows how the Piagetian procedure and Yost and colleagues' procedure differ.

Table 4
Description of Experimental Conditions

Piagetian condition	*Decision-making condition*
modified Piaget method	decision making technique
one box	two boxes
color choice	container choice
no control for color preference	control for color preference
no reinforcement other than knowledge of outcome	reinforcement

Adapted from Yost, Siegel & Andrews (1962)

Specifically, they tested the hypothesis that children observed in a probabilistic situation that utilizes a decision-making technique exhibit an understanding of probability greater than that exhibited by a control group of participants observed with a technique similar to Piaget's (p. 773). They administered the probabilistic decision-making procedure to 20 preschool children. At the beginning, three paired comparisons of sample chips were presented to determine children's most and least preferred colors. After a subject's preference had been determined, an equal number of most preferred chips and an unequal number of least preferred chips were put in two transparent containers. Both containers were shaken to randomize the position of the chips, and the child had to choose which container he would want reach into (without looking) to get a chip of a specific color. The subject's choice was classified as correct if he chose the container with the greatest proportion of the target color chips (although tangible rewards were given only for actually drawing a chip of the specified color from the correct container).

Yost and colleagues found that children exhibited a greater understanding of probability in the decision making procedure than in the Piagetian procedure. Specifically, 75% of children responded correctly; they concluded that "4-year-olds do have some understanding of probability" (p. 779) and this result contrasted with the Piagetian view that a child under seven years is not able to respond consistently to the quantitative proportions of elements.

Other authors (e.g., Hoemann & Ross, 1971) found, instead, that four-year-olds have greater difficulties making probability judgments than magnitude judgments in a two-sample spinner task. They concluded that the precocious competence young children showed in probability judgments of Yost and colleagues might be due to estimations of magnitudes rather than a real understanding of probabilities. They argued that researchers may have overestimated children's true probability judgment competence.

Chapman (1975) criticized Yost and colleagues' study, claiming that the specific items they employed required only that the child attend to the relative number of one class of objects in the two containers (p. 141) and it is possible that children solved the comparison without comparing ratios. To investigate this possibility, Chapman (1975) systematically varied the numbers of elements and the difference in the proportions to be evaluated. For example, each container in the two-container condition contained a mixture of brown and yellow M&Ms. The container with the greater proportion of the target color (brown) contained: a) more brown M&Ms (2B2Y versus 1B2Y), b) an equal number of brown M&Ms (2B2Y versus 2B1Y), c) or fewer brown M&Ms (2B1Y versus 3B3Y) than the other container. These three conditions were labeled type 1, 2 and 3 items. The probability of drawing one brown M&Ms differed for each of the three comparisons; specifically, if children base their choice on only the quantity of the target color rather than the correct proportion, then they would commit errors on type 3 items. First graders, third graders, fifth graders and college students were tested, and children performed significantly better on type 1 items than type 2 items and significantly better on type 2 items than type 3 items. Overall, first graders and third graders gave a high proportion of correct responses on type 1 and 2 items replicating the result that Yost and colleagues (1962) obtained. But a more valid measure of children's failure to comprehend proportions was the difficulty that first graders and third graders experienced on type 3 items. Interestingly, Chapman also excluded the possibility that children failed the most difficult items because of an inability to calculate proportions, and he concluded, as did Piaget and Inhelder, that children are not capable of evaluating proportions until formal operations have been attained.

Neo-Piagetian theorists also shifted toward information-processing explanations of age variability. The information processing approach focused on cognitive change rather than stages of development, and a host of new constructs became available to explain variations in children's performance: perceptual processes, working-memory capacity, executive functions, scripts, and so on. Human mind is considered a complex cognitive system, analogous in some ways to the operations of a computer; this *mind-as-computer* metaphor allowed researchers to re-think the human mind as a system that manipulates or processes information coming in from the environment or information already stored within the system. In addition, information-processing developmentalists focused on cognitive activities much more than on cognitive structures, and limitations in information processing were generally viewed as *non-conceptual* deficits (Reyna & Brainerd, 1994).

As Flavell, Miller and Miller (2002) explained, the information-processing approach tried to provide an explicit, testable, detailed understanding of what a child's cognitive system actually *does* when dealing with some task or problem (p. 13). Moreover, information-processing developmentalists questioned whether different reasoning steps work in parallel or serially, and such terminology (i.e. system or processing) also plays a role in contemporary dual-process accounts.

Information-processing researchers reanalyzed various Piagetian concepts. For example, Siegler (1996) investigated the developmental sequence in which children acquire various rules of reasoning in Piaget's balance-scale task, but he also tested the development of probabilistic reasoning using the two-urn-choice approach (Siegler, 1981).

The basic assumption underlying Siegler's (1981) so-called rule assessment approach is that cognitive development is characterized as the acquisition of increasingly powerful rules for solving problems. Siegler hypothesized that children would apply the following rules:

- Rule I: the urn with the greater number of favorable colors is chosen;
- Rule II: with an equal number of favorable and unfavorable colors, the urn with the smaller number of unfavorable events is selected;
- Rule III: the difference between the number of favorable and unfavorable events is calculated for each urn and the one with the greater difference is selected;
- Rule IV: the ratio between favorable and unfavorable events is the basis for the choice.

He expected that there would be changes in rule usage as cognitive development progressed, although the rule applied would also be influenced by task structure. Siegler (1981) tested these rules against the responses of 3- to 20-year-olds to six two-urn-choice tasks. The sample spaces in the pairs of urns were designed in such a way that responses could be linked to each of the four rules. Although the expected applications of Rule I and IV were confirmed, response patterns were not expected for Rules II and III. For example, preschoolers used only Rule I, but the 8 year-olds tended to skip Rule II and use mostly Rule IV and sometimes Rule III. It appears that the hierarchy of rules is not applicable to this probability task although it was applicable to a weight-on-balance-beams task. Overall, most 3-year-olds do not respond in accord with any apparent rule; most 5-year-olds predict that they are more likely to get a marble of the desired color from the set with the larger number of marbles of that color, regardless of the number of undesired color marbles; and most children 8 years and older consider both the number of desired and the number of undesired color marbles and answer correctly (Siegler, 1981).

At the same time, and contrary to Piagetian theory, other researchers (Brainerd, 1981) showed that young children (4- and 5-years-olds) also processed frequency information. He concluded that children lacked sufficient working memory capacity to accommodate frequency information as well as the other information that they maintained about problems. Many studies appeared to demonstrate that frequency information was critical for probability judgments.

Thus, two viewpoints emerged from this research (Reyna & Brainerd, 1994): probability judgments are based on simple magnitude estimation (for children lacking logical competence) and, alternatively, that frequency processing is limited by working memory and retrieval failure. This conflict continued in a third phase of research we call the intuitive process period.

2.1.3 The intuitive process period

The period of the 1970s and 1980s saw a continuation of Piaget's work with a strong interest in the nature of probabilistic conceptions. Other psychologists were concerned with the heuristics or strategies that people use to make probabilistic judgments. Older children and adults exhibit a number of systematic errors in reasoning that place their competence in question. Contrary to the view that cognitive development proceeds in a linear fashion, an increasing number of studies showed that varying tasks and experimental conditions lead to different patterns of results, and several researchers introduced a distinction between quantitative processing that involves explicit or analytic calculation and intuitive processing or estimation (Dixon & Moore, 1996).

In an ingenious study, Acredolo, O'Connor, Banks and Horobin (1989) presented a clear plastic bag to children containing different colored jellybeans. In one condition, for example, the bag contained three red jellybeans, two green jellybeans, and one yellow jellybean. They asked fifth graders, third graders and fifth graders to indicate the likelihood of randomly drawing a yellow jellybean by sliding a marker along a scale with a sad face at one end and a happy face at the other. They demonstrated that first graders, third graders and fifth graders can take into account variations in the numerator and in the denominator, as well as the interaction of these two dimensions. In particular, they demonstrated that the error of assigning a higher probability to a display having a larger numerator but a smaller probability was prevalent when the probability ratio was the same (i.e., by preferring 5/10 rather than 3/6). Instead, when the two probabilities were different but close (i.e. 5/8 versus 4/6), there was no significant tendency to assign higher estimates to either the display having the larger numerator or to the display having the higher probability. Finally, when the two probabilities were very discrepant, children reliably assigned a higher estimate to the display having the higher probability.

Thus, children assigned a significantly higher estimate, for example, to 3/6 than 4/10. They concluded that children may lack the ability to generate precise mathematical computations but they nevertheless possess the capacity to approach an accurate solution by appropriately integrating cues to generate rough estimates (pp. 944). Their finding that children do rather well when tasks involve graded judgments of probabilities was replicated by other researchers (e.g., Anderson & Schlottmann, 1991, Wilkening & Anderson, 1991). The sliding scale response method may enable young children to use intuitive problem solving strategies that are more likely to be correct than more explicit strategies (Boyer, 2007; Falk & Wilkening, 1998; Reyna & Brainerd, 1994; Schlottmann, 2001).

Fischbein's contribution to the development of intuition and to dual-process theories

Fischbein, Pampu and Manzat (1970) investigated the evolution of probabilistic behavior including the ages of the three major Piagetian stages: the preoperational, the concrete operational, and the formal operational. Preschool children, third graders and sixth graders were tested in 18 trials of the two-urn-choice task with three categories of ratios presented. In six of the 18 trials the correct response was to point to the box on the right side; in another six, point to the box on the left side; and in the final six, state their equivalence. Results revealed that the number of correct responses increased from younger to older children. Specifically, the most important developmental changes were revealed in the category of trials with equal ratios (e.g., a comparison of 1-in-2 and 2-in-4), and the authors concluded that these trials may be considered indicators of the level of intellectual development. Similarly, Goldberg (1966) found that preschool children committed more errors on two-container items in which the correct container had an equal rather than greater number of target color. Furthermore, Carlson (1970) reported that second graders had difficulty managing an equal number of target chips.

Participants were also asked to explain their responses, and the authors identified participants as belonging to one of three types of solution procedures (p. 385): a) Choice was based on a simple binary comparison, where the comparison is reduced to one ratio, although not necessarily using numbers (for instance, "I chose this box because there are more black marbles here than in the other box"); b) Transitional stage in which the reasoning took into account all four terms, but the numerical ratios were not indicated (for instance, "Here it is easier to pick a white marble for there are more white than black marbles as compared with the other box"); and c) Correct comparison of the numerical ratios in which both ratios are considered, and their numerical relations are set forth explicitly (for instance, "In this box there are three times and in the other two times more black than white marbles").

Preschool children based their judgments primarily on simple binary relations, whereas most sixth graders based their decisions on explicit correct estimations (relations between ratios). The finding that 9- to 10-year-old children, after brief instruction, were able to perform chance estimates by comparing numerical ratios and to understand the concept of proportionality, caused Fischbein and colleagues to hypothesize that instruction might be able to set up structures corresponding to formal operations as early as the concrete operational stage with much greater ease and more stability than would be the case for the transition from the preoperational to the operational stage.

Although Fischbein's early work was predicated on Piaget's research on the development of children's probabilistic thinking, he was from the outset concerned about the way that children responded in instructional settings. Contrary to Piaget, Fischbein was the first to claim that intuitions of probability and chance are basic adaptive tools for living organisms and that even young children have functional probabilistic intuitions. This interest in both development and teaching led to his seminal work on primary and secondary probabilistic intuitions (Fischbein, 1975). This distinction is based on the origin of these kinds of intuitions; primary intuitions may be either pre-operational or operational and refer to those cognitive beliefs that develop in individuals on the basis of normal everyday experience and independently of any systematic instruction as an effect of their personal experience (Fischbein, 1987, p. 64). An example of primary intuition is the intuition of causality. For instance, Piaget (1930) asked: “Why do the clouds move more or less quickly?” The child answered: “Because of the wind. They move along by the wind". Another example is that increasing the number of conditions imposed on an expected event diminishes its chances.

Secondary intuitions, by contrast, are those that are acquired through some educational intervention and mainly in school (p. 71). Such intuitions are not produced by the natural, normal experience of an individual. For example, the statement "the sum of the angles of a triangle is 180°" is not self-evident. We learn that this is true but we do not learn it from experience. Primary and secondary intuitions are relative concepts because they depend on the cultural environment of the individual. Moreover, both primary intuitions and secondary intuitions are not innate; they are learned cognitive capacities in the sense that they are always the product of ample practice in some field of activity.

According to Fischbein (1975), it is important to emphasize that new, correct intuitions do not simply replace primitive, incorrect ones. Primary intuitions are usually so resistant that they may coexist with new, superior, scientifically acceptable ones. Moreover, very often secondary intuitions contradict the natural attitude towards the same question. That situation can generate inconsistencies in the student's reactions depending on the nature of the problem. A student may understand logically and intuitively that when tossing a coin several times, each outcome has the same probability. Nevertheless he may still feel intuitively, that, after getting *tails* 3 to 4 times in succession, there is a greater likelihood of getting *heads* on the next toss.

It is interesting to note that this *interaction* between the two types of intuitions generates what dual-process theories in more recent decades called cognitive conflict between two opposite reasoning processes: the heuristic process and the analytic process. In another relevant example that Fischbein considers the following question (p. 14): "Two liters of juice cost 3 dollars. What is the price for 4 liters?" Contrary to the Piagetian approach, Fischbein explains that the intuitive, correct, answer is 6 dollars. Instead, consider the question, "One liter of juice costs 2 dollars. What is the price to be paid for 0.75 liters?" The required multiplication in this case is not an intuitive (direct, global) solution. He concluded that the property of globality (or self-evidence) serves to distinguish intuitive and analytical thinking where the analytic thinking consists, as in this example, of making calculations. Fischbein's view of intuitive cognition as characterized by self-evidence, extrapolativeness, coerciveness, and globality resembles the FTT dual-process theory (Reyna & Brainerd, 1995) and its view of gist processing. It is noteworthy that some of the most prominent and recent theories in reasoning, judgment and decision making share concepts and terminology that come from but rarely mention developmental and educational psychology.

Finally, Fischbein (1975, p. 204) explains which mechanisms participate in the process of generating intuitions and describes those factors that are automatically elicitated to contribute to the intuitions' effect of immediacy. Among these factors, he deals with the notions of: a) *availability*, which happens when one tends to consider a certain element as being more frequent if it belongs to a class that can be more easily detected than others; b) *anchoring*, which happens when a certain salient feature may become decisive in an individual's intuitive interpretation not because it is objectively decisive but merely because it is more salient; and c) *representativeness*, which happens when the probability of an uncertain event or sample belonging to a class may be determined not by objective considerations but by its superficial, apparent representativeness of the respective category. For example, the sequence ABABB is considered intuitively to be more representative of a random process than, for instance, the sequence AAAAA, and therefore more likely to appear. Availability, anchoring and representativeness are better known as cognitive shortcuts or heuristics in judgment and decision making (Tversky & Kahneman, 1974) but, as Fischbein underlines, such heuristics account for systematic biases in probabilistic judgments as well as in intuitive judgments.

Building on Fischbein's work, Stavy and Tirosh (2000) developed the intuitive rules theory; intuitive rules are self-evident and global, and people strongly rely on them, often disregarding their formal education. According to intuitive rules theory, students' responses when solving a wide variety of conceptually non-related mathematical and scientific tasks that share some common, salient, external features are affected by a small number of intuitive rules (Babai, Brecher, Stavy & Tirosh, 2006). Stavy and Tirosh (2000) defined three intuitive rules: two relate to comparison tasks (*more A-more B* and *same A-same B*), and one to subdivision tasks (*everything can be divided endlessly*). Responses in line with the rules are often correct. But sometimes they are in conflict with concepts and reasoning in mathematics and science, and then may lead to incorrect responses.

Babai and colleagues (2006) investigated the first intuitive rule (*more A-more B*) in a probability task, the two-urn-choice task. In this task the intuitive rule *more A-more B* means that participants show a systematic preference for the urn with the greater absolute number of elements inside. This choice is considered congruent when the probability ratio for the urn with more elements is greater than the probability ratio for the urn with fewer elements, whereas the choice is considered incongruent when the urn that contains more elements is not the one with the higher probability for picking the target element. According to intuitive rules theory, responses to congruent trials should be more accurate than responses to incongruent trials.

Sixty-one 16- to 17-year-olds were tested in 20 congruent and 20 incongruent probability trials and, in order to understand the cognitive processes related to the impact of the intuitive rule, participants' response times were also investigated. Babai and colleagues (2006) hypothesized that students who apply only the intuitive rule strategy should respond correctly to the congruent trials and incorrectly to the incongruent trials. In both of these cases, their responses should be relatively fast. Students who apply an appropriate strategy should, by contrast, respond correctly to all trials and their mean response times should be relatively long for both conditions.

Babai and colleagues (2006) found that students provide more accurate responses in the congruent trials, and students with a high level of mathematical instruction perform better than those with a low level of mathematical instruction in the incongruent trials. The mean response time of correct responses was significantly longer in the incongruent condition than in the congruent condition. Moreover, with respect to the incongruent condition, the response time was significantly shorter for incorrect responses than correct responses.

They interpreted their results as evidence that all participants consciously applied an appropriate strategy to solve the comparisons, but they also, at the same time, unconsciously attended to the difference in the number of target balls in the two urns. In the incongruent trials, the higher response time was interpreted as the time necessary to overcome the intuitive response, underlining the relevance of control mechanisms for successful problem solving. In other words, they detected the conflict between the two possible responses in the incongruent trials and spent more time to inhibit the intuitive response (De Neys, 2006).

2.2 The *Ratio-bias* phenomenon

When reading contemporary scientific literature on the ratio bias, most of the articles we reviewed (i.e., Amsel, Close, Sadler, & Klaczynski, 2009; Alonso & Berrocal, 2003; Bonner & Newell, 2008; Ferreira, Garcia-Marques, Sherman, & Sherman, 2006; Price & Matthews, 2009) describe the ratio bias as *the tendency for people to judge a low probability event as more likely when presented as a large numbered ratio, such as 10 in 100, than as a smaller numbered but equivalent ratio, such as 1 in 10* (Kirkpatrick & Epstein, 1992). Some authors refer to this phenomenon as a systematic manifestation of irrationality (Dale, Rudski, Schwarz, & Smith, 2007). Other researchers refer to the same phenomenon as *denominator neglect* (Klaczynski, 2001b; Kokis, Macpherson, Toplak, West, & Stanovich, 2003; Stanovich Toplak, & West, 2008) although denominator neglect is an alternative explanation (Brainerd & Reyna, 1990b; Reyna, 1991) for the ratio bias provided by FTT (discussed in more detail later). The ratio bias task introduced by Kirkpatrick and Epstein (1992) is an experimental paradigm that has been investigated in several modified versions and is related to individual differences such as numeracy and need for cognition (Pacini & Epstein, 1999b; Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2003). Moreover, the bias toward large-numbered ratios in low probability events is a robust effect that has been replicated many times.

We will argue, however, that the ratio bias task has been modified by researchers during the last two decades from its original formulation as a probabilistic judgment task (i.e., Alonso & Berrocal, 2003; Bonner & Newell, 2008; Ferreira, Garcia-Marques, Sherman, & Sherman, 2006; Morsanyi, Primi, Chiesi, & Handley, 2009).

Kirkpatrick and Epstein (1992) tested whether CEST could account for some findings reported by proponents of norm theory (NT, Kahneman & Miller, 1986). Specifically, Kirkpatrick and Epstein (1992) used an experimental paradigm adapted from a study by Miller, Turnbull and McFarland (1989) in which subjects had been shown to regard certain outcomes as more likely than others despite the equality of their objective probabilities. As we will explain, the ratio bias, which Denes-Raj and Epstein identified for the first time in 1994, refers also to high probability events. In the next section, we will present the original studies on ratio bias with particular attention to high probability events, which are the topic of this dissertation.

2.2.1 The original studies of the ratio bias by CEST researchers

According to norm theory (NT; Kahneman & Miller, 1986), when judging equal low-probability events with ratios differently numbered (for example 1-in-10 versus 10-in-100), it is necessary to distinguish between pre-computed judgments of probability and post-computed judgments of normality. It is assumed that subject's pre-computed probability judgments are identical for the two events. After learning the outcome, instead, people engage in post-outcome counterfactual thinking and imagine other ways in which the outcome might have occurred. The more the event is abnormal, the more it generates surprise and suspicion; consequently, the outcomes of two equal probability events are judged as differently as possible despite their objective probabilities being identical.

In one study, Miller, Turnbull, and McFarland (1989) presented a vignette to undergraduates in which a child successfully draws a preferred cookie from a jar containing either 1 preferred cookie and 19 non preferred ones, or 10 preferred cookies and 190 non preferred ones. After participants read the scenario, they were asked to indicate how suspicious they would be that the child peeked into the jar before selecting the cookie. Participants reported that they were suspicious that the child had cheated (by peeking) more in the former condition (1 out of 20) than in the latter condition (10 out of 200) even though the probabilities associated with each scenario were identical. Miller and colleagues (1989) reported similar findings for a variety of other vignettes and they concluded that, if norms are computed only after the event, as they presume, then the higher level of suspicion for one jar can occur only in post-outcome judgments. They ruled out a subjective-probability explanation by using a control condition that tested for such an effect and found that it did not occur. Miller and colleagues (1989) concluded that the results could not be explained in terms of differences in pre-computed subjective probabilities.

Kirkpatrick and Epstein (1992) argue that, according to the CEST perspective, it is possible to entertain two different estimates of probability at the same time. For example, a person might understand objective probability rationally but nevertheless feel, at the same time and experientially, that the odds described by one ratio are more favorable than the other ratio. According to CEST, there are two principles of experiential system, the concretive and experiential principles.

The concretive principle causes the experiential system to encode events primarily in the form of concrete representations and is, accordingly, particularly responsive to such representations. Since absolute numbers are more concrete than ratios, it follows that people are more responsive to frequencies than to ratios. Moreover, ratios between small numbers are easier to comprehend than ratios between large numbers; small numbers are more concrete in the sense that they are easier to visualize (Paivio, 1991) and they are well articulated in memory (Holyoak, 1978). For example, people can represent 1 versus 10 items in memory fairly accurately, but the same is not true for 10 versus 100 (Kirkpatrick & Epstein, 1992, p. 536). According to CEST, the subjective probability of a low probability event should be judged as greater when the ratio is composed of larger numbers.

According to the experiential learning principle, people often experience the phrase “1 in X odds” in everyday life and this phrase is generally understood to mean “unlikely”. As a consequence, and particularly in the case of past significant personal experiences, the subjective probability of a 1-in-10 outcome is perceived as smaller than that of 10-in-100 outcome. To test NT and CEST predictions, Kirkpatrick and Epstein (1992) showed participants the following vignette:

> Imagine that you are presented with two bowls of folded tickets. One bowl contains 1 ticket marked "winner" and 9 blank tickets. The other bowl contains 10 tickets marked "winner" and 90 blank tickets. You must draw one ticket (without peeking, of course) from either bowl: if you draw a ticket marked "winner" you win $8.00, otherwise you win nothing and the game is over.

Participants were asked to say which bowl they would choose from. This vignette represents a low probability event, but Kirkpatrick and Epstein (1992) varied the probability levels, and participants were also presented with a vignette describing a high probability event; the scenario was identical for the 90% win condition except that the respective bowls were described as containing (a) 9 winners and 1 blank, and (b) 90 winners and 10 blanks. With regards to the 90% win condition, contrary to the modified NT which assumes that people should prefer the large bowl because it contains more winners, Kirkpatrick and Epstein, according to CEST, hypothesized the following: *subjects should favor the small bowl because 1-in-10 describes the odds of drawing a nonwinner from the bowl, an outcome that is subjectively regarded as highly unlikely according to the concretive and experiential learning principles* (p. 536).

Finally, Kirkpatrick and Epstein (1992) varied also the valence of the outcome and two additional conditions were constructed in which winning tickets were replaced by losing tickets. In this dissertation, we will not mention the losing frame anymore because it is beyond our purpose.

Table 5 summarizes predictions, experiments and results both for the low-probability event and the high-probability event. As Table 5 shows, results differed slightly from predictions. In Experiment 1, the authors explain the weakness of results with a failure of the task to adequately tap the experiential system. Instead, according to CEST predictions, the other experiments worked to engage subjects' experiential system in the low-probability winning frame, in particular in the real-life situations in which the bowls were real and participants could win or lose real money.

Kirkpatrick and Epstein (1992) interpreted their results as strong support for CEST theory, observing that, contrary to NT, participants also preferred the small bowl in the 90% win condition (p. 542) and they made choices that, by their own account, they recognized were irrational. They knew their behavior was irrational, but they nevertheless felt something different, even though the ratios were the same. This is consistent with the view that the experiential system, as a relatively concretive system, is responsive to perceptual phenomena (figure-ground relationships) and is more responsive to absolute numbers than to ratio (p. 543).

Table 5

Summary of Predictions, Experiments and Results of Kirkpatrick & Epstein's (1992) Study

Frame	Predictions	Experiments	Results
Low probability event (Win condition 10%)	**The large bowl** • The subjective probability is greater • 1-in-10 outcome perceived as smaller	Experiment 1 • stimulus material in a written form • subjects' own bowl choices	*no significant difference* between the large (49.4%) and the small (51.6%) bowl
		Experiment 2 • stimulus materials in a written format • other's choice perspective	*significant difference* between the large (64.1%) and the small (35.9%) bowl
		Experiment 3a • Real life situation (real bowls) • subjects' own bowl choices	*significant difference* between the large (76.9%) and the small (23.1%) bowl
		Experiment 3b • Real life situation (real bowls) • other's choice perspective	*significant difference* between the large (94.2%) and the small (5.8%) bowl
High probability event (Win condition 90%)	**The small bowl** • The probability of drawing a nonwinner is 1 in 10 which is considered as highly unlikely	Experiment 1 • stimulus material in a written form • subjects' own bowl choices	*no significant difference* between the large (47.5%) and the small (52.5%) bowl
		Experiment 2 • stimulus materials in a written format • other's choice perspective	*no significant difference* between the large (56.3%) and the small (43.7%) bowl
		Experiment 3a • Real life situation (real bowls) • subjects' own bowl choices	*significant difference* between the large (36.5%) and the small (63.5%) bowl
		Experiment 3b • Real life situation (real bowls) • other's choice perspective	*no significant difference* between the large (51.9%) and the small (48.1%) bowl

We argue, however, that the difference of preference between the small bowl and the large bowl in the 90% win condition was not statistically significant in three experiments out of four and these results were not properly interpreted. Instead, the preference for the small bowl in the self-perspective condition is also coherent with recent studies showing that people, like bees, rely on relative small samples in their experience-based choices (Hertwig & Pleskac, 2010). Small samples make choices easier even though less accurate.

Denes-Raj and Epstein (1994) extended Kirkpatrick and Epstein's study (1992) to understand how extreme people's irrational behavior can be. Specifically, they investigated irrational decisions in the low-probability frame by asking participants to draw from a bowl that they could recognize as offering less favorable objective probabilities than an alternative bowl. Participants selected probabilities of 9, 8, 7, 6 and even 5% (presented as 9-in-100, 8-in-100, 7-in-100, 6-in-100, and 5-in-100 respectively), in preference to 10% (presented as 1-in-10).

Denes-Raj and Epstein (1994) found that a considerable majority (82%) of participants made one or more non-optimal choices on the win trials; moreover, 8% of the subjects made non-optimal responses on all five win trials. Participants showed a considerable preference for the large bowl: for example, almost half of participants preferred the large bowl containing 8-in-100 (8% probability of winning) to the small bowl (10%). Furthermore, 40% of participants preferred a 7% over a 10% probability. As in Kirkpatrick and Epstein's research (1992), participants experienced a conscious conflict between the normative response and the biased response, and Denes-Raj and Epstein (1994) concluded that the experiential system can override the rational system even when people are aware of normative rules (p. 823). These two systems *pull* for different solutions. This study adds strong empirical evidence to the strength of the ratio bias effect with low probability events, which was the result emphasized by Kirkpatrick and Epstein (1992).

What happens with high probability events remained an open question, however, until Pacini and Epstein (1999b) investigated the ratio-bias phenomenon varying the frame (win or lose), the choice perspective (self-perspective or others-perspective) and the probability level (low or high) as Kirkpatrick and Epstein (1992) had done. Participants were shown five different trials in which two trays had the same proportion of target items (10%, 30%, 50%, 70%, or 90%) but differed in absolute numbers, similarly to Denes-Raj and Epstein (1994). There were always 10 target items in the small tray and 100 in the large tray.

According to Kirkpatrick and Epstein (1992) and to the relative CEST principles that the authors explained, one might expect that participants would prefer the small tray when comparing 9-in-10 to 90-in-100. Pacini and Epstein (1999) explained that participants should prefer the large tray (90-in-100). According to the first facet of the concretive principle (the experiential system is more responsive to numerosity rather than ratios) preference for the large tray should increase as probability-ratios increase (p. 308). However, Kirkpatrick and Epstein (1992) found no ratio bias at all in the 90% win condition. To explain this null-result, Pacini and Epstein (1999) called for a second facet of concretive thinking: the small-numbered effect (the experiential system comprehends small numbers better than large numbers) as also Kirkpatrick and Epstein (1992) did. The combined effect of these two facets of the concretive principle when operating simultaneously is the same as predicted by Kirkpatrick and Epstein (1992) in low-probability win conditions (10%): people should have a preference for the large tray because the two facets work in the same direction. In high-probability win conditions (90%), however, the combined effect of these two facets should yield behavior that contradicts the predictions of Kirkpatrick and Epstein (1992). The ratio bias should decrease or disappear because the two facets work in opposition to each other (Pacini & Epstein, 1999, p. 308). As they explained, the numerosity effect pushes toward the large tray whereas the small-numbered effect favors the small tray. Finally, Pacini and Epstein (1999) hypothesized a gradient of the ratio-bias effect which means, for example, that in the win condition, the ratio bias should be strong in the 10% condition and become progressively weaker along the increasing probability dimension. In the 90% condition the ratio bias should disappear and participants should have no preference. Table 6 summarizes predictions and results of Pacini and Epstein's (1999) study in win condition.

As Table 6 displays, predictions and results in the others-perspective condition match. In the self-perspective condition, contrary to Pacini and Epstein's predictions but nevertheless consistent with Kirkpatrick and Epstein's (1992) predictions and results in their Experiment 3a (see Table 5), participants prefer the small tray more than the large tray. Pacini and Epstein (1999) did not explain this effect and stated that *because the effect is small, was not predicted, and did not occur in the others-perspective responses, it should not be taken seriously until replicated* (p. 317). We argue that two different studies (Kirkpatrick & Epstein, 1992; Pacini & Epstein, 1999) that obtained the same result are a good reason to take this result seriously. Moreover, as Reyna and Brainerd

(2008) pointed out, the predictions for high probability problems are still somewhat mixed.

Table 6

Predictions and Results of Pacini and Epstein's (1999) Study in the Win Condition

	Predictions	Results
	10% → 90%	10% → 90%
Others-perspective response	From a strong ratio bias (preference for the large tray) in the 10% probability condition to no preference, or very weak preference for either tray, in the 90% probability condition	• The preference for the large tray (ratio bias) is greatest in the 10% probability condition and becomes progressively weaker along the increasing probability dimension; • No preference in 70% and 90% probability conditions
Self-perspective response	Ratio bias weaker than in the others-perspective condition; possibly, no ratio bias at all in any of probability conditions	• The preference for the large tray (ratio bias) much weaker than in the others-perspective condition; • Weak negative linear trend of responses; • The preference for the small tray in 90% condition is statistically significant (p. 317)

Note. The notation "10% → 90%" means the following probabilities: 10%, 30%, 50%, 70%, 90%

Other studies investigated the ratio bias only for low-probability events. For example, Denes-Raj, Epstein, and Cole (1995) replicated the ratio bias for low-probability events across a variety of conditions including (Study 3) a high-emotional-impact scenario that involved the development of AIDS in a loved one. They found that the ratio-bias phenomenon is highly general, and they replicated previous findings such as attributing a ratio-bias effect to others more than to ourselves and being prone to the ratio bias more in real-life situations than in simulated situations. Finally, the ratio-bias experimental paradigm with low-probability events has been investigated to test the CEST assumption that the experiential system responds to visual imagery in a way similar to the way it does to real experience. Epstein and Pacini (2001) presented two versions of the ratio bias in the 10% winning condition. One version consisted of a real-life situation with monetary payoffs; participants were trained to vividly visualize the situation. The other control version was a simulated situation in the form of a verbal description without payoffs. They found that the visual-imaging group exhibited the ratio bias both in the others-perspective condition and in the self-perspective condition. Instead, the control group exhibited the ratio bias only in the others-perspective condition. Interestingly, there was no ratio bias (as predicted) in a third experimental condition, which Epstein and Pacini introduced for the first time. The condition was to say how a completely logical person would choose (logical-perspective). The absence of ratio-bias effect follows the CEST assumption that most people are aware of the rational way of responding.

As we mentioned previously, the tendency to judge a 10-in-100 probability as more probable than 1-in-10 (ratio-phenomenon) is often described in terms of denominator neglect. Denominator neglect (Reyna & Brainerd, 1991) basically explains this tendency as consequence of an exclusive focus on the size of numerators rather than comparing ratios. For example, consider the task of determining which of two lotteries is most likely to yield a winner. In Lottery A, one winning ticket will be pulled from a total pool of 10 tickets. In Lottery B, there are 10 winning tickets in a pool of 100 tickets. To neglect the denominator, according to FTT, means to rely on the low-level gist representation “compare number of winners” rather than on the more precise “compare ratios”. As a consequence, an option with 10 winning tickets better than an option with 1 winning ticket.

Ratio bias and denominator neglect are considered synonymous in the literature on judgment and decision-making. For example, Stanovich (2008) attributes denominator neglect to Epstein and colleagues (p. 265), or Kokis, Macpherson, Toplak, West and Stanovich (2002). While presenting Klaczynski’s results (2001b), Stanovich stated that *he found that the denominator neglect that accounts for the ratio-bias phenomenon discovered by Epstein (1994; Denes-Raj & Epstein, 1994; Kirkpatrick & Epstein, 1992) was constant from early adolescence to young adulthood...*(p. 29).

We argue that there are some differences between ratio bias and denominator neglect: they are not interchangeable. The ratio-bias effect was initially investigated by researchers who proposed CEST as its explanation. Denominator neglect, on the other hand, is a theoretical concept introduced by researchers advocating FTT. Reyna and Brainerd (2008) explicitly argued that people misunderstand simple ratio and decimal representations in many decisions due to overweighting numerators and neglecting denominators. Denes-Raj and Epstein (1994, p. 827) and Denes-Raj, Epstein and Cole (1995, p. 1084) indirectly mentioned denominator neglect when they explained that, according to the concretive principle, the ratio-bias phenomenon might be due to attending to the absolute number of winning targets (i.e., comparing 1 winning target in the 1-in-10 ratio to 10 winning targets in the 10-in-100 ratio) and ignoring the ratios. However, they never called this effect denominator neglect. In addition, as described in Chapter 1, although CEST and FTT are both dual-process accounts, they nevertheless differ in many critical points and theoretical assumptions. As we described in Chapter 1, CEST and FTT differ regarding development.

CEST (Kirkpatrick & Epstein, 1992) focuses on adults and affirms that rational abilities improve with age; FTT, instead, claims that intuition and analytical competence are present early in development and improve with age. Therefore, CEST hypothesizes that errors are caused by the incapacity of the rational system to override the experiential one because of the characteristics of the task, the lack of a normative response, lack of time to reflect, or the need of specific training (Agnoli, 1991; Agnoli & Krantz, 1989; Jacobs & Klaczynski, 2002; Klaczynski, 2001; Kuhn, 2000). To the contrary, FTT predicts that reasoning errors are low-level bookkeeping mistakes that are made even late in development among advanced reasoners (Brainerd & Reyna, 1995). Children rely more than adults on precise and verbatim details, whereas adults show a *fuzzy-processing preference* that is able to avoid systematic errors caused by poor verbatim memory.

Epstein and colleagues explain that participants attend to the target items or the numerator because they are more salient, *standing out as figure against ground*. Moreover, large numbers in the denominator are perceived as more abstract than small numbers and are less articulated in memory. Directing attention to the denominator requires a step back from the immediate focus on the numerator (Denes-Raj, Epstein, & Cole, 1995, p. 1084). This explanation coherently supports what happens with low probability events but it does not explain the results that Epstein and colleagues obtained with high probability events.

It is, however, undoubtedly true that CEST provided consistent evidence that people are biased when judging probability ratios and this tendency may depend on many reasons such as the situation (real or simulated), the subjective probability, the emotional impact of decisions, and some relevant individual differences in personality. We will discuss the relevance of individual differences on judgment and decision-making in a following section, but CEST attributed great relevance to the relative degree and effectiveness with which individuals use the two modes of information processing (experiential and rational). As we will present, Epstein and colleagues were the first to construct an inventory (the Rational-Experiential Inventory, REI) to assess individual differences in thinking styles with regard to both heuristic processing and analytic processing. Moreover, they applied this inventory to the ratio-bias phenomenon.

In subsequent research the ratio-bias effect has been widely investigated mainly in the low-probability condition, and, as we will describe, the original task and the relative ratio-bias phenomenon started to be considered a probability-judgment experimental paradigm that testifies on the one hand to humans' irrationality, and on the other hand to the capacity of the analytic system to override the heuristic system. Research on thinking dispositions and cognitive styles investigated the ratio-bias phenomenon to understand whether the tendency to respond normatively rather than heuristically can be predicted by individual differences.

2.2.2 Ratio bias or denominator neglect?

Proponents of FTT investigated the development of decisional processes from childhood to adulthood using the two-urn-choice-task. FTT assumes, contrary to traditional theories (Inhelder & Piaget, 1958), that reasoning develops from computational to intuitive thinking. Specifically, FTT presumes that wrong responses are linked to a particular phenomenon called *denominator neglect*. Denominator neglect is a particular effect that characterizes class-inclusion problems (Reyna & Brainerd, 2007; Reyna, Lloyd, & Brainerd, 2003), and FTT explains that the ratio bias occurs because people understand that probability is a function of frequencies in both the numerator and the denominator but still tend to pay less attention to the denominator as a default.

In class-inclusion tasks (Reyna, 1991; Reyna & Brainerd, 1990, 1993, 1994) children and adults lose track of classes, when they exist in part-whole relation to one another, and they focus on target classes in numerators, neglecting denominators. Moreover, the availability of a salient and compelling gist leads to the failure to retrieve relevant knowledge and use it to inhibit the wrong response. The *marble game task* may be considered a particular problem of class inclusion. Specifically, when children have to choose between two urns they prefer to draw a marble from the urn with more marbles of the target class even when they recognize the equivalent probabilities of the two samples (e.g., 1 out of 10 and 10 out of 100). They exhibit a telltale bias toward samples with larger numerators.

Because of children's tendency to choose the urn with the larger absolute number of elements, this is also called the numerosity effect. Because the numerosity effect can be seen as a particular phenomenon within the class-inclusion problem, it can be explained using three principles (Reyna & Brainerd, 2007). Given that the denominator and numerator represent overlapping classes and that the retrieved knowledge on the ratio (verbatim traces) is unable to inhibit the wrong response, then children focus on the salient compelling fuzzy information (the numerator), neglecting the denominator.

In *the marble game task*, therefore, even if children have a concept of probability, the confusion created by the overlap between the numerator (e.g., the number of winning marbles) and denominator (the total number of marbles in the urn) leads people to focus on salient gist (comparisons between numerators), underweighting the denominators.

For example, with highly probable events (90% winning) in which children have to choose between Urn A with 9 red (winning) marbles out of 10 and Urn B with 90 red marbles out of 100, FTT hypothesizes that children choose Urn B. This happens because numerators (9 and 90) and denominators (10 and 100) are two overlapping classes; children mistakenly report the horizontal relationship between subsets (numerators) rather than the vertical one between set and subset (between numerators and denominators).

In addition, children focus on the compelling, salient, gist information (numerators, 9 and 90), and they neglect the seemingly irrelevant information provided by the denominators (10 and 100). Even if children know the concept of ratio (the relationship between numerators and denominators), this knowledge is not able to inhibit the gist traces leading to the wrong response (the preference for the largest urn, Urn B). This is essentially what ratios are: part–whole relations.

2.3 Studies on the ratio bias with children and adolescents

Building on Epstein and colleagues' results on the *ratio bias* (Denes-Raj & Epstein, 1994; Kirkpatrick & Epstein, 1992; Pacini & Epstein, 1999a, 1999b), other researchers (Agnoli, Dellai, Furlan, & Stragà, 2007; Dellai, 2007; Babai, Brecher, Stavy, & Tirosh, 2006; Macpherson, 2001) focused on children's and adolescents' behavior to understand developing trajectories of reasoning abilities and individual differences in statistical reasoning.

Results from the *marble game task* were re-evaluated by Babai, Brecher, Stavy and Tirosh (2006) in light of their new model: intuitive rules theory. According to these authors (Babai, Brecher, Stavy, & Tirosh 2006; Osman & Stavy, 2006; Tirosh & Stavy, 1999), subjects' preference for one of the two urns was guided by the automatic application of a specific intuitive rule: *more A- more B* because *more* marbles means *more* probabilities of drawing the winning element (Stavy & Tirosh, 2001). In addition, Babai and colleagues showed the importance of control mechanisms for successful *problem solving*: evidently, the correct response requires, in addition to knowledge of an appropriate strategy, developed executive control mechanisms that overrule the impact of the intuitive interference.

Dellai (2007) conducted four experiments to clarify and expand some aspects of the *ratio bias* in children. Specifically, she wanted to understand the processes underlying decisions and they used the *marble game task.* In the first experiment, she investigated the decision-making behavior of 8- and 10-year-old children by asking them to decide from which of two containers (with equal win probabilities) they would want to draw a marble; children were presented with three response options consisting of each of the two containers (wrong responses) and the normative response, “it does not matter from which of the two urns you decide to draw the marble”. The second and the third experiment stimulated further activation of the experiential system by asking subjects to imagine another child performing the task (experiment 2) and by eliminating the normative response (experiment 3). Finally, in the last experiment, Dellai (2007) presented scenarios with unequal probabilities.

Results indicated that younger children are oriented preferentially toward the urn with the largest absolute number of marbles, but older children exhibited the same pattern of responses predicted by CEST. Nevertheless, Dellai (2007) examined verbal explanations, she found that older children were not influenced by the salient “marbles that are in the minority” (Kirkpatrick & Epstein, 1992). The most common motivation was in line with the application of the intuitive rule *more A-more B*: children claimed to choose the container B, because *more* marbles meant *more* probability of success.

In accordance with the intuitive rule, children looked for the urn containing the largest number of salient stimuli, neglecting the numerical ratios between the different marbles (Klaczynski, 2001b). As Reyna (2004) argued, children base their decisional behavior on qualitative-gist representations rather than on quantitative-verbatim traces (Brainerd & Reyna, 2001; Reyna, 2004). Therefore, the *marble game task* led to a prevailing formulation of intuitive responses, without changes during development (Babai, Brecher, Stavy, & Tirosh, 2006; Macpherson, 2001), and knowledge of the normative response did not necessarily lead to its application. For these reasons the experiential system, influencing subjects to choose the urn with the largest absolute number of marbles, led to an immediate response that was difficult to inhibit (Stavy, Goel, Critchley, & Dolan, 2006).

In 2001, Macpherson used the *marble game task* to examine the development of statistical reasoning abilities in children and adolescents and to verify that statistical reasoning is linked to age and cognitive ability. In her study, Macpherson (2001) did not offer any incentive, so that motivation should not have been a factor. Students 10-, 11-, and 13-years-old were included in the experiment, because they provided a sufficient age range to reveal potential developmental differences in reasoning ability.

In addition to differing age groups, Macpherson included a second group of gifted children and she examined the role of cognitive capacity in the development of statistical reasoning abilities.

Participants had to choose a white marble from either a small or a large container. Each container had a different number of white (winning) and blue marbles (null). The small container always contained 10 marbles (1 white and 9 blue), and thus presented a 10% chance of selecting a white marble. The large container always contained 100 marbles, but the number of blue and white marbles was varied slightly over trials (9 white marbles and 91 blue marbles providing a 9% chance of selecting a white marble or 8 white marbles and 92 blue marbles providing an 8% chance of selecting a white marble).

In this experiment, the small container (with 1-in-10 odds of picking a white marble) was deemed the analytic choice because it offered higher probabilities of drawing the winning marble (10% versus 9 % or 8%); whereas the large container (with either 9-in-100 or 8-in-100 odds of picking a white marble) was deemed the heuristic choice. In general, the results for the *marble game task* did not reveal an increase of analytic responses during development. The preference for the analytic choice, instead, was most pronounced among participants in the gifted program. Macpherson (2001) concluded that performance on the statistical reasoning task is associated with cognitive capacity and, using the Thinking Disposition Questionnaire, she found (2001) that *Need for Cognition* and *Superstition* were significant predictors of better reasoning performance.

2.3.1 Contradictions between adults and children

Considering together results from adults' performance (Pavan, 2005; Denes-Raj & Epstein, 1994; Kirkpatrick & Epstein, 1992; Pacini & Epstein, 1999a) and from children and adolescents' performance (Agnoli, Dellai, Stragà, & Furlan, 2007; Babai, Brecher, Stavy & Tirosh, 2006; Macpherson, 2001) for the *marble game task*, it follows that there is no universal pattern of responses that remains constant throughout childhood, and development of reasoning is not unidirectional.

In the condition with high probabilities of winning, Epstein and colleagues (Kirkpatrick & Epstein, 1992; Denes-Raj & Epstein, 1994; Pacini & Epstein, 1999a) hypothesized that people should choose the urn with the smallest absolute number of elements because of the *concretive principle* (that leads people to focus on the most concrete information) and the *experiential learning principle* (that leads subjects to evaluate a single event among many as an unlikely event). To the contrary, Babai, Brecher, Stavy and Tirosh (2006) found that children always choose the urn with the largest absolute number of elements, as predicted by FTT, because of the automatic application of the intuitive rule *more A-more B*: *more* marbles means *more* likely to draw a winning element (Stavy & Tirosh, 2001).

Other research explored the relation between rational and experiential thinking styles and performance on the *marble game task.* Epstein and Pacini (1999b) showed that only rationality is a determining factor in the degree of non-optimal responding. Instead, Macpherson (Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Macpherson, 2001), *showed that good performance subtends great cognitive ability, high Need for Cognition and low Superstition.*

2.4 Reasoning fallacies and perceptual illusions are not comparable: on the relative contribution of individual differences

We started Chapter 1 by describing how perceptual illusions were originally compared to judgment biases because our cognitive system systematically and consciously fails in some cases, much like our visual system. Baron (2008) argued that this analogy is a poor one if only half, or less, of the people show a bias. Nevertheless, some people show biases a lot more than others and other people do not show them at all. For example, as Pacini and Epstein (1999b) claim, some people in their studies always perform optimally whereas others behave somewhat non-optimally, and the responses of the latter vary in degree of non-optimality. Furthermore, some people show the reverse of the usual bias even though these people are rare. Instead, everyone with normal vision sees most visual illusions. Visual illusions may be hard wired into our nervous systems in ways that judgment biases are not.

Together with an increasing interest in cognitive processes involved in reasoning, the 1990s saw an increasing focus on individual differences as a noticeable feature of behavioral decision research (Weber & Johnson, 2009). Researchers underlined that decisions are moderated by individual differences that are outside general intelligence or cognitive ability. For example, Stanovich (1999) clearly presented the distinction between dual-process theories and cognitive styles. He argues that processes whose effectiveness is correlated with individual differences in general intelligence are part of System 2 and processes that are uncorrelated with individual differences in general intelligence are part of System 1. Having taken into account the correlation with general intelligence, he then analyzes *residual variance*, which he shows to be related to dispositional factors inherent in different cognitive styles.

Furthermore, some persons are more able than others to take decisions based on rational and formal reasoning, but independently of this, some persons are also more inclined to reason analytically (Evans, 2007). Analytic thinking is undoubtedly linked to general cognitive ability, as has been proposed by a number of authors, including Reber (1993) and Stanovich (1999). Individual differences in working memory capacity, reasoning ability, and general intelligence scores are all very closely intercorrelated (Kyllonen & Christal, 1990). Evans (2006) explained that those characteristics that determine the analytic system intervention, other than cognitive ability, are *dispositional* (p. 383).

As Weber and Johnson (2009) clearly explain in their review, among the characteristics of the decision maker there are some important individual differences such as cognitive styles (i.e., need for cognition, cognitive reflection) and numeracy which is the ability to process numerical information and probability concepts (Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2003; Reyna & Brainerd, 2008).

2.4.1 Individual differences in the ratio bias

The ratio bias and the Rational-Experiential Inventory (REI)

The basic assumption of CEST is that there are individual differences in personality and thinking styles: some people rely more on rational thought processes, whereas others rely more on intuitive processes rooted in concrete experience. To investigate this aspect of personality, Epstein and colleagues constructed a self-report test, the Rational-Experiential Inventory (REI; Epstein, Pacini, Denes-Raj, & Heier, 1996). Since the REI was intended to measure both the rational processing mode and the experiential processing mode, Epstein, Pacini, Denes-Raj and Heier (1996) adopted different dimensions to capture the specificity of each processing mode. Rationality was nicely captured by a self-report measure of the need for cognition (Cacioppo & Petty, 1982), whereas experientiality was investigated through a self-report measure of the faith in intuition. The term *need for cognition* was defined by Cohen, Scotland, and Wolfe (1955) as “a need to understand and make reasonable the experiential world” (p. 291). Cacioppo and Petty (1982) adopted this term and proposed that need for cognition was a stable (although not invariant) individual difference in the tendency to engage in and enjoy effortful cognitive activity.

Pacini and Epstein (1999b) developed a new and improved version of the test, the REI-40 (in Appendix B), with 20 items each for the rationality and experientiality measures that were again found to be orthogonal to one another. Each measure (for rationality and experientiality) is divided into Ability and Engagement subscales (10 items for each). Rational Ability is a high level of ability to think logically and analytically (e.g., “Using logic usually works well for me figuring out problems in my life”), whereas Rational Engagement is reliance on and enjoyment of thinking in an analytic, logical manner (e.g., “I enjoy thinking in abstract terms”). Experiential Ability is a high level of ability with respect to one’s intuitive impressions and feelings (e.g., “I trust my initial feelings about people”), whereas Experiential Engagement is a reliance on and enjoyment of feelings and intuitions in making decisions (e.g., “I tend to use my heart as a guide for my actions”).

In their original study, Pacini and Epstein (1999b) found that rational and experiential scales each exhibited good internal consistency (Cronbach’s α = .90 and .87 respectively). The structural validity of the REI-40 has been found to be adequate in other languages, also. For example, Björklund and Bäckström (2008) translated the REI-40 into Swedish and found that rationality and experientiality are also separate factors in the Swedish translation. Moreover, REI-40 predicted performance in framing effect problems.

There is also good evidence demonstrating that individuals high in Need for Cognition (NFC) and low in Faith in Intuition (FI) are less likely to fall prey to common decision errors and reasoning fallacies (Pacini & Epstein, 1999; Shiloh, Salton, & Sharabi, 2002). REI-40 has been investigated also in personality. For example, Kemmelmeier (2010) showed that authoritarianism is related to FI, but not to NFC; both lower need for cognition and higher faith in intuition, respectively, have been associated with creativity, paranormal beliefs and use of complementary and alternative medicine (Lindeman & Aarnio, 2006; Raidl & Lubart, 2000/2001; Wheeler & Hyland, 2005).

As Sladek, Bond and Phillips argue (2010), age differences in need for cognition and faith in intuition have rarely been reported or investigated across the lifespan. They investigated 520 participants ranging in age from 20 to 74 years and found that increasing age was associated with decreasing preference for experiential processing and rational processing. They suggested that further research using younger (< 20 years) samples may contribute to an enhanced understanding of the REI scale as a measure of personality traits relevant to thinking and reasoning processes. However, Davies (2006) administered the REI to 433 Australian high school students and reported Cronbach's α of .50 and .58 for the rational and experiential scales respectively, values that are substantially lower than those reported in adult studies.

Marks, Hine, Blore and Phillips (2008) reported that many adolescents in their study thought that the REI was difficult to understand, and they suggested that the adult REI may not be developmentally appropriate for use with adolescents. They developed the 32-item Rational-Experiential Inventory for Adolescents (REI-A) and suggested using it to investigate dispositional preferences for rational and experiential cognition in adolescent populations because of REI-A-32's high internal consistency and excellent temporal stability. However, since Pacini and Epstein (1999) investigated the REI-40 to examine the relation between rational and experiential thinking styles in the ratio-bias phenomenon, we will use the REI-40 in the present investigation to be adherent to Pacini and Epstein's (1999b) original study.

Pacini and Epstein (1999b) observed that the ratio bias can result from strong experiential processing, weak rational processing, or both (p. 981). When ratios are different (participants had a choice between 1-in-10, 7-in-100, and 9-in100), they predicted that rationality scores would be inversely related and/or experientiality scores would be directly related to the number of nonoptimal responses. Instead, when people had to choose between two equal ratios, 1-in-10 versus 10-in-100, they predicted that rationality scores would be unrelated to the ratio bias because there is no rational basis for choosing between two equivalent alternatives; experientiality scores would be directly related to the ratio bias, which means a preference for the large tray for the low-probability condition.

Overall, Pacini and Epstein (1999) found that only rationality scores were negatively related to heuristic processing that leads to the ratio bias. Instead, with regard to sensitivity to numerosity, they did not find a significant main effect of experientiality. They explained this absence of effect as a natural and automatic tendency of people to focus on numerosity roughly at the same degree and concluded that rationality is the determining factor in the degree of non-optimal responding. Rational processing serves to moderate inappropriate experiential processing (p. 985).

Alonso and Berrocal (2003) investigated the relationship between rational thinking style and ratio bias using the Need for Cognition Scale; participants were presented with one of the following three experimental conditions: a) 1-in-10 versus 10-in-100 (10% condition); b) 2-in-10 versus 10-in-100 (20% condition) ; and c) 3-in-10 versus 10-in-100 (30% condition). They also varied the perspective (self-perspective, others-perspective, or logical-perspective) and the scenario, presenting a non-gambling task that Amsel, Close, Sadler and Klaczynski (2009) called “the employment task.” The scenario is as follows (alternatives are presented in brackets):

> Imagine that you have finished your studies and you need to find a job. You are looking through the newspaper and you read an advert from a company that is looking for people like you. This company offers two types of job positions: Type P and Type Q. Both are of the same category and you like them equally. Therefore, you quickly go to the company to present your application to work in either of them. Once there, they tell you that you cannot request both at the same time, you have to opt for one of them: P or Q. They also tell you that: - For the Type P job, 1 [2, or 3] people are needed and only 10 candidates are admitted (one of them would be you). For the Type Q job, 10 people are needed and only 100 candidates are admitted (one of them would be you).

Participants were asked to state which of the two options they would choose. Results showed that participants preferred the more numerous option (10 people out of 100 candidates) only in the others-perspective condition when comparing equal probability ratios (1-in-10 versus 10-in-100). Amsel, Close, Sadler and Klaczynski (2009) argue that this result is due to the sample used, because high school students have little, if any, experience with the task domain of employment. For this reason, the employment task is difficult to compare directly with previous studies.

Instead, in 20% and 30% conditions, participants judged the more numerous ratio (10-in-100) as more probable than 2-in-10 and 3-in-10 both in the self-perspective condition and in the logical-perspective condition. In both conditions, there were no significant differences in the others-perspective. Interestingly, in the 10% condition they also found that people who preferred the 10-in-100 alternative had lower scores on the Need for Cognition Scale than people who preferred the 1-in-10 alternative in both the self-perspective and in the logical-perspective. However, they found no significant difference in the others-perspective condition (that was the only one in which people showed ratio bias). Moreover, the relationship between non-optimal responses (10-in-100) and low degree of rationality was not significant in both the 20% condition and the 30% condition. Finally, the total score of optimal responses (which ranges from 0 to 3) was not correlated with Need for Cognition scores.

A high cognitive ability is related to normative responses in the ratio-bias task

Cognitive capacities, such as working memory capacity, perceptual speed, or discrimination accuracy, refer to the type of cognitive processes studied by information processing researchers who are seeking the underlying cognitive basis of performance on IQ tests (Stanovich & West, 1997, p. 344). As proxies for these processes, Stanovich, West and colleagues used two global indicators of cognitive ability: Scholastic Aptitude Test (SAT) scores and a vocabulary test known to correlate with a number of cognitive subprocesses (Carroll, 1993).

One of the first questions asked in the heuristic and biases literature exploring individual differences in probability judgment was whether cognitive ability and intelligence predict differential responding (Stanovich & West, 1998a, 2000). It is intuitive to think that the more rational and intelligent you are, the more you are able to solve abstract and concrete problems correctly and adaptively. For example, people with cognitive and personality characteristics more conducive to deeper understanding (higher SAT scores, higher Need for Cognition) are supposed to be more likely to favor a correct normative model (Stanovich, 1999). However, despite suspicions that traditional IQ tests miss some of the most important aspects of real-world intelligence, research has indicated that cognitive ability is modestly related to performance on several tasks from the heuristic and biases literature. In addition, Stanovich and West (2008) found evidence that the conjunction fallacy is not correlated with cognitive ability.

Instead, West, Toplak and Stanovich (2008) found positive correlations in 793 undergraduates between the ratio bias (they prefer to call it denominator neglect) and: a) cognitive ability (correlation = .19) measured with self-reported SAT scores; b) thinking dispositions (correlation = .09) measured through the Actively Open-minded Thinking composite (AOT, Stanovich & West, 1997, 2007) and Need for Cognition (Cacioppo, Petty, Feinstein, & Jarvis, 1996); and c) belief bias syllogism (correlation = .15), which is a measure of the tendency for judgments of logical validity to be contaminated by prior knowledge of the world. Although these correlations are statistically significant, in such a large sample relatively small correlations are significant, as the authors also noted (p. 935). We argue that these correlations are small and the findings are weak in terms of practical significance (Kline, 2004).

A positive correlation (.28) between a similar ratio-bias task and cognitive ability was also found in research investigating children of grades 5 to 8 (Kokis, Macpherson, Toplak, West, & Stanovich, 2002). Moreover, Kokis and colleagues (2002) found that the ratio bias was positively correlated with Need for Cognition (.25) but not with the AOT composite (.17). This result might explain the reason why West, Toplak and Stanovich (2008) obtained such a low correlation between the ratio bias and thinking dispositions (.092) in which AOT composite and Need for Cognition were intermixed.

On the relevance of metacognitive skills in the ratio-bias phenomenon

Amsel, Klaczynski, Johnston, Bench, Close, Sadler and Walker (2008) showed the relevance of metacognitive knowledge about the processing sources of responses on the ratio-bias task. Specifically, metacognitive status predicted ratio-bias judgments in the traditional low-probability condition (1-in-10 versus 10-in-100) independently of ACT (American College Test) scores, which were treated as measures of general cognitive skills. Metacognitive awareness is central to the effective regulation of dual processes (Amsel, Close, Sadler & Klaczynski, 2009). In their first study, Amsel and colleagues (2009) replicated the ratio-bias task with equal low probabilities (1-in-10 versus 10-in-100) in a forced-choice selection, and asked participants to rate on a 4-point scale how certain they were that the response option was rational. The metacognitive status of participants was categorized into one of three metacognitive status styles: competent (representing only the analytically based correct response option as rational), conflicted (representing both analytically based correct and experientially based incorrect response options as rational), or poor (misrepresenting an incorrect experientially based response option as the only rational one). Overall they found no significant difference in the preference between the small tray (1-in-10) and the large tray (1-in-100), but participants rated as more rational their willingness to pay for the privilege of choosing the large tray rather than the small tray. More irrational and biased decisions were taken by metacognitively poor participants.

In Study 2, Amsel and colleagues (2009) compared the ratio bias lottery (Kirkpatrick & Epstein, 1992) and employment (Alonso & Berrocal, 2003) tasks in the traditional low-probability condition (1-in-10 versus 10-in-100). All participants performed both tasks and, finally, they recorded their rational certainty judgments for each response option on a slightly altered 4-point Likert-type scale ranging from 1 (not at all sure) to 4 (very sure). In this study, the no-preference option was presented. As predicted, there was a significant task-effect; participants selected Job A (1-in-10) more often than Job B (10-in-100) in the employment task, whereas the same participants selected Jar B (10:100) more than Jar A (1:10) on the lottery task. Regarding metacognitive statuses, Study 2 replicated the Study 1 findings that most participants could be successfully categorized as competent, conflicted, or poor. Those adopting the poor style, whom we assumed to have limited metaknowledge and inhibitory control skills, were shown to be more vulnerable than others to judgmental biases. Because of their exclusive use of intuition, metacognitively poor students were particularly influenced by task context on their judgments (Study 2). Despite making subjective and biased judgments, these participants represented their judgments as rational (p. 314). Amsel and colleagues (2009) concluded by noting how vulnerability to irrational gambling-related decision making and behavior is associated with metacognitive status independently of various background and demographic variables. They challenge future research to investigate the relationship between metacognitive status and skills for inhibition. There is much evidence that supports the relation between inhibitory control, as measured by neuropsychological tasks, and mathematical performance independent of other cognitive and individual difference factors (Bull & Scerif, 2001; Kwon, Lawson, Chung, & Kim, 2000; Peterson et al., 2003). Kwon et al. (2000) in particular found that inhibitory control predicted performance on proportional reasoning tasks, independently of calculation skills, memory capacity, planning ability, and locus of control.

Numeracy and ratio bias

As Weber and Johnson (2009) clearly explain in their review, among the characteristics of the decision maker there are important individual differences such as cognitive styles (i.e., need for cognition, cognitive reflection) and numeracy, which is the ability to process numerical information and probability concepts (Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2003; Reyna & Brainerd, 2008). Low numeracy increases susceptibility to biases in judgment and decision making, and research on this topic is essential both for risk perception and for risk communication (for a review see Reyna, Nelson, Han, & Dieckmann, 2009).

In recent years a variety of instruments have been developed that assess, for example, health numeracy (Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, & Smith, 2007; Lipkus, Samsa, & River, 2001), and several researchers investigated the relation between cognitive ability, thinking styles, numeracy, cognitive reflection and reasoning fallacies (Campitelli & Labollita, 2010; Klaczynski & Cottrell, 2004; Pacini & Epstein, 1999; Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2003, Stanovich & West, 2000). As Reyna, Nelson, Han and Dieckmann (2009) argue, a limit of these measures is that that they do not assess understanding of risk and probability, and adequate understanding of risk and probability is critical for decision making (p. 953). Moreover, no single measure appears to capture the totality of numeracy as a construct, and we argue that great attention should be devoted to understanding the cognitive processes that underlie numerical processing.

With regard to the ratio-bias phenomenon, Peters, Västfjäll, Slovic, Mertz, Mazzocco, and Dickert (2006) investigated whether numeracy, the ability to comprehend and manipulate probability numbers, relates to performance on judgment and decision tasks. To assess numeracy, they used the 11-item Numeracy Scale (NS) developed by Lipkus, Samsa and River (2001), which is currently the most extensively used in research on numeracy in decision making settings. Studies have shown that even highly educated laypersons and health professionals have an inadequate understanding of probabilities, risks, and other chance-related concepts (Estrada, Barnes, Collins, & Byrd, 1999; Nelson, Reyna, Fagerlin, Lipkus, & Peters, 2008; Sheridan & Pignone, 2002). Consequently, Lipkus and colleagues (2001) extended the Schwartz, Woloshin, Black, and Welchet (1997) numeracy assessment to test it in highly educated populations. In Study 3, Peters and colleagues (2006) investigated the relation between numeracy and ratio bias according to the experimental paradigm developed by Denes-Raj and Epstein (1994) in which participants have to compare a small bowl (1-in-10) to a large bowl (9-in-100) and decide which bowl they would prefer to draw from in order to win $5.

As predicted, less numerate participants showed a preference for the large bowl (ratio bias) more than high numerate participants, suggesting that less numerate people are more influenced by irrelevant information and affect source; moreover, they are less likely to retrieve and use appropriate numerical principles and transform numbers presented in one frame to a different frame (p. 412). Curiously, Dale, Rudski, Schwarz and Smith (2007) investigated the ratio-bias effect as a measure for innumeracy related to variations in monetary incentive. They confirmed the existence of the ratio-bias phenomenon and they found that the presence of a monetary incentive affects behavior disregarding its magnitude and reducing the frequency of suboptimal choices.

Chapter 3

Experiment 1: The ratio-bias task in a mathematical context

3.1 Introduction

The ability to manage and understand numerical information is fundamental in everyday life, and low numeracy increases susceptibility to biases in judgment and decision-making (Reyna, Nelson, Han, & Dieckmann, 2009). Ratio concepts—fractions, decimals, percentages, and probabilities—are especially difficult to process, as observed in national and international surveys, as well as in many numeracy assessments. As we described in previous chapters, even mathematically able adults can understand ratios in principle but still have difficulty manipulating them in practice and, they respond based on frequencies, for example, instead of ratios.

According to dual-process theories, the ratio-bias phenomenon is a prominent example of heuristic processing that is the source of biases and errors in numerical processing. To the contrary, analytic processing is the source of objective and optimal numerical decisions, and this is a basic assumption of dual-process theories. Consequently, dual-process theories predict that people high in analytical thinking but low in intuition should be less susceptible to ratio bias than people low in analytical thinking and high in intuition. Unfortunately, as we described, several critical tests of this prediction yielded weak and inconsistent results (e.g., Pacini & Epstein, 1999a, b; see also Alonso & Berrocal, 2003; Reyna & Brainerd, 2008), especially for high probability events. In addition, a long research tradition shows that good decisions are not always arrived by linear and sequential reasoning, but by intuition (Djiksterhuis, Bos, Nordgren, & van Baaren, 2006; Reyna & Farley, 2006; Seligman & Kahana, 2009).

From a developmental perspective children generally perform quite poorly on problems involving numerical ratios until at least 7 or 8 years of age (e.g. Dixon & Moore, 1996; Fischbein, 1990; Kieren, 1988; Mack, 1990; Moore, Dixon & Haines, 1991; Nunes, Schliemann, & Carraher, 1993; Piaget & Inhelder, 1975; Reyna & Brainerd, 1994; Singer, Cohn, & Resnick, 1997). Nevertheless, there are findings that even relatively young children can attend to probability (Acredolo et al., 1989; Boyer, 2007; Schlottmann, 2001; Schlottmann & Anderson, 1994). This is coherent with dual-process theories of development, but these theories also predict that the tendency for analytic processing to override heuristic processing increases with age. Consequently, a clear prediction for the ratio bias is that a developmental increase in normative responses should parallel a decrease in susceptibility to the ratio bias. However, findings on development of the ratio bias (Klaczynski, 2001b) show, for example, that normatively correct responses increase and heuristic responses remain stable from 12-year-olds to young adults.

Some developmental theorists explain that explicit logical and computational operations improve with age at the same time that people acquire a greater variety of heuristics (Jacobs & Klaczynski, 2002). FTT proposes that relative reliance on intuitive gist representations over exacting verbatim representations increases with development, resulting in increased cognitive efficiency (Brainerd & Reyna, 2001; Reyna, Lloyd, & Brainerd, 2003). For example, according to Reyna and Brainerd (1994), this tendency to respond based on frequencies instead of ratios, or denominator neglect, is present early in development (as early as first grade) and persists in adolescence and adulthood. Although young children have the competence to compare ratios, they tend to focus on target classes, contained in numerators, especially when ratios are equal. Confusion arises when classes of objects or events being compared overlap or are nested within one another (Reyna & Brainerd, 2008).

Overall, we argue that on the one hand research on the development and interaction of analytic and heuristic processing is fundamental to understand the origins of some reasoning errors, such as the ratio bias. On the other hand, if we take together results on adult decision-making and results on the development of the two processes, then we conclude that developmental trajectories clearly differ depending on the bias/heuristic and the specific task at issue. We totally agree with Jacobs and Klaczynski (2002) who say that it is difficult to predict which developmental trend should appear under which conditions (Jacobs & Klaczynski, 2002), because developmental results show different trajectories and phenomena such as the ratio bias show different patterns in studies with somewhat different methodologies. We argue that developmental trajectories should not differ depending on the specific task at issue, but rather depending on reasoning processes under the specific task. This implies that experimental variations of the task should be strictly controlled before generalizing results.

In the previous chapters, we showed that two theories explain the *ratio bias*: the Cognitive-Experiential Self-Theory (Epstein, 1990) and the fuzzy-trace theory (Brainerd & Reyna, 1990; Reyna 2004, 2008). Predictions of the two theories are similar when dealing with low probability events: for example, with regard to the *two-urn-choice-task,* both CEST and FTT predict that people show a tendency to judge according to absolute numerosity (e.g., 10-in-100 is more probable than 1-in-10) rather than probability ratio.

Instead, the high-probability winning condition represents a critical point because CEST and FTT have different predictions and the results that CEST's investigators found are mixed. CEST predicts a preference for the ratio composed of smaller numbers (e.g., 9-in-10) whereas, according to FTT, people would prefer the ratio composed of larger numbers (e.g., 90-in-100) because people tend to rely on the gist "more is better than less". In addition, the high-probability winning condition has never been investigated from a developmental point of view. Our purpose is to understand whether seventh graders, middle adolescents, and adults exhibit the *ratio bias* in the standard *two-urn-choice-task* adapted from Kirkpatrick and Epstein (1992), which we call the *mathematical scenario*.

The rationale to choose seventh graders as the younger age group is twofold. On the one hand, seventh graders are able to understand both the problem and the numerical information presented. They have already received formal instruction on fractions and probabilities. Because we seek to understand whether the heuristic processing generates the same bias across ages, participants must be able to comprehend ratios. This is a guarantee that any systematic errors of reasoning that we observe can be attributed to a bias rather than just a failure in analytic processing.

We assume that 12-year olds are in a critical phase of cognitive development with regard to analytic processing and heuristic processing. Increasing experience, socially shared responses, and the tendency to rely on gist processing increase with age, become more deeply entrenched in memory, and are applied automatically to decisions. In other words, the increasing influence of heuristic processing corresponds to an increasing tendency to contextualize information, and, in our case, numerical information. As Morsanyi and Handley (2008) point out, children ranging from 5- to 11-years-old may lack the necessary cognitive resources for contextualizing some decision-making problems in the first place. This is relevant because the ability to decontextualize requires contextualized representations.

The ability to decontextualize (to extract a generalized representation distinct from its specific context) is essential for analytic reasoning competencies to be used to evaluate decision-making problems (Klaczynski, 2000; Stanovich, 1999). In addition, it is only during the second decade that children increasingly develop the potential to manage their cognitive resources in consciously controlled and purposefully chosen ways (Kuhn, 2009). Consequently, if the ratio bias results from a failure of analytic processing to inhibit a contextualized/gist representation generated by the heuristic process, then both abilities related to the analytic processing and contextualized representations related to the heuristic processing must be sufficiently developed to correctly address the problem. Reyna and Narter (1991) found that the verbatim-to-gist shift is observed in framing problems with children as young as 10 years old. In this age group the framing reversal was attenuated. These children were old enough to process the quantitative information, although their execution of relevant computations was less reliable than adults' execution.

With regard to probability level, we presented participants the well-studied condition with equal probabilities in which participants compare 9-in-10 to 90-in-100 (Kirkpatrick & Epstein, 1992). In this condition, we expect that a small proportion of participants in all age groups will exhibit the ratio bias. With equal probabilities (9-in-10 versus 90-in-100), we argue that the correct response (it is the same) might depend on an automatized knowledge due to the habit born of practice simplifying fractions.

Consequently, we think that ratios of equal probability may not be the best test of ratio bias.

At this point, it is important to remember that a bias is a systematic error. In other words, every time a comparison between a small-numbered ratio and a large-numbered ratio is presented, if there is a systematic preference toward a specific quantity (i.e., the large-numbered ratio) than the choice must *always* be directed toward that quantity regardless of the ratio. Consequently, we introduced two new conditions in which the two probability ratios differ slightly and are difficult to compute. If there is a bias, than the preference must be systematic disregarding the normative answer. Solving these problems with unequal ratios requires good computational ability. For this reason, we expect normative responses to increase with age.

However, when the difference between two ratios is too small to detect by a rough computational ability, it is reasonable to think that decisions will be based on estimation or intuitive cognition (Fischbein & Schnarch, 1997). Consequently, unequal ratios are more likely to generate heuristics and exhibit biases. We expect that the bias will be stronger with unequal ratios than with equal ratios in all age groups.

With regard to the direction of bias, if participants are biased toward the ratio composed of larger numbers (as predicted by FTT), then they will show a systematic preference toward the large-numbered ratio. To the contrary, if participants are biased toward the ratio composed of smaller numbers (as predicted by CEST), then they will show a systematic preference toward the smaller-numbered ratio. The mathematical problem that we used is decontextualized and should not activate specific representations related to every-day life experience. Consequently, we expect that the direction of the bias will be the same in all age groups and, specifically, toward the large-numbered ratio because of the activation of the lowest level gist: “more is better than less”.

Finally, we will attempt to clarify which individual differences explain participants’ performance on the ratio-bias task presented as a mathematical problem. In recent years a variety of instruments have been developed that assess numeracy (Fagerlin, et al., 2007; Lipkus, 2001) and several researchers investigated the relation between cognitive ability, thinking styles, numeracy, cognitive reflection, and reasoning fallacies (Campitelli & Labollita, 2010; Klaczynski & Cottrell, 2004; Pacini & Epstein, 1999; Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2003, Stanovich & West, 2000).

Research has demonstrated that individuals with faster and greater cognitive capacity tend to perform better on reasoning problems (Evans, Newstead, & Byrne, 1993; Stanovich & West, 1998). However, there is also evidence (Klaczynski et al., 1997) that cognitive capacity makes a unique contribution to reasoning performance. In addition, according to dual-process theories the correct resolution in the ratio-bias task, as we discussed in Chapter 2, should be related to analytic processing (i.e., mathematical ability and numeracy) and to some rational thinking dispositions such as need for cognition (Cacioppo, Petty, Feinstein, & Jarvis, 1996) or cognitive reflection. Consequently, we hypothesize that good performance on the ratio-bias task presented as a mathematical problem is related to high level of mathematical ability and analytic competence in all age groups.

On the other hand, a large corpus of research debates whether biased decisions should be attributed to an inhibition failure or a conflict detection failure per se (De Neys & Glumicic, 2008; Houdè, 2007). If the inhibitory process operates during the resolution of the ratio-bias task, than the correct solution should be related to the ability to suppress or inhibit the erroneous answer that springs impulsively to mind because of heuristic processing. We will measure this ability in both middle adolescents and adults using the Cognitive Reflection Task (Frederick, 2005). However, because the mathematical problem is decontextualized and participants will have all the time they need to make calculations by hand, we hypothesize that the inhibition process is not involved, and measures of inhibition ability are not related to a good performance.

3.2 Method

Participants. Participants were 94 seventh graders (43 females, 51 males) ranging in age from 12 to 14 years (mean age = 12.94 years, *SD* = 0.42 years), 58 middle adolescents (32 females, 26 males) ranging in age from 16 to 18 (mean age = 16.47 years, *SD* = 0.57 years), and 30 adults (17 females, 13 males) ranging in age from 18 to 29 (mean age = 20.85 years, *SD* = 1.83 years). Seventh graders and middle adolescents were recruited from public schools in the northeast part of Italy; all participants were typically developing children of middle socioeconomic status. Adults were undergraduate students at Cornell University (Ithaca, NY) and were recruited in psychology courses and via campus postings. All participants gave written informed consent and the Institutional Review Board of Cornell University approved the part of the research conducted at Cornell University.

Material and procedure. This experiment (and the following experiments) was designed and run using Qualtrics.com online survey software (Qualtrics Labs Inc., Provo, UT). Qualtrics supports personalized and controlled experimental conditions (i.e., with reference to logic, randomization of questions and blocks of questions, and branches and control of the reading times with a visible countdown programmed in Java script). Qualtrics also supports an advanced online questionnaire with many practical advantages such as downloading the dataset directly according to a response code.

All respondents participated in the experiment online by accessing Qualtrics's secure survey site. Seventh graders and middle adolescents performed the experiment in their school's computer lab during lesson time. Cornell undergraduates signed up for the experiment using *Susan* (http://susan.psych.cornell.edu/), which is a web page that allows experimenters to list their studies, students to participate, and professors to observe their students' participation and generate end-of-term extra-credit lists. Cornell undergraduates received one class credit for completing the experiment. To protect participants' anonymity a sequentially generated participation number identified their data; their names were not recorded.

All participants performed the *added-constant task*. This is a probabilistic reasoning task modeled on the ratio-bias task introduced by Kirkpatrick and Epstein (1992) and constructed following a pattern similar to that used by Kokis, Macpherson, Toplak, West, and Stanovich (2002). The added-constant task consists of three trials in which a problem is held constant but the number of marbles in container B varies. The mathematical problem appears below with the manipulated information in brackets:

> *Two containers, labeled A and B, are filled with red and blue marbles in the following quantities.*
> *Container A contains 10 marbles, 9 red and 1 blue.*
> *Container B contains 95[100 or 105] marbles, 85 [90 or 95] red and 10 blue.*
> *You must draw a marble (without looking, of course) after choosing one of the containers. If you draw a red marble, you win, otherwise you win nothing and the game is over.*

The small container (container A) always contains 10 marbles (9 red and 1 blue) for each of the three trials, and thus presents a 90% chance of selecting the red winning marble. Both the total number of marbles and the proportion of red winning marbles vary in the large container (container B). In Trial 1, the large container held 95 marbles (85 red and 10 blue); thus, the probability of drawing a red winning marble was equal to 89.5%. In Trial 2, the large container had 100 marbles (90 red and 10 blue); thus, the probability of drawing a red winning marble was equal to 90%. Finally, in Trial 3, the large container held 105 marbles (95 red and 10 blue); thus, the probability of drawing a red winning marble was equal to 90.5%.

We manipulated the odds of winning by adding a constant $k = 5$ to both the numerator (85, 90, 95) and the denominator (95, 100, 105), which resulted in three different ratios (85:95, 90:100 and 95:105) to be compared with 9 out of 10 (container A). Consequently, the probability of winning was: a) higher for container A (90%) than container B (89.5%) in Trial 1; b) the same for both containers (90%) in Trial 2; c) and higher for container B (90.5%) than container A (90%) in Trial 3. All participants performed all three trials and presentation order of the trials was randomized. Note that the difference in probability ratios is small (.005) in both conditions with different probability ratios (Trials 1 and 3).

First participants read the problem. On the next page, they were asked to choose the container that gave a better chance of winning. They had to choose one of three answers: a) container A (9:10); b) “it would not matter to me; chances are the same”; c) container B (85:95) in Trial 1 (90:100 in Trial 2 and 95:105 in Trial 3). The order of presentation of the three answers was randomized. Participants had as much time as they wanted to make their choice and reason about the problem; moreover, they were allowed to calculate but only by hand.

We controlled reading time in accordance with the reading speed ability of each age group. According to standards for normal reading, adults' reading rate for learning is 100-200 wpm. Because the number of words in the instructions and the problem summed to 92, we applied an average of 150 wpm and gave participants 40 sec to read the problem. The time available to read the problem was specified in the instructions and a countdown was visible at the bottom of the page indicating how much time participants had remaining to read and understand the problem. Middle adolescents and adults were given the same amount of time.

Seventh graders had 60 seconds to read the problem because Italian children of that age require an average of 22 hundredths of a second to read each syllable (Cornoldi & Colpo, 1995).

After participants decided which container gave a better chance of winning, a new page appeared that asked them to explain their choice. The purpose of this page was to explore the reasoning process that led to their choice. The question was: "Can you explain the basis for your response?" After completing all three trials of the task, seventh graders and middle adolescents were asked about age, gender, and their practice with computers or videogames. Cornell undergraduates were asked about age, gender, ethnicity, and student status.

3.2.1 Individual differences measures

Seventh graders

Seventh graders' mathematical ability was assessed using the AC-MT 11-14 (Cornoldi & Cazzola, 2003). This test investigates various aspects of mathematical learning including written and oral mathematical computation ability, symbolic arithmetical reasoning ability, computational speed, and problem solving skills. We assessed written calculation, comprehension and production, and arithmetical reasoning.

Total score in written calculation (range: 0-10) is obtained by summing correct answers to two subscales that measure: a) the ability to calculate additions, subtractions, multiplications and divisions (range: 0-8); and b) the ability to solve arithmetical expressions that involve ratios (range: 0-2). Total score in comprehension and production (range: 0-20) is obtained by summing correct answers to three subscales that measure: a) the ability to recognize which number is the biggest between five options presented as ratios, or decimals, or exponentials (range: 0-4); b) the ability to compose a number in digital form that is expressed in terms of the number of thousands, hundreds, tens, units, tenths, and hundredths that it contains. (range: 0-8); and c) the ability to express in digits a number that is written in words (range: 0-8).

Total score in arithmetical reasoning is obtained by summing correct answers to two subscales (range: 0-32) in which: a) participants have 2 minutes to solve as many operations as they can (range: 0-16); and b) participants identify relationships between entries in two columns. The left column contains 16 solved mathematical operations. The right column contains 16 parallel mathematical operations without calculations. Participants extract the mathematical rule in the right column that underlies the operation in the left column (range: 0-16). Participants have 2 minutes to solve as many operations as they can.

According to Cornoldi and Cazzola (2003), the overall arithmetical ability (range: 0-82) is calculated by summing the total score in comprehension and production, the total score in arithmetical reasoning, and three times the total score in written calculation.

Middle adolescents and college students (adults)

Middle adolescents performed the Primary Mental Ability test (P.M.A.; Thurstone & Thurstone, 1963). We used three subscales: verbal meaning (range: 0-50; time to complete the test equal to 4 minutes), calculation ability (range: 0-70; time to complete the subscale equal to 6 minutes) and spatial relations (range: 0-54; time to complete the subscale equal to 5 minutes).

Adults completed a 12-item short form of the Raven Advanced Progressive Matrices (RAPM) test (Arthur & Day, 1994) as a measure of cognitive capacity that is designed for adults with above-average general intelligence.

Middle adolescents and adults completed several scales assessing numeracy, thinking style and cognitive inhibition. The adults also performed the go/no-go task

(Garavan, Ross, Murphy, Roche & Stein, 2002) to test their ability in inhibitory control. A detailed description of these measures follows.

Objective numeracy scales. Participants answered the Lipkus, Samsa and Rimer's (2001) objective numeracy scale (NS). This 11-item scale includes questions covering orders of magnitude, probability, converting metrics, and arithmetical computation. NS is divided into two subscales: the General Numeracy subscale (range: 0-3) consists of three questions involving probability (i.e.," Imagine that we rolled a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even?"). Instead, the Expanded Numeracy subscale (range: 0-8) frames questions about probability within the context of health risks (i.e., "If Person A's chance of getting a disease is 1 in 100 in ten years, and person B's risk is double that of A's, what is B's risk?").

Subjective numeracy scales. Participants completed the 8-item Subjective Numeracy Scale (SNS, Fagerlin et al., 2007; Zikmund-Fisher, Smith, Ubel, & Fagerlin, 2007) to assess people's perception of their numerical competence. The SNS is divided in two subscales: in the Cognitive abilities subscale, participants rate their capacity to manage with fractions, percentages and calculation (i.e., "How good are you at figuring out how much a shirt will cost if it is 25% off?") on a 6-point scale that ranges from 1 (*not at all good*) to 6 (*extremely good*). The second subscale is the Preference for display of numerical information (range: 0-4) in which participants rate their preference for managing numbers rather than words expressing probability in different contexts (i.e., "When people tell you the chance of something happening, do you prefer that they use words ('it rarely happens') or numbers ('there's a 1% chance')?"). Participants rate all the items on a 6-point scale that ranges from 1 to 6. In our experiment, the range of answers varies from 1 to 5 in both subscales because of measurement coherence with the other rating scales.

Cognitive reflection task. The cognitive reflection task (CRT; range: 0-3) measures cognitive impulsivity or one's reliance on more intuitive (e.g. automatic) versus deliberative (e.g. effortful and subjectively controlled) cognitive processing (Frederick, 2005). This task consists of three problems in which an intuitive answer springs quickly to mind, but this *impulsive* answer is wrong.

Rational thinking style. As a measure of rational thinking style we used the Need for Cognition Scale (NCS) in its long version (20 items), which Pacini and Epstein (1999) adjusted from the original version of the Cacioppo and Petty scale. In this scale, 10 items belong to the Rational Ability subscale and another 10 items refer to the Rational Engagement subscale. Participants filled in this longer version of the NCS and rated all the items on a 5-point scale that ranged from 1 (*definitely not true for myself*) to 5 (*definitely true for myself*).

Faith in intuition. Faith in intuition (FI) is related to Type 1 processes and focuses on degree of confidence that feelings and impressions are a basis of one's decisions and actions. This scale (which is the second part of the Rational Experiential Inventory) consists of 20 items according to the version adjusted by Pacini and Epstein (1999b), and it is divided in two subscales. Experiential Engagement (10 items) refers to reliance on and enjoyment of feelings and intuitions in making decisions (e.g., I tend to use my heart as a guide for my actions). Instead, Experiential Ability (10 items) refers to reports of a high level of ability with respect to one's intuitive impressions and feelings (e.g., "I can usually feel when a person is right or wrong, even if I can't explain how I know") . Participants rated all 20 items on a 5-point scale ranging from 1 (*definitely not true of myself*) to 5 (*definitely true of myself*).

Need for Cognition (20 items) and Faith in Intuition (20 items) were used together by Pacini and Epstein (1999b) as the Rational Experiential Inventory (REI scale). Even though Marks and colleagues (2008) developed the 32-item Rational-experiential inventory for adolescents (REI-A), we prefer to administer the original REI-40 (Pacini & Epstein, 1999b) to our sample of middle adolescents because the REI has never been used or validated for an Italian sample.

Inhibitory control. The inhibitory control ability of adults was tested with the Go/no-go task (Garavan, Ross, Murphy, Roche & Stein, 2002). This task's software was hosted by Columbia University and compiled using E-Prime TM software (Psychology Software Tools, Inc., Pittsburg, PA; Schneider, Eschman, & Zuccolotto, 2002) by the Center for the Decision Sciences of Columbia University. Participants were directed to an external secure link at the end of the experiment and accessed the task through a personal ID number, which was randomly generated during the survey. Participants first completed a trial version of the go/no-go task and then were tested in two blocks of stimuli. Both blocks required subjects to respond as quickly and accurately as possible by pressing the 'h' key every time the "X" (go cue) appeared and not to respond to the "K" (no-go cue). Stimuli were presented in the center of the screen for 500 ms. Each block contained 140 stimuli, of which 112 (80%) were go cues and 28 (20%) were no-go cues. The interstimulus interval (ISI) was 500 ms and the presentation order of go cues and no-go cues was pseudorandomized to discourage anticipatory responses. A fixation cross was displayed in the center of the screen during the ISI. Instructions were displayed on the computer screen at the beginning of each block and subjects pressed the spacebar when ready to begin. Go/no-go task duration was up to 8 minutes.

3.3 Results

3.3.1 Statistical analysis method

Traditional analysis approaches—ordinary least squares regression on the means of items computed by aggregating data over participants—is not appropriate for this experiment because the dependent variable is dichotomous (Jaeger, 2008). Moreover, traditional analyses do not take into account the variability from both participants and items. If we consider, for example, a reaction-time experiment in which all participants respond to three trials in each condition, their reaction times will be different in each trial. A repeated-measure analysis of variance (RM-ANOVA) would lose the information related to individual variability.

There are several reasons why mixed-effects models are preferred over a conventional RM-ANOVA (Baayen, Tweedie, & Schreuder, 2002; Quené & van den Bergh, 2004). Mixed-effects models are often more powerful than the univariate or multivariate approaches compared to RM-ANOVA, especially if the sphericity assumption is violated, as it often is in real data. First, variances and covariances may be modeled explicitly to take into account potential heterogeneous variances and collinearity. Second, both discrete and categorical predictors can be included in a single model.

In the present experiment, participants performed three trials; some participants might give the correct answer in one trial, others might give the correct response in all three trials. There is individual variability that should be accounted for in the model.

Generalized Linear Mixed Models (GLMM) extend Generalized Linear Models (McCullagh & Nelder, 1989; Nelder & Wedderburn, 1972) and represent a class of fixed effects regression models for several types of dependent variables such as binary variables (e.g., correct versus incorrect responses) and factors with ordinal levels (e.g., low, middle, and high educational levels). Generalized Linear Models are an important extension to ordinary least squares regression models. Parameter estimation, however, is not based on minimizing the sum of squared errors. Instead, parameters are chosen such that, given the data and our choice of model, they make the model predicted values most similar to the observed values. This general technique is known as maximum likelihood estimation. Parameters that describe the effect of a factor in ordinary linear models are called fixed effects.

In contrast, random effects apply to a sample. For an experiment with repeated measurements, like our experiment, a cluster is a set of observations for a particular participant, and the model contains a random effect term for each participant. The random effects refer to a sample of clusters from all the possible clusters and, contrary to fixed effects, which are an unknown constant that we try to estimate from the data, a random effect is a random variable (Faraway, 2006).

GLMM have the great advantage of including random effects as a predictor and they describe an outcome as the linear combination of fixed effects and conditional random effects associated with subjects and items (Jaeger, 2008). While fixed effects are modeled by means of contrasts, random effects are modeled as random variables with a mean of zero and unknown variance.

The data of Experiment 1 were analyzed using mixed logit models (Generalized Linear Mixed Models) for binomially distributed outcomes (Agresti, 2002; Bates & Sarkar, 2007; Breslow & Clayton, 1993; DebRoy & Bates, 2004) because we categorized the response given to the added constant task as correct or incorrect.

Analyses were performed using R software with the function glmer in package lme4 (Bates, 2005). This is an extension package to R, an open-source package for statistical analysis available from http://www.r-project.org (R Development Core Team, 2008).

The Akaike information criterion AIC (Akaike, 1974; Burnham & Anderson, 2002) was employed as the model-selection method (Myung, Forster, & Browne, 2000; Wagenmakers & Waldorp, 2006). A baseline model was constructed with all the possible interactions and main effects, and the best-fitting model was defined as the one minimizing the AIC. This approach finds the model that explains the data with a minimum of free parameters; the AIC selection criterion balances between a good fit and a simple model. In order to verify that this most simple model is justified, we carry out a likelihood ratio test (e.g., Pinheiro & Bates, 2000, p. 83) that compares the most specific model to the more general.

3.3.2 Analysis of correct responses

Table 7 presents the proportion of responses given by each age group to each trial. Adults gave a high proportion of correct answers in all three trials (.80, .77, and .80), showing no evidence of bias. The middle adolescents also achieved a high proportion of correct answers (.74) when the ratios were equal for Containers A and B, suggesting that they quickly recognized that simplifying 90/100 yields 9/10. Seventh graders were more challenged by this trial, achieving a proportion correct of only .55. The trials with unequal ratios are the critical tests of alternative theories. Middle adolescents strongly preferred Container A (the smaller container), selecting this option in 67% of the trials in which the ratio for Container A was larger (9:10 versus 85:95) and 40% of the trials in which it was smaller. Seventh graders also preferred Container A, selecting this option in 56% of the trials in which the ratio for Container A was larger and 29% of the trials in which it was smaller.

Table 7

Proportion of Preferences (Correct Responses in Bold) in Three Age Groups

Age group	Trial	Preference: Container A (9-in-10)	Preference: Container B (more numerous)	Preference: No difference
Seventh graders (N = 94)	9:10 versus 85:95	**.56**	.31	.14
	9:10 versus 90:100	.19	.26	**.55**
	9:10 versus 95:105	.29	**.47**	.24
Middle adolescents (N = 58)	9:10 versus 85:95	**.67**	.14	.19
	9:10 versus 90:100	.16	.10	**.74**
	9:10 versus 95:105	.40	**.29**	.31
Adults (N = 30)	9:10 versus 85:95	**.80**	.10	.10
	9:10 versus 90:100	.10	.13	**.77**
	9:10 versus 95:105	.13	**.80**	.07

A Generalized Linear Mixed Model was run with age (seventh graders, middle adolescents or adults), and trial (9-in-10 versus 85-in-95, 9-in-10 versus 90-in-100 or 9-in-10 versus 95-in-105) as fixed effects. Participants were included in the model as a random factor to take into account the dependence between our observations due to repeated measures. The random intercept for participants allowed the intercept to vary between participants.

The best-fit model is summarized in Table 8, including regression coefficients, their standard errors, and the corresponding Z scores. The right side of Table 8 shows the odds ratios, their confidence intervals and the relative significance of main effects and the interaction (Chi square). We compared the three age levels by fixing the seventh graders as the reference and comparing middle adolescents and adults to the seventh graders. We chose seventh graders as the reference category because previous findings (i.e., Jacobs & Potenza, 1991; Reyna & Ellis, 1994) showed that young children perform better than older children and adults in some decision-making tasks. In addition, we compared the three trial levels by fixing the trial 9-in-10 versus 90-in-100 as the reference.

The model shows that both adults (B = 1.08, Z = 2.15, p < .05) and middle adolescents (B = .84, Z = 2.30, p < .05) perform better than seventh graders. As shown in Table 7, the pattern of preferences is different for seventh graders and middle adolescents in the trials with unequal ratios, and seventh graders perform better than middle adolescents when judging 9-in-10 versus 95-in-105 compared to 9-in-10 versus

90-in-100. The present analysis confirms the pattern; seventh graders perform 4.90 (inverse of the .20 odds ratio) times better than middle adolescents when judging 9-in-10 versus 95-in-105 compared to 9-in-10 versus 90-in-100 (B = -1.59, Z = -3.13, p < .01).

Table 8

Generalized Linear Mixed Model Parameters and Effect Sizes Predicting Correct Answers in Experiment 1

Best-fit model	B	(SE)	Z	OR	95% CI	Chisq (df)
Age						22.10 (2)***
Middle adolescents	0.84	(.36)	2.30*	2.32	[1.14, 4.69]	
Adults	1.08	(.50)	2.15*	2.94	[1.11, 7.85]	
Trial						22.09 (2)***
9:10 versus 85:95	< 0.001	(.29)	< 0.001	1.00	[.57, 1.77]	
9:10 versus 95:105	-0.34	(.29)	-1.17	0.71	[.40, 1.26]	
Age x Trial						13.27 (4)***
Middle adolescents x 9:10 versus 85:95	0.33	(.50)	0.66	1.39	[.52, 3.71]	
Adults x 9:10 versus 85:95	0.23	(.74)	0.31	1.26	[.30, 5.37]	
Middle adolescents x 9:10 versus 95:105	-1.59	(.51)	-3.13**	0.20	[.08, .55]	
Adults x 9:10 versus 95:105	0.34	(.71)	.48	.48	[.35, 5.65]	

Note. *p < .05, **p < .01, *** p < 0.001; Baseline categories: seventh graders for Age; and 9-in-10 versus 90-in-100 for Trial.

Overall, results suggest that the trajectory of correct responses increases with age. Adults are clearly able to make correct decisions based on ratios. Seventh graders were surprisingly better in performance than middle adolescents when comparing 9-in-10 to 95-in-105. Did seventh graders reason more normatively than middle adolescents? If so, then we might have expected seventh graders to perform better than middle adolescents with the other unequal ratio (9-in-10 versus 85-in-95), but this was not the case.

A possible explanation is that correct responses are not necessary related to normative reasoning. Instead, it is possible that the correct responses are due to alternative reasoning strategies. For example, seventh graders might give a high proportion of correct answers when comparing 9-in-10 to 95-in-105 because, in accordance with FTT, they neglect the denominator and judge 95 bigger than 9. We argue that the analysis of correct responses does not shed light on a crucial question: are correct responses really dependent on mathematical reasoning or on heuristics? We address this question in the next section.

3.3.3 Analysis of biased responses

To understand whether seventh graders have a bias toward the large ratio and whether this bias changes or disappear with age, we analyzed the responses the responses to each trial (small container, large container, or *it is the same*). We used multinomial logistic regression because the dependent variable is polytomous (Agresti, 2007; Yee & Mackenzie, 2002). In this analysis, selecting the small container (9-in-10) is the comparison category and selecting both the large container and *it is the same* are compared with it, yielding estimates of the effects of the predictor variables (age and trial) on the probabilities of the responses. We compared: a) the three age levels by fixing the seventh graders as the reference and comparing middle adolescents and adults with the seventh graders; and b) trials by fixing trial 1 (9-in-10 versus 85-in-95) as the reference and comparing trial 2 (9-in-10 versus 90-in-100) and trial 3 (9-in-10 versus 95-in-105) with trial 1.

The Akaike information criterion AIC (Akaike, 1974; Burnham & Anderson, 2002) was employed as the model-selection method (Myung, Forster, & Browne, 2000; Wagenmakers & Waldorp, 2006). The best-fit model took into account the interaction between age and trial. The best-fit model is summarized in the left side of Table 9, including the regression coefficients, their standard errors, and the corresponding *Z* scores. The right side of Table 1 shows the odds ratios, which measure effect size, and their 95% confidence interval (CI). The odds ratios are the antilog (i.e., exponentiated values) of the model coefficients.

The top half of Table 9 compares the choice of the small container (9-in-10) to the choice of the large container (85-in-95, or 90-in-100, or 95-in-105). This comparison is particularly relevant because it allows us to test directly systematic preferences for the small container or the large container. Comparing trial 3 (9-in-10 versus 95-in-105) with trial 1 (9-in-10 versus 85-in-95), there was a statistically significant preference (B = 1.07, Z = 3.18, p < .001) for the small container. Adults responded the small container more often than seventh graders in trial 1 and the large container more often in trial 3 (B = 2.80, Z = 3.17, p < .001). Of more interest, both middle adolescents (B = -1.00, Z = -2.21, p < .05) and adults (B = -1.50, Z = -2.28, p < .05) have a stronger preference than seventh graders for the small container (9-in-10) or, conversely, seventh graders have a stronger preference for the larger container. This results suggests that seventh graders performed better than middle adolescents in trial 3 because of a bias toward the large container instead of greater mathematical ability.

Table 9

Multinomial Logistic Regression Model Parameters and Effect Sizes (N = 182)

Best-fit model	*B* (SE)	*Z*	OR	95% CI
Selecting the small container (9-in10) versus selecting the large container				
Age (Adolescents)	-1.00 (.45)	-2.21*	.37	[.15, .89]
Age (Adults)	-1.50 (.65)	-2.28*	.22	[.06, .80]
Trial (9-in-10 versus 90-in-100)	-.87 (.39)	-2.24*	2.39	[1.11, 5.13]
Trial (9-in-10 versus 95-in-105)	-1.07 (.34)	-3.18***	2.92	[1.50, 5.68]
Age (Adolescents) X Trial (9-in-10 versus 90-in-100)	.31 (1.76)	.40	1.36	[.04, 42.93]
Age (Adults) X Trial (9-in-10 versus 90-in-100)	1.49 (1.05)	1.42	4.44	[.57, 34.74]
Age (Adolescents) X Trial (9-in-10 versus 95-in-105)	.21 (.60)	.34	1.23	[.38, 4.00]
Age (Adults) X Trial (9-in-10 versus 95-in-105)	2.80 (.88)	3.17***	16.44	[2.93, 92.28]
Selecting the small container(9-in-10) versus selecting the response it is the same				
Age (Adolescents)	.12 (.46)	-.26	1.13	[.46, 2.78]
Age (Adults)	.69 (.69)	-1.01	.50	[.13, 1.94]
Trial (9-in-10 versus 90-in-100)	2.45 (.41)	-5.92***	11.59	[5.19, 25.88]
Trial (9-in-10 versus 95-in-105)	1.23 (.42)	-2.92**	3.42	[1.50, 7.79]
Age (Adolescents) X Trial (9-in-10 versus 90-in-100)	.38 (.65)	.59	1.46	[.41, 5.23]
Age (Adults) X Trial (9-in-10 versus 90-in-100)	1.67 (.96)	1.74	5.31	[.81, 34.87]
Age (Adolescents) X Trial (9-in-10 versus 95-in-105)	.21 (.63)	-1.33	.81	[.24,2.79]
Age (Adults) X Trial (9-in-10 versus 95-in-105)	0.16 (1.14)	.14	1.17	[.13, 10.96]

Note. OR = odds ratio; CI = confidence interval; *p < .05, **p <.01, ***p <.001

This finding also shows that the direction of the bias changes with age. In accordance with the predictions of FTT, seventh graders have a preference toward the large container, whereas, in accordance with CEST, adolescents and adults prefer the small container.

The bottom half of Table 8 compares preference for 9-in-10 to preference for the response *it is the same*. The main effect of trial is significant with regard to the comparison between 9-in-10 versus 90-in-100 and 9-in-10 versus 85-in-95 ($B = 2.45$, $Z = 5.92$, $p < .001$) because the response *it is the same* is the correct answer in trial 2. Also the comparison between 9-in-10 versus 95-in-105 and 9-in-10 versus 85-in-95 ($B = 1.23$, $Z = 2.92$, $p < .01$) is significant, but this result depends on the higher proportion of preferences for 9-in-10 that participants gave in trial 1.

3.3.4 Analysis of written explanations associated with the container preference

To understand how participants reasoned, we analyzed the explanations that participants gave to their responses after each trial. Specifically, we coded their explanations in four categories:

1. Mathematical reasoning: participants who reasoned mathematically and explained their response through considerations of proportions and probabilities. For example "I solved the problem mathematically," or "statistically, one is better than the other," or "probabilities are the same/differ,"
2. Denominator neglect (as defined by FTT): participants who wrote that they responded according to the absolute number of winning marbles. For example "the large container is better because there are more winning balls."
3. Reasoning based on the absolute number of losing balls (according to CEST): participants who stated that they responded according to the smaller number of losing balls. For example "in the small container there is 1 losing ball (1-9) whereas in the large container there are 10 losing balls (10-90). 1 losing ball is better than 10";

4. Intuition: participants who explained their response by referring to intuition, gut feelings, sensations, or were unable to provide a formal explanation. For example, “I felt that the 9:1 ratio had a better probability” [9-in-10 versus 85-in-95].

Table 10 shows the proportion of correct and wrong responses associated with each explanation for each trial. In trial 2 (9-in-10 versus 90-in-100), most participants explained their responses through mathematical reasoning, referring to probabilities and mathematical computation. The proportion of correct responses associated with mathematical reasoning increased with age, whereas the application of early-acquired strategies (denominator neglect or reversed reasoning) decreased with age. This result suggests that seventh graders, middle adolescents and adults are able to compare 9-in-10 with 90-in-100, and their correct responses can be attributed to analytic reasoning.

In trials with unequal ratios the proportion of correct answers associated with mathematical reasoning again increased with age. These proportions were, however, much lower for seventh graders and middle adolescents than in trial 2. About a fourth of seventh graders and middle adolescents responded correctly in Trial 1 (9-in-10 and 85-in-95) and justified their response, as predicted by CEST, based on the number of non-winning balls. Nearly as many (.17 seventh graders and .21 middle adolescents) responded incorrectly in Trial 3 with this same justification. Only about 10% of seventh graders and middle adolescents relied only on numerators to explain their choices in these two trials. We consider explanations based on denominator neglect and on the absolute number of non-winning balls to be *early-acquired strategies*.

Seventh graders and middle adolescents were more likely to refer to mathematical reasoning as a justification of their responses when the ratios in the two containers were equal. When ratios were unequal, they were more likely to rely on early-acquired strategies. Note that attending to the number of non-winning balls (the reversed reasoning strategy) yielded the correct response in the comparison between 9-in-10 and 85-in-95 but the incorrect response in the comparison between 9-in-10 and 95-in-105.

Table 10

Proportions of Correct and Wrong Responses Associated With Different Explanations Used by Participants of Each Age Group in Three Trials

9-in-10 versus 90-in-100	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (N = 94)	.48	.04	—	.12	.01	.13	.06	.16
Middle adolescents (N = 58)	.62	.02	.02	.05	—	.10	.10	.08
Adults (N = 30)	.74	.07	.03	.07	—	.07	—	.03
9-in-10 versus 85-in-95	**Mathematical reasoning**		**Denominator neglect**		**Reversed reasoning**		**Intuition**	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (N = 94)	.16	.12	—	.13	.26	—	.14	.20
Middle adolescents (N = 58)	.22	.19	—	.07	.23	.02	.23	.05
Adults (N = 30)	.67	.10	—	.03	.03	—	.10	.06
9-in-10 versus 95-in-105	**Mathematical reasoning**		**Denominator neglect**		**Reversed reasoning**		**Intuition**	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (N = 94)	.17	.18	.12	.02	.01	.17	.17	.16
Middle adolescents (N = 58)	.10	.29	.05	.02	—	.21	.14	.19
Adults (N = 30)	.53	.03	.03	—	—	.07	.23	.10

This demonstrates that correct responses cannot always be attributed to correct analytic reasoning. Moreover, according to dual-process theories, even if participants knew the correct response, they may be triggered to give a wrong response because of a conflict caused by the heuristic system. None of these participants, however, referred to any conflict in their explanations or mentioned anything related to experience or beliefs.

3.3.5 Predictors of reasoning performance

To understand the dependence of correct responses on reasoning processes, we investigated participants' individual differences. Do relationships exist between good reasoning and cognitive capacity as well as between good reasoning and thinking dispositions? We present the results separately for each age group because participants of different ages performed different tests.

Seventh grade reasoning performance

After participants performed the mathematical problem, their mathematical ability was assessed using the AC-MT 11-14 (Cornoldi & Cazzola, 2003). We investigated their written computation ability, their comprehension and production ability, and their arithmetical reasoning ability. According to traditional cognitive development theories and traditional dual-process accounts, the seventh graders' performance in the ratio-task should improve as their mathematical competence improves. Table 11 presents the minimum, maximum, mean, standard deviation, and normative scores for each measure and for the AC-MT 11-14 total score.

Table 11

Descriptive Statistics for the AC-MT 11-14 (N = 94)

	Min	Max	*M*	*SD*	Normative scores
Written calculation	0	10	5.83	2.59	6.34
Comprehension and production	1	20	13.09	4.71	14.40
Arithmetical reasoning	3	28	15.88	5.82	19.09
AC-MT 11-14 total score	6	75	46.46	14.04	52.26

According to the performance standards established by Cornoldi and Cazzola (2003), 12.8% of the participants need immediate intervention, 27.7% need attention, 55.3% performed sufficiently, and only 4.3% had excellent mathematical ability. Figure 7 shows the density distribution of the total number of correct responses in the ratio-bias task (range = 0-3) for each of these four levels of performance. The four distributions overlap, and it is interesting that seventh graders with the highest mathematical ability and seventh graders with the lowest mathematical ability achieved similar performance in the ratio-bias task.

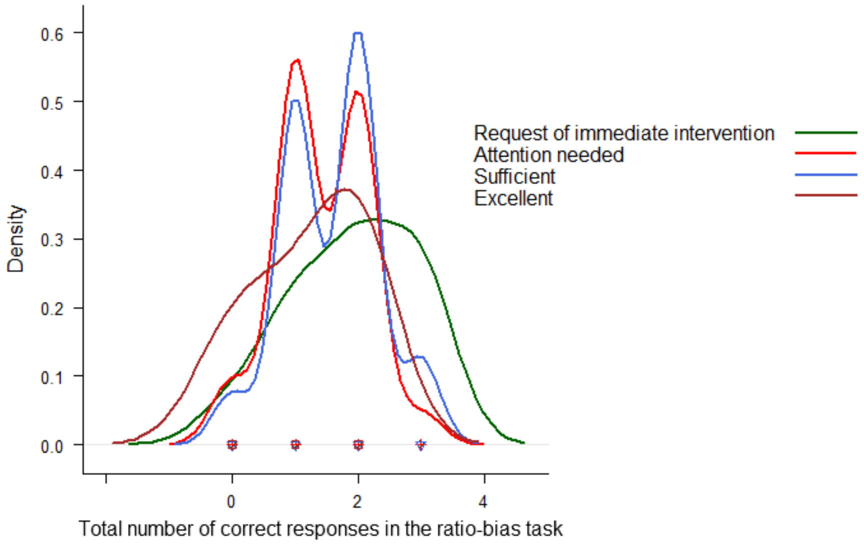

Figure 7. Density plots for correct responses grouped by seventh grade mathematical level according to the AC-MT 11-14

This pattern is confirmed by Table 12 that shows the correlations between accuracy in each trial of the ratio-bias task, the total correct responses in the ratio-bias task, and the AC-MT (subscales and total score).

Table 12

Tetrachoric, Polychoric and Pearson Correlations Between the Primary Variables in AC-MT 11-14 and Accuracy in the Ratio-bias Task (N = 94)

Variables	1	2	3	4	5	6	7	8
Ratio-bias task								
1. 9-in-10 versus 85-in-95	—							
2. 9-in-10 versus 90-in-100	.22*	—						
3. 9-in-10 versus 95-in-105	-.21*	-.21*	—					
4. Total correct responses	.61**	.61**	.40**	—				
AC-MT 11-14								
5. Written calculation	.08	.08	-.20*	-.02	—			
6. Comprehension and production	.14	.25*	-.29*	.54**	.56**	—		
7. Arithmetical reasoning	.01	-.11	-.04	-.08	.33**	.17	—	
8. AC-MT 11-14 total score	.10	.09	-.22*	-.03	.88**	.72**	.66**	—

Note. *p < .05; **p < .01

The three trials of the ratio-bias task are correlated, suggesting that they activate common reasoning processes. The negative correlation between trial 1 (9-in-10 versus 85-in-95) and trial 3 (9-in-10 versus 95-in-105) is evidence that some participants who responded correctly in one trial also responded incorrectly in the other trial. This is consistent with evidence of a bias in the seventh grade responses.

Total correct responses in the ratio-bias task are positively correlated with comprehension and production ability (r =.56, p < .01). This is the ability to recognize the magnitude of numbers and transform written numbers as digits. This ability is relevant to manipulating ratios. It is paradoxical that trial 3 (9-in-10 versus 95-in-105) is negatively correlated with written calculation (r = -.20, p < .05), comprehension and production (r = -.29, p < .05), and total AC-MT score (r = -.22, p < .05). Mathematical ability is not predictive of performance in trial 1 (9-in-10 versus 85-in-95) and trial 2 (9-in-10 versus 90-in-100).

Middle adolescent reasoning performance

The tests used with middle adolescents were the PMA, objective numeracy, subjective numeracy, cognitive reflection ability, need for cognition and faith in intuition. Table 13 summarizes descriptive statistics for each scale.

Table 13

Means, Standard Deviations and Reliabilities of the PMA, Objective Numeracy, Subjective Numeracy, Need for Cognition, Faith in Intuition, and CRT (N = 58)

	Min	Max	*M*	*SD*	Cronbach's *α*
PMA					
Spatial ability	6	52	24.95	12.11	—
Verbal ability	10	42	26.45	7.71	—
Calculation ability	6	43	20.83	9.78	—
Objective numeracy total score	0	11	7.03	2.22	.63
General numeracy	0	3	1.45	1.01	.44
Expanded numeracy	0	8	5.59	1.63	.57
Subjective numeracy general mean	1.50	4.75	3.17	.52	.52
Cognitive ability	1.74	5.00	3.28	.72	.63
Preference for information display	1.25	4.50	3.05	.65	.30
Cognitive reflection task total score	0	3	0.43	0.77	—
Need for cognition general mean	2.10	4.65	2.92	5.82	.76
Rational ability	2.00	4.60	2.93	0.51	.69
Rational engagement	1.90	4.70	2.91	0.50	.58
Faith in intuition general mean	2.55	4.65	3.36	0.42	.76
Experiential ability	2.30	4.80	3.25	0.50	.58
Experiential engagement	2.40	4.50	3.28	0.44	.67

Table 14 shows the correlations between the responses for each trial of the ratio-bias task and each scale with its relative subscales.

Table 14.

Middle Adolescent Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Mathematical Problem (N = 58)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ratio-bias task																				
1. 9-in-10 vs. 85-in-95	—	-.05	.43**	.70**	-.04	.05	-.11	-.14	-.17	-.09	-.40**	-.54**	-.02	-.26	-.07	-.03	-.09	.01	.01	-.03
2. 9-in-10 vs. 90-in-100		—	.13	.55**	.30*	.14	.10	.35*	.36*	.26	-.05	-.08	.01	.57**	.19	.24	.08	-.11	-.04	-.14
3. 9-in-10 vs. 95-in-105			—	.74**	.17	-.22	-.01	.19	.01	.26	-.07	-.09	-.22	.10	.05	-.08	.18	.14	.22	.05
4. Total correct responses				—	.17	-.01	-.01	.16	.08	.17	-.25	-.27	-.09	.17	.07	.05	.07	.02	.10	-.05
PMA																				
5. Spatial ability					—	.22	.03	.20	.24	.12	.12	.18	-.01	.60**	.35*	.26	.35*	-.13	-.01	-.22
6. Verbal ability						—	.01	.25	.15	.24	.01	.03	-.01	.22	.05	.16	-.07	-.07	-.05	.07
7. Calculation ability							—	-.04	.02	-.07	.13	.19	.01	.12	.18	.03	.29*	.04	-.03	-.10
Objective Numeracy Scale																				
8. Objective numeracy general score								—	.73***	91***	.16	.25	-.01	.48**	-.06	.01	-.12	.20	.13	.21
9. General numeracy									—	.37*	.04	.15	-.10	.26	.08	.08	.06	.09	.02	.13
10. Expanded numeracy										—	.19	.24	.04	.48**	-.13	-.03	-.21	.21	.16	.21
Subjective Numeracy Scale																				
11. Subjective numeracy general score											—	.79**	.74**	.49**	.27	.21	.26	-.03	.09	-.12
12. Cognitive ability												—	.17	.43**	.23	.14	.27	.14	.18	.08
13. Information display													—	.30*	.18	.19	.13	-.20	-.07	-.28
Cognitive Reflection Task																				
14. Cognitive reflection general score														—	.33*	.39*	.22	-.30*	-.19	-.31*
Need for Cognition Scale																				
15. General need for cognition score															—	.87***	.87***	.05	.01	-.09
16. Rational ability																—	.52**	.04	.07	.01
17. Rational engagement																	—	-.12	-.05	-.16
Faith in Intuition Scale																				
18. Faith in intuition general score																		—	.87**	.90***
19. Experiential ability																			—	.58**
20. Experiential engagement																				—

Note. *p < .05; **p < .01

Only trials 1 and 3 of the ratio-bias task (the trials with unequal ratios) were correlated, suggesting that these two trials activate common reasoning processes. Total correct responses in the ratio-bias task was not significantly correlated with any reasoning performance abilities, but subjective numeracy general score and cognitive ability were correlated with accuracy in trial 1 (r = -.40 and r = -.54, respectively).

Adult reasoning performance

The tests used with adults were the Raven Advanced Progressive Matrices (RAPM), objective numeracy, subjective numeracy, cognitive reflection ability, need for cognition and faith in intuition. Table 15 summaries the descriptive statistics for each scale.

Table 15

Means, Standard Deviations and Reliabilities of the RAPM, Objective Numeracy, Subjective Numeracy, CRT, go/no-go task, Need for Cognition, and Faith in Intuition (N = 30)

	Min	Max	*M*	*SD*	Cronbach's *α*
RAPM					
Total score	4	12	9.63	2.14	—
Objective numeracy total score	7	11	10.03	1.07	.51
General numeracy	1	3	2.40	.72	.37
Expanded numeracy	6	8	7.63	.61	.21
Subjective numeracy general mean	2.38	4.88	3.71	.63	.73
Cognitive ability	1.50	5.00	3.67	.89	.85
Preference for information display	1.75	4.75	3.74	.71	.58
Cognitive reflection task total score	0	3	1.67	1.21	—
Go/no-go task					
Reaction time (ms)	167.42	492.37	302.53	56.42	—
Correct answers	66	275	244.64	35.76	—
False alarms	2	47	20.94	9.40	—
Need for cognition general mean	3.05	4.85	3.75	0.50	.87
Rational ability	2.80	4.90	3.68	0.46	.69
Rational engagement	2.90	4.90	3.82	0.61	.83
Faith in intuition general mean	2.10	4.50	3.39	0.53	.88
Experiential ability	2.10	4.50	3.40	0.53	.81
Experiential engagement	2.10	4.30	3.36	0.56	.79

Table 16 shows the correlations between the accuracy in each trial of the ratio-bias task and each scale with its relative subscales. Accuracy for trials 1 and 2 were correlated (r = .52), but correlations among other pairs of trials were not significant.

Adult Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Mathematical Problem (N = 30)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Ratio-bias task																					
1. 9-in-10 vs. 85-in-95	—	.52**	.01	.57**	-.19	-.21	-.27	-.05	.04	.25	-.29*	.28*	.18	-.04	-.07	.21	.19	.21	.39*	.40*	-.06
2. 9-in-10 vs. 90-in-100		—	.22	.73**	-.11	-.10	-.05	-.12	-.11	.06	-.31*	-.07	.09	-.08	-.12	-.30*	-.23	-.36*	.15	.16	.04
3. 9-in-10 vs. 95-in-105			—	.72**	-.03	.33*	.25	.27	.32*	.20	.31*	-.42*	.04	-.70**	.28	.06	.17	-.09	-.28*	-.28*	-.33*
4. Total correct responses				—	-.06	.07	.06	.04	.24	.30*	.01	-.09	.05	-.46*	.24	.01	.10	-.13	-.02	-.02	-.23
RAPM																					
5. RAPM total score					—	.08	.19	-.08	.35*	.34*	.20	.38*	.34*	-.14	-.13	.48*	.52**	.35*	-.17	-.18	-.13
Objective Numeracy Scale																					
6. Objective numeracy general score						—	.84***	.75**	.40*	.29*	.34*	.19	.04	-.14	-.07	.47*	.43*	.46*	-.28*	-.28*	-.50**
7. General numeracy							—	.26	.42*	.38*	.28*	.19	.04	-.11	-.08	.46*	.42*	.46*	-.34*	-.35*	-.40*
8. Expanded numeracy								—	.19	.06	.27*	.11	.02	-.11	-.03	.27*	.26*	.25*	-.08	-.07	-.40*
Subjective Numeracy Scale																					
9. Subjective numeracy general score									—	.83***	.72***	.35*	-.12	.15	.31*	.60**	.56**	.57**	-.04	-.04	-.18
10. Cognitive ability										—	.22	.39*	.02	.10	.21	.58**	.55**	.53**	.14	.14	.02
11. Information display											—	.12	-.25	.13	.29*	.33*	.29*	.33*	-.24	-.24	-.32*
Cognitive Reflection Task																					
12. Cognitive reflection general score												—	.05	.46*	-.08	.44*	.38*	.45*	-.21	-.21	-.42*
Go/no-go task																					
13. Reaction Time													—	-.49*	-.71**	.16	.16	.14	.15	.15	-.10
14. Proportion of correct responses														—	.10	-.02	-.08	.08	.06	.06	.17
15. Proportion of false alarm															—	-.04	-.02	-.07	-.09	-.09	.17
Need for Cognition Scale																					
16. General need for cognition score																—	.95***	.92***	-.07	-.08	-.34*
17. Rational ability																	—	.76**	-.13	-.14	-.37*
18. Rational engagement																		—	.01	.02	.26
Faith in Intuition Scale																					
19. Faith in intuition general score																			—	.99***	.72**
20. Experiential ability																				—	.72**
21. Experiential engagement																					—

Note. $^{*}p < .05$; $^{**}p < .01$

The only measures significantly correlated with total accuracy in the ratio-bias task were cognitive ability for adults ($r = .30$, $p < .05$) and the proportion of correct responses in the go/no-go task ($r = -.46$, $p < .05$). This latter correlation is surprisingly negative. Correct responses at the go/no-go task depend on the ability to respond as quickly and accurately as possible by pressing the ‘h’ key every time the “X” (go cue) appeared and not to respond to the “K” (no-go cue). Nevertheless, the proportion of correct responses in the go/no-go task is negatively correlated to the RT in the go/no-go task ($r = -.49$, $p < .05$) and positively correlated to the cognitive reflection general score ($r = .46$, $p < .05$). This pattern of results suggests that the ability to inhibit the wrong answer depends on the capacity to detect the conflict and that this process is probably very fast.

3.4 Discussion

In the present experiment, we investigated the ratio-bias phenomenon in a decontextualized situation to test predictions of CEST and FTT and to understand whether ratios that are difficult to compute generate a higher proportion of biased answers compared to equal ratios.

In trial 9-in-10 versus 90-in-100 most of middle adolescents (.74) and adults (.77) gave the correct answer suggesting that participants quickly recognize that simplifying 90/100 yields 9/10. Only seventh graders gave a low proportion of correct answers (.55) but this result might depend on a low mathematical ability rather than on a systematic preference toward one container.

With unequal ratios we hypothesized that participants are more likely to exhibit heuristics and biases, but we also expected that performance improves with age. Cornell undergraduates performed better than adolescents and seventh graders, and they showed no bias at all. Instead, seventh graders exhibited a bias toward the large container performing better than adolescents in trial 9-in-10 versus 95-in-105. At the same time, according to CEST, middle adolescents showed a preference for the smaller container (Denes-Raj & Epstein, 1994). We remind that CEST theorists claim that people prefer small numbers because of the concretive principle: people would prefer 9-in-10 because

9-in-10 winning balls are more concrete and easy to visualize than 90-in-100. In other words, what is familiar is good when thinking intuitively (Reyna & Brainerd, 2008, p. 91).

However, the analysis of explanations revealed that a small proportion of seventh graders (12% on average) justified their choice neglecting the denominator and focusing only on the absolute numbers of winning balls. With unequal ratios both seventh graders and middle adolescents who preferred the small container justified their response differently from what is expected: seventh graders and adolescents relied on the absolute number of losing balls rather than on winning balls. That is: *less* number of losing balls (9-in-10) means *less* probabilities to lose.

These findings show that, contrary to what we hypothesized, the direction of the bias is not *always* toward the large container. Seventh graders apply different strategies to solve the ratios, and, according to Siegler (1981), it is possible that when seventh graders have difficulties to calculate ratios, they apply rules previously acquired during development. Nevertheless, these rules are rational and we might define them as *formal heuristics* related to analytic process since participants relied on numbers and never mentioned to have experienced a conflict between the correct answer and something irrational or gut feeling.

Contrary to dual-process theories and to what we expected, individual differences did not shed light on what is predictive of a good performance in the added-constant task. The total number of correct answers in the added-constant task: a) is linked only to comprehension and production in seventh graders; b) is independent from all the individual differences measures in adolescents; and c) is independent from all the individual differences measures in adults except for cognitive ability measured with the Subjective Numeracy Score as a self-report of cognitive ability.

It is possible that as adults need to think carefully to solve the ratio-bias task, the more likely they are to make errors in the go/no-go task. This task requires the capacity to process the information very fast and without thinking whereas the time-paced condition in the present experiment allows analytic thinking and calculation. If this is true, than we expect that when participants are required to solve the ratio-bias task in a time-pressure condition, correct answers would depend on correct intuition which is by definition very fast. Consequently, the correlation between total number of correct responses in the added-constant task and the proportion of correct responses in the go/no-go task should be positive. We will test this prediction in Experiment 3.

Chapter 4

Experiment 2: The ratio-bias task in everyday-life scenarios

4.1 Introduction

There is evidence (Kirkpatrick & Epstein, 1992) that real-life simulations in which participants judge the probabilities of an event from a self-perspective increase the tendency to exhibit the ratio bias. This is coherent with dual-process theories because real-life situations activate contextualized representations related to heuristic processing. In previous research, for example, Alonso and Berrocal (2003) used an everyday-life scenario that Amsel, Close, Sadler and Klaczynski (2009) called the *employment task* (see Chapter 2). As Amsel and colleagues (2009) showed, participants preferred the small-numbered option because they focused on the numerator of the ratio, which was the number of employment competitors. Participants explained that *they believed that fewer competitors for a job would increase their chances of being noticed and hired* (p. 309). Amsel and colleagues (2009) also noted that the ratio-bias job task does not parallel the ratio-bias lottery task because only the latter task actually involves a random draw.

According to dual-process theories, context influences the strength of the representation generated by heuristic processing. When representations are more context-based and correspond more closely with experienced reality, heuristic processes are more likely to be triggered and yield a biased response. The strength of heuristic representations should increase with age as a consequence of increased experience. As De Neys and Vanderputte (2010) point out, heuristic thinking develops with age, but how it develops has been scarcely investigated by dual-process theorists.

Biased responses should increase in everyday-life contexts compared to a decontextualized scenario such as the mathematical problem in Experiment 1.

According to CEST, participants comparing 9-in-10 versus 90-in-100 in a high-probability frame will be biased toward the small-numbered ratio because of concretive and experiential principles if the scenario activates heuristic processes. According to FTT, participants exhibit denominator neglect and will have a systematic preference for the large-numbered ratio because participants tend to pay less attention to the denominator as a default: participants rely on the gist *more is better than less*. In addition, when ratios are unequal and difficult to compute (such as those we used in Experiment 1: 9-in-10 versus 85-in-95 and 9-in-10 versus 95-in-105), we expect that in everyday-life scenarios participants rely on heuristics more than with equal ratios (9-in-10 versus 90-in-100).

From the perspective of traditional dual-process theories, the proportion of heuristic responses should decrease with age. On the one hand, the heuristic process activates contextualized representations that are expected to be highly compelling in everyday-life scenarios. If seventh graders, middle adolescents, and adults activate the same representation, then heuristic answers should decrease with age because the ability of analytic processing to override heuristic answers increases with age. However, according to Klaczynski (2001b), it is also possible that heuristic answers remain constant with age because such answers are heuristic *products* related to the specific culture.

On the other hand, according to FTT, children change from quantitative reasoners (verbatim-based) to qualitative reasoners (gist-based) as they grow to adulthood. Qualitative reasoners process categorical (e.g., more money is better than less money, no risk is better than some risk) gist (Reyna, 2004). Consequently, seventh graders should rely on numerical and detailed information more than middle adolescents and adults. Thus, biased answers should increase with age. One might argue that unequal ratios are difficult to compute and non-optimal answers, especially in seventh graders, might be related to mere performance errors. However, if non-optimal responses depend on performance errors rather than on systematic reasoning errors, then the overall participants' responses on the two unequal ratios would be at the chance level.

According to the traditional definition of bias as a systematic error *toward one direction*, we must assume that the direction of the bias (toward the small-numbered ratio as predicted by CEST, or toward the large-numbered ratio as predicted by FTT) should *not* change across different contexts. The influence of the context on the direction of the bias has never before been investigated, neither in adults nor from a developmental point of view.

On the one hand, if the direction of the bias changes according to different contexts, either we conclude that there is no bias at all or we have to presume that biased answers are not irrational. Instead, biased answers depend on the activation of different experience-based memory representations that generate rational decisions consistent with the surface of the reasoning problem. In other words, according to FTT (Reyna, 2004; Reyna & Farley, 2006), people decide on the basis of a consistency criterion that is adherent to reality. We hypothesize that different contexts generate different patterns of biased answers.

On the other hand, in Chapter two we described some counterintuitive findings from a developmental point of view in which younger children perform better than adults. These results have been explained by theorists as related to something *not-already-developed*, such as stereotypes or the ability to contextualize a certain situation. These are post-hoc explanations that, nevertheless, share the common denominator that heuristic processing develops with age. If heuristic processing develops with age and participants have sufficient time to access age-related and experience-based memory representations, then we predict that in everyday-life situations that activate different levels of expertise and familiarity: a) younger children (seventh graders) will exhibit the same direction of bias regardless of the context and b) middle adolescents and adults will show a context-based bias.

Finally, we hypothesize that if participants rely on quantitative reasoning to answer in everyday-life scenarios, then measures related to analytic processes (i.e., numeracy, need for cognition) will be positively associated with correct responses. Conversely, if participants rely on heuristic representations, then measures related to analytic process will be unrelated to correct answers.

We devised two scenarios, a *volleyball problem* and a *vaccination problem*. The volleyball problem requires deciding which of two teams to join, a problem that should be familiar to all age groups. According to Alonso and Berrocal (2003) and to Amsel and colleagues (2009), this type of scenario should trigger the heuristic that an alternative with fewer applicants is better, because fewer alternatives offer more chances to be noticed. Participants in all age groups should prefer to join a smaller team.

The vaccination problem requires choosing between two clinics. This is a health-related decision, and we expect that seventh graders are less familiar than middle adolescents and adults with everyday-life heuristics related to medical environments. Adults often think, for example, that bigger clinics perform more surgeries. In addition, a big clinic is considered better than a small one because it will have greater experience. Thus, in the vaccination problem, middle adolescents and adults should be triggered to choose a clinic that treats more patients. If seventh graders are inexperienced making medical decisions, then this heuristic would not be activated and seventh graders would base their answers on ratios or exhibit the same bias as in the volleyball problem.

4.2 Method

Participants. Participants were 157 seventh graders (83 females, 74 males) ranging in age from 12 to 14 years (mean age = 12.72 years, *SD* = 0.45 years), 131 middle adolescents (56 females, 75 males) ranging in age from 15 to 17 (mean age = 15.80 years, *SD* = 0.55 years) and 69 adults (40 females, 29 males) ranging in age from 19 to 36 (mean age = 21.28 years, *SD* = 3.18 years). Seventh graders and middle adolescents were recruited from public schools in the northeast part of Italy; all participants were typically developing children of middle socioeconomic status. Adults were undergraduate students at Cornell University (Ithaca, NY) and were recruited in some psychology courses and via campus postings. All participants gave written informed consent and the part of the study that took place at Cornell University was approved by the Institutional Review Boards of Cornell University.

Material and procedure. The experiment was designed and run using Qualtrics.com online survey software (Qualtrics Labs Inc., Provo, UT). All respondents participated in the experiment online by accessing Qualtrics's secure survey site. Seventh graders and middle adolescents performed the experiment in their school's computer lab during lesson time. Cornell undergraduates signed up for the experiment using "Susan" (http://susan.psych.cornell.edu/) and received one class credit for completing the experiment. To protect participants' anonymity their data was identified by a sequentially generated participation number; their names were not recorded.

All participants performed the same three trials as in Experiment 1. Consequently, they solved trial 1 (9-in-10 versus 85-in-95), trial 2 (9-in-10 versus 90-in-100), and trial 3 (9-in-10 versus 95-in-105). The presentation order of trials was randomized. Participants were divided into two groups that were assigned two different everyday-life problems, the *volley problem*, which is an adaptation of the employment task (Alonso & Berrocal, 2003), and the *vaccination problem*. The two problems are as follows:

> The volleyball problem. *A Selection Committee invited you to participate in first-round volleyball recruitment. You like two teams of equal skill. Your volleyball skills are equally good for each team, but you can only apply for one. For Team A there are 10 applicants (one of them would be you), 9 of whom will be selected. For Team B, there are 95 [100 or 105] applicants (one of them would be you) and 85 [90 or 95] will be selected.*

> The vaccination problem. *You are called because you must get a vaccine. There are two clinics of equal competence where you can go to get vaccinated and you must select one of them. You are told that there are 10 people vaccinated per week (including you) at Clinic A, and 9 vaccines are expected to be effective. At Clinic B, instead, there are 95 [100 or 105] people vaccinated per week (including you), and 85 [90 or 95] vaccines are expected to be effective. The waiting time to get the vaccine is the same at both clinics.*

First participants read the problem. Then on the next page they were asked to say which team gave a better chance of being selected or which clinic gave a better chance that the vaccine is effective. They were asked to choose one of three answers: a) Team/Clinic A (9:10); b) "it would not matter to me; chances are the same"; or c) Team/Clinic B (85:95) in the case of Trial 1 (90:100 in Trial 2 and 95:105 in Trial 3) according to the problem that each participant solved. Participants had as much time as they wanted to make their choice and reason about the problem; moreover, they were allowed to make calculations but only by hand. The use of electronic devices was explicitly not allowed.

As in Experiment 1, we controlled the reading time in accordance with the reading ability of each age group. Seventh graders were given 80 seconds to read the volleyball problem and 68 seconds to read the vaccination problem, which is about 3.5 sec per syllable. Middle adolescents were given 40 sec to read the volleyball problem and 42 sec to read the vaccination problem, which is about 150 wpm. Adults were given 40 sec and 50 sec to read the English-language version of the volleyball and vaccination problems, respectively. The time available to read the problem was specified in the instructions and a countdown was visible at the bottom of the page indicating how much time participants had remaining to read and understand the problem.

After participants selected a response, they were asked to explain this choice. They were asked, “Can you explain the basis for your response?” Finally, demographic questions requested information about Cornell undergraduates’ age, gender, ethnicity, and student status. Seventh graders and middle adolescents were asked about age, gender, and their practice with computers or videogames.

4.2.1 Individual differences measures

The individual differences measures were exactly the same as in Experiment 1. Seventh graders’ mathematical ability was assessed using the AC-MT 11-14 (Cornoldi & Cazzola, 2003). Middle adolescents performed the Primary Mental Ability test. Adults completed the Raven Advanced Progressive Matrices (RAPM) test and the go/no-go task to test inhibitory control. Both middle adolescents and adults completed several scales assessing numeracy and thinking style, including objective numeracy, subjective numeracy, cognitive reflection, rational thinking style, and faith in intuition.

4.3 Results

4.3.1 Analysis of correct responses

Figure 8 compares the density plots of the total number of correct responses (range: 0-3) that the three age groups gave in each problem. As Figure 8 shows, most adults responded correctly. Instead, seventh graders’ and middle adolescents’ performance are similar and their density curves overlap.

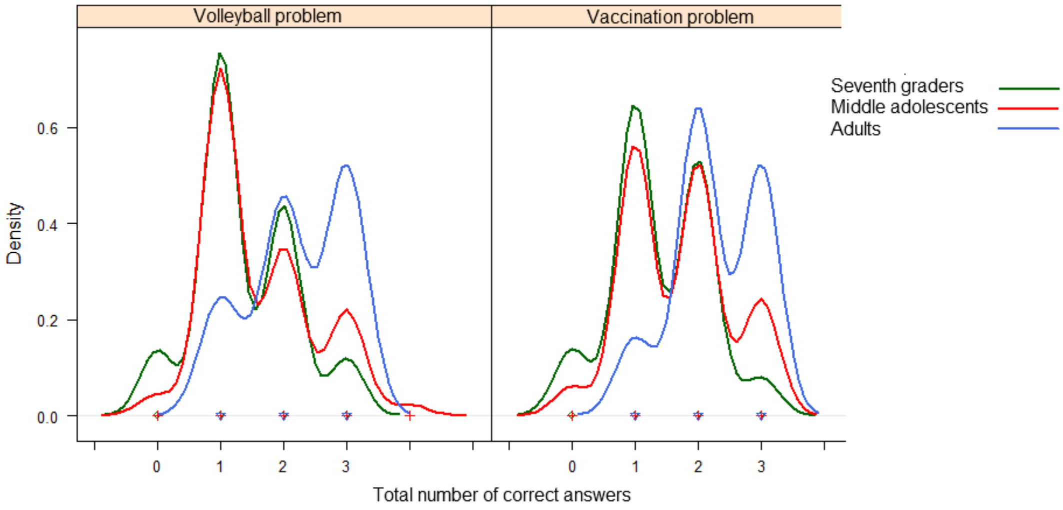

Figure 8. Density plots of the number of correct responses in each trial grouped by age for each problem

Table 17a and Table 17b present the proportion of responses given by each age group to each trial in the volleyball problem and the vaccination problem.

In the volleyball problem, seventh graders responded correctly to 46%, middle adolescents to 51%, and adults to 75% of all trials. In the vaccination problem, seventh graders responded correctly to 47%, middle adolescents to 57%, and adults to 75% of all trials. Within each age group performance was about the same for seventh graders and adults, but middle adolescents responded more accurately in the vaccination problem.

With equal ratios in trial 2 (9-in-10 versus 90-in-100) middle adolescents selected the correct response (*it makes no difference*) more often than seventh graders in both scenarios. Furthermore, these two age groups responded correctly more often in the vaccination problem than the volleyball problem.

When ratios were unequal the pattern of results was somewhat different for the two scenarios. In the volleyball problem, middle adolescents responded correctly (.63) more often than seventh graders (.50) in trial 1 (9-in-10 versus 85-in-95), but seventh graders (.40) responded correctly more often than middle adolescents (.27) in trial 3 (9-in-10 versus 95-in-105). Interestingly, this pattern is reversed in the vaccination problem. In this problem middle adolescents (.49) responded correctly less often than seventh graders (.58) in trial 1 and more often (.41) than seventh graders (.20) in trial 3

Table 17a

Proportion of Preferences (Correct Responses in Bold) in the Volleyball Problem

Age group	Trial	Preference		
		Team A (less numerous)	Team B (more numerous)	No difference
Seventh graders (N = 86)	9:10 vs. 85:95	**.50**	.26	.24
	9:10 vs. 90:100	.23	.28	**.49**
	9:10 vs. 95:105	.39	**.40**	.21
Middle adolescents (N = 62)	9:10 vs. 85:95	**.63**	.18	.19
	9:10 vs. 90:100	.26	.10	**.64**
	9:10 vs. 95:105	.37	**.27**	.36
Adults (N = 35)	9:10 vs. 85:95	**.78**	.11	.11
	9:10 vs. 90:100	.09	.14	**.77**
	9:10 vs. 95:105	.11	**.71**	.18

Table 17b

Proportion of Preferences (Correct Responses in Bold) in the Vaccination Problem

Age group	Trial	Preference		
		Clinic A (less numerous)	Clinic B (more numerous)	No difference
Seventh graders (N = 71)	9:10 vs. 85:95	**.58**	.15	.27
	9:10 vs. 90:100	.28	.09	**.63**
	9:10 vs. 95:105	.45	**.20**	.35
Middle adolescents (N = 69)	9:10 vs. 85:95	**.49**	.25	.26
	9:10 vs. 90:100	.15	.05	**.80**
	9:10 vs. 95:105	.30	**.41**	.29
Adults (N = 34)	9:10 vs. 85:95	**.73**	.24	.03
	9:10 vs. 90:100	.03	.18	**.79**
	9:10 vs. 95:105	.12	**.73**	.15

Figure 9a and Figure 9b present mean proportions of correct answers graphically.

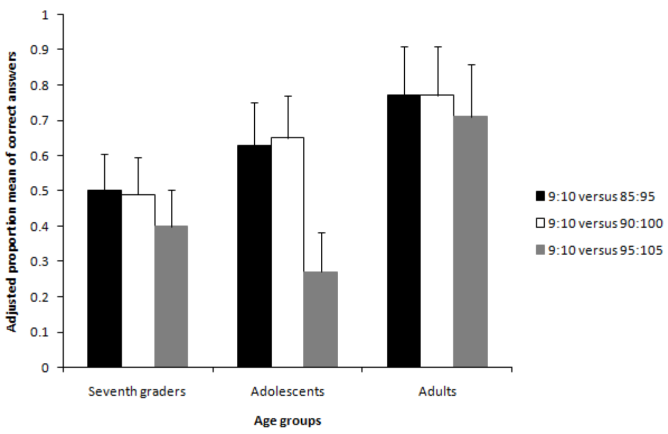

Figure 9a. Mean proportion of correct answers in the volleyball problem

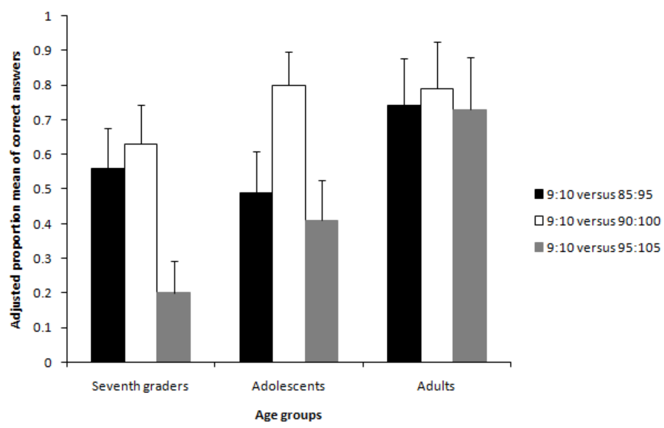

Figure 9b. Mean proportion of correct answers in the vaccination problem

A Generalized Linear Mixed Model (GLMM) was computed to investigate how age, trial, and problem influence the proportion of correct responses. As in Experiment 1, the data were analyzed using a GLMM for binomially distributed outcomes (correct response or wrong response). A GLMM was run with age, gender, trial, and problem as fixed effects. Participants were included in the model as a random factor to allow us to take into account the dependence between our observations due to repeated measures. The random intercept for participants allowed the intercept to vary between participants.

The Akaike information criterion AIC was employed as the model-selection method. A baseline model was constructed (with all possible interactions and main effects), and the best-fitting model was defined as the one minimizing the AIC. In order to verify that this most simple model is justified, we carried out a likelihood ratio test (e.g., Pinheiro & Bates, 2000, p. 83) that compares the most specific model with the more general model.

The best-fit model is summarized in Table 18, including regression coefficients, their standard errors, and the corresponding *Z* scores. The right side of Table 18 shows the odds ratios, their confidence intervals and the relative significance that the main effects and the interaction have in the model (Chi square). We compared the three age levels by fixing the seventh graders as the reference, comparing middle adolescents and adults with seventh graders; in addition, we compared the two problems by fixing the volleyball problem as the reference. Finally, because we observed in Table 18 the reversed pattern of seventh graders' and middle adolescents' answers with unequal ratios, we compared the three trial levels by fixing trial 9-in-10 versus 85-in-95 as the reference.

The three-way interaction between age, problem and trial and the two-way interaction between age and trial are both statistically significant. The two-way interaction (B = -1.08, Z = -2.17, p < .05) confirms that seventh graders are 2.94 times more likely than middle adolescents to give the correct answer in trial 3 (9-in-10 versus 95-in-105) compared to trial 1 (9-in-10 versus 85-in-95). The three-way interaction (B = 2.38, Z = 3.33, p < .001) confirms that the performance of seventh graders and middle adolescents in trial 1 and 3 is strongly mediated by the problem. Middle adolescents perform better than seventh graders in trial 3 compared to trial 1 of the vaccination problem. Conversely, seventh graders perform better than middle adolescents in trial 3 compared to trial 1 of the volleyball problem.

Table 18

Generalized Linear Mixed Model Parameters and Effect Sizes Predicting Correct Answers in Experiment 2

Best-fit model	*B*	(*SE*)	*Z*	OR	95% CI	Chisq (*df*)
Age						53.65 (2)***
Middle adolescents	0.52	(.34)	1.55	1.68	[.86, 3.28]	
Adults	1.22	(.46)	2.67***	3.39	[1.37, 8.34]	
Trial						54.48 (2)***
9:10 versus 90:100	-0.05	(.30)	-.15	0.95	[.53, 1.71]	
9:10 versus 95:105	-0.43	(.31)	-1.38	0.65	[.35, 1.19]	
Problem						0.48 (1)
Vaccination problem	0.25	(.32)	0.79	1.28	[.69, 2.40]	
Age x Trial						9.81 (4) *
Middle adolescents x 9:10 versus 90:100	0.12	(.48)	0.24	1.13	[.44, 2.89]	
Adults x 9:10 versus 90:100	0.05	(.65)	0.07	1.05	[.29, 3.76]	
Middle adolescents x 9:10 versus 95:105	-1.08	(.49)	-2.17*	0.34	[.13, .89]	
Adults x 9:10 versus 95:105	0.12	(.63)	0.20	1.13	[.33, 3.88]	
Age x Problem						.53 (2)
Middle adolescents x vaccination	-0.81	(.48)	-1.69	0.44	[.17, 1.14]	
Adults x vaccination	-0.45	(.65)	-0.70	0.64	[.18, 2.28]	
Problem x Trial						7.35 (2) *
Vaccination x 9:10 versus 90:100	0.34	(.46)	0.74	1.50	[.57, 3.46]	
Vaccination x 9:10 versus 95:105	-1.23	(.49)	-2.51*	0.29	[.11, .76]	
Age x Problem x Trial						26.70 (12) ***
Middle adolescents x vaccination x 9:10 versus 90:100	0.99	(.71)	1.40	2.69	[.67, 10.82]	
Adults x vaccination x 9:10 versus 90:100	-0.01	(.93)	-0.01	0.99	[.16, 6.13]	
Middle adolescents x vaccination x 9:10 versus 95:105	2.38	(.71)	3.33***	10.80	[2.69, 43.45]	
Adults x vaccination x 9:10 versus 95:105	1.53	(.92)	1.67	4.62	[.76, 28.03]	

Note. *p < .05, **p < .01, *** p <.001; Baseline categories: seventh graders for Age; 9-in-10 versus 85-in-95 for Trial; and volleyball for Problem

Note that the main effects of age, trial, and problem were not significant, except for the main effect comparing adults and seventh graders. Performance depended on a combination of the problem context, and different age groups exhibited different patterns of responses. These results suggest that participants modulate their reasoning strategies according to the specificity of the problem; if participants apply systematic reasoning errors, then correct answers may sometimes result from a biased reasoning process rather than quantitative reasoning.

4.3.2 Analysis of biased answers

To understand whether the bias changes with age and context we considered participants' choice after each trial and we categorized responses according to three levels: preference for less numerous (9-in-10), preference for more numerous (85-in-95, 90-in-100, or 95-in-105), and preference for *it is the same*. We used multinomial logistic regression to analyze the data. In the analysis, preference for less numerous is the comparison category and the preferences for more numerous and for *it is the same* are each compared with this reference, yielding estimates of the effects of the predictor variables (age, scenario, and trial) on the probabilities of the responses. We also compared the three age levels by fixing the seventh graders as the reference and comparing middle adolescents and adults with the seventh graders; the two scenarios by fixing the volleyball scenario as the reference; and trials by fixing trial 1 (9-in-10 versus 85-in-95) as the reference and comparing trial 2 (9-in-10 versus 90-in-100) and trial 3 (9-in-10 versus 95-in-105) with trial 1.

The Akaike information criterion AIC was employed as the model-selection method. The best-fit model took into account the main effects and the interactions between age and scenario and between age and trial. The best-fit model is summarized in the left side of Table 19, including the regression coefficients, their standard errors, and the corresponding *Z* scores. The right side of Table 19 shows the odds ratios, which measure effect size, and their 95% confidence interval (CI). The odds ratios are the antilog (i.e., exponentiated values) of the model coefficients.

The top half of Table 19 compares preference for the less numerous alternative (9-in-10) to preference for the more numerous alternative (85-in-95, or 90-in-100, or

95-in-105). This comparison is particularly relevant because it allows us to test directly whether the bias changes with age and scenario. The interaction between age and scenario clarifies this point. Middle adolescents (B = 1.48, Z = 3.84, $p < .001$) are more biased toward the less numerous (9-in-10) alternative than seventh graders in the volleyball problem but less biased in the vaccination problem. Adults (B = 1.29, Z = 2.54, $p < .01$) appear to have some bias toward the more numerous in the vaccination problem and little or no bias in the volleyball problem, whereas seventh graders are biased toward the less numerous alternative in both scenarios.

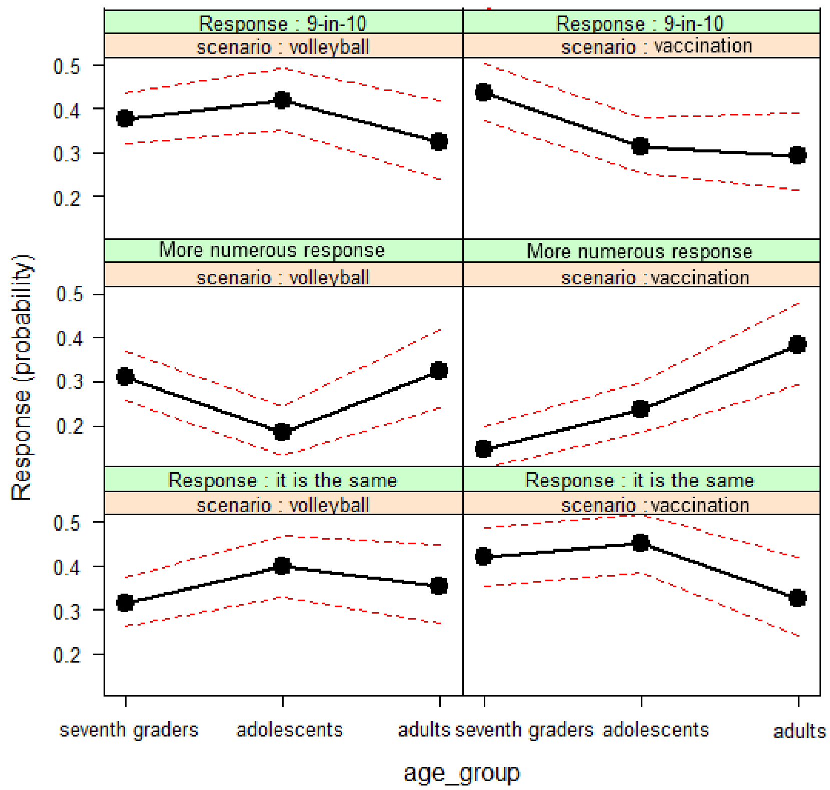

Figure 10. Graphical representation of the response choice made in each trial of both problems by all three age groups

Figure 10 shows the interaction between age and scenario. Black lines represent the probability of each response (less numerous, more numerous, or it is the same) according to age groups. Dashed red lines, instead, represent the confidence intervals of the odds ratios. Figure 10 clearly shows that in the volleyball problem the preference for the less numerous alternative is high especially in seventh graders and middle adolescents whereas the preference for the more numerous alternative decreases from seventh graders to middle adolescents.

Table 19

Multinomial Logistic Regression Model Parameters and Effect Sizes (N = 357)

Best-fit model	*B* (SE)	*Z*	OR	95% CI
Preference for 9-in-10 versus preference for the large ratio				
Age (Adolescents)	-.67 (.36)	-1.88	.51	[.25, 1.04]
Age (Adults)	-1.07 (.46)	-2.34*	.34	[.14, .85]
Trial (9-in-10 versus 90-in-100)	.63 (.32)	1.97*	1.88	[1.00, 3.52]
Trial (9-in-10 versus 95-in-105)	.63 (.28)	2.23*	1.88	[1.08, 3.25]
Scenario (Vaccination)	-.91 (.26)	-3.52***	.40	[.24, .67]
Age (Adolescents) X Trial (9-in-10 versus 90-in-100)	-.61 (.54)	-1.12	.54	[.19, 1.57]
Age (Adults) X Trial (9-in-10 versus 90-in-100)	1.84 (.74)	2.49*	6.30	[1.48, 26.85]
Age (Adolescents) X Trial (9-in-10 versus 95-in-105)	.36 (.42)	.87	1.43	[.63, 3.26]
Age (Adults) X Trial (9-in-10 versus 95-in-105)	2.69 (.58)	4.67***	14.73	[4.73, 45.92]
Age (Adolescents) X Scenario (Vaccination)	1.48 (.38)	3.84***	4.39	[2.09, 9.25]
Age (Adults) X Scenario (Vaccination)	1.29 (.51)	2.54**	3.63	[1.34, 9.87]
Preference for 9-in-10 versus preference for the response it is the same				
Age (Adolescents)	-.31 (.34)	-.91	.73	[.38, 1.43]
Age (Adults)	-1.53 (.57)	-2.68**	.22	[.07, .66]
Trial (9-in-10 versus 90-in-100)	1.52 (.27)	5.61***	4.57	[2.69, 7.76]
Trial (9-in-10 versus 95-in-105)	.31 (.27)	1.13	1.36	[.80, 2.31]
Scenario (Vaccination)	.16 (.22)	.74	1.17	[.76, 1.81]
Age (Adolescents) X Trial (9-in-10 versus 90-in-100)	.68 (.41)	1.65	1.97	[.88, 4.41]
Age (Adults) X Trial (9-in-10 versus 90-in-100)	3.42 (.75)	4.57***	30.57	[7.03, 132.95]
Age (Adolescents) X Trial (9-in-10 versus 95-in-105)	.55 (.41)	1.32	1.73	[.78,3 .87]
Age (Adults) X Trial (9-in-10 versus 95-in-105)	2.35 (.71)	3.29***	10.49	[2.61, 42.17]
Age (Adolescents) X Scenario (Vaccination)	.30 (.33)	.90	1.35	[.71, 2.58]
Age (Adults) X Scenario (Vaccination)	-.14 (.54)	-.25	.87	[.30, 2.51]

Note. OR = odds ratio; CI = confidence interval; *p < .05, **p <.01, ***p <.001

Instead, in the vaccination problem, seventh graders show the same preference for the less numerous alternative whereas the preference for the more numerous alternative increases with age. These findings suggest that, contrary to seventh graders, middle adolescents and adults rely on different scenario-based representations to give their response.

The bottom half of Table 19 compares preference for less numerous (9-in-10) to preference for the response *it is the same*. The main effect of trial is significant with regard to the comparison between 9-in-10 versus 90-in-100 and 9-in-10 versus 85-in-95 (B = 1.52, Z = 5.61, p < .001). This finding shows that, disregarding age and scenario, participants give a higher proportion of correct answers (it is the same) in trial 2 comparing 9-in-10 versus 90-in-100. In addition, the interaction between age and trial suggests that adults are 5 times better than seventh graders to solve trial 2 correctly compared to trial 1 (B = 3.42, Z = 4.57, p < .001).

Overall, these findings suggest that different processes are employed at different ages and in different contexts. To clarify which heuristics are activated, we present results about participants' written justifications.

4.3.3 Analysis of written preference explanations

To understand how participants reasoned, we analyzed the explanations that participants gave to their responses after each trial. Specifically, we coded their explanations in four categories:

1. Mathematical reasoning: participants who reasoned mathematically and explained their response through considerations of proportions and probabilities. For example "I solved the problem mathematically," or "statistically, one is better than the other," or "probabilities are the same/differ,"

2. Denominator neglect: participants who stated that they gave their response according to the absolute number of applicants or vaccinations. For example "higher sample seems safer" or "I would give myself a better chance of selection by being selected from a pool of 100 instead of 10";
3. Reasoning based on the absolute number of negative events: participants who stated that they gave their response according to the

smaller number of applicants/vaccinations. For example "in the small clinic the probabilities that the vaccination is not effective are less than in the big clinic";

4. Intuition: participants who explained their response referring to intuition, gut feelings, sensations or they were not able to provide a formal explanation. For example, "Just seems better" or "decided by gut feeling".

Table 20a and 20b show the proportions of correct and wrong responses associated with each category of explanation in the volleyball problem and the vaccination problem, respectively.

Table 20a

Volleyball Problem: Proportions of Correct and Wrong Responses Associated with Explanations Given by Participants of Each Age Group in Three Trials

9-in-10 versus 90-in-100	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 86)	.44	.06	—	.14	.01	.19	.03	.13
Middle adolescents (*N* = 62)	.61	.02	—	.05	.02	.18	.05	.15
Adults (*N* = 35)	.77	—	—	.14	—	.09	—	—
9-in-10 versus 85-in-95	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 86)	.09	.21	.01	.16	.31	.01	.08	.12
Middle adolescents (*N* = 62)	.25	.26	.02	.08	.15	.02	.23	.02
Adults (*N* = 35)	.54	.14	.03	.03	.11	.03	.09	.03
9-in-10 versus 95-in-105	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 86)	.14	.24	.12	.01	.01	.21	.13	.15
Middle adolescents (*N* = 62)	.14	.40	.03	.02	.02	.16	.08	.14
Adults (*N* = 35)	.63	.17	.03	—	.03	.07	.03	.03

Table 20b

Vaccination Problem: Proportions of Correct and Wrong Responses Associated with Explanations Given by Participants of Each Age Group in Three Trials

9-in-10 versus 90-in-100	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 71)	.61	.09	—	.04	—	.16	.03	.08
Middle adolescents (*N* = 69)	.77	—	—	.05	—	.10	.03	.08
Adults (*N* = 34)	.76	.03	—	.15	.03	.03	—	—
9-in-10 versus 85-in-95	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 71)	.17	.25	.01	.04	.30	.01	.08	.13
Middle adolescents (*N* = 69)	.20	.33	.01	.06	.07	.01	.20	.10
Adults (*N* = 34)	.53	.09	.03	.06	.12	.03	.06	.08
9-in-10 versus 95-in-105	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 71)	.10	.44	.03	.01	.01	.20	.06	.15
Middle adolescents (*N* = 69)	.19	.36	.03	.02	.01	.09	.17	.13
Adults (*N* = 34)	.56	.15	.09	.03	.03	.03	.06	.06

First, consider the categories of explanations for volleyball problem answers shown in Table 20a. In Trial 2 (9-in-10 versus 90-in-100), most participants of all three age groups explained their responses through mathematical reasoning, referring to probabilities and mathematical computation. For example, participants explained their correct choice of *it is the same* by saying, “because for both teams there was the same probability of joining,” or that “from a mathematical point of view, both answers lead to the same result.” Seventh graders and adolescents included calculations in their explanations more often than adults, writing, for example, “9 out of 10 is equal to 90 out of 100 and 900 out of 1,000, etc.”

The proportion of correct responses associated with mathematical reasoning increased with age, whereas denominator neglect associated with wrong answers was the same (14%) in seventh graders and adults but lower in adolescents (5%). A participant wrote, for example, "because there are 90 instead of 9 who can join the team, there is a good possibility that I can also join." Another participant wrote, "the larger group appeals to me more, I'm not sure why."

The reversed reasoning explanation according to CEST is due to the appeal of the less numerous option. We expected explanations such as, "the probabilities are the same but it is easier to see myself in a 10-person team than a 100-person team," or that "there are fewer participants in team A that I would have to compete with." Instead, explanations generally focused on the absolute number of candidates who do *not* pass the selection. For example, one wrote that 9-in-10 is better than 90-in-100 because "in the first case only one person would not be selected whereas in the second case 10 people will be left out." Another wrote, "of 10 athletes 9 will be selected and therefore only 1 will be left out, while in the other there are 100 athletes of which 90 will be left out." This kind of reasoning was most often applied in trials with unequal ratios, suggesting that participants compared 1-in-10 to 9-in-85 or 9-in-95 by focusing on the numerators and neglecting the denominators. FTT might explain these results as due to inverse denominator neglect: With highly favorable probability events, people reason about the low probability of unfavorable events.

Finally, and nicely, we noticed that seventh graders mentioned more than adolescents and adults that "if there are fewer players then it is easier or more likely to play." Some adolescents preferred the small team more than seventh graders and adults because they think the small team is more challenging. One wrote, for example, "I gave this response because if few people are selected and they select you, that means that you really have talent."

Second, consider the categories of explanations for vaccination problem answers shown in Table 20b. Again, most participants relied on mathematical reasoning. With unequal ratios (trials 2 and 3) seventh graders and middle adolescents made mistakes in their calculations. They wrote, for example, "according to my mathematical calculations, I found the answer …." A small proportion of participants in all three trials preferred the larger clinic and offered a denominator neglect explanation. One wrote, for example, "I would like to go to clinics with more experience," and another wrote, "where there are more vaccinations the probability is greater that the vaccine will have an effect."

The reversed reasoning explanation was given more often with unequal ratios than with equal ratios, and given more by seventh graders than middle adolescents and adults. Participants justified their preference for the small clinic by reasoning about the probability that the vaccine will not be effective. One participant said, "out of 10 there is only 1 vaccine that will not be effective, but in the second clinic there are 10 that will not be effective."

4.3.4 Predictors of reasoning performance

To understand the dependence of correct responses on reasoning processes, we investigated participants' individual differences. Do relationships exist between good reasoning and cognitive capacity as well as between good reasoning and thinking dispositions? We present the results divided by age group because we applied different measures.

Seventh grade arithmetical reasoning performance

After participants performed the assigned problem, their mathematical ability was assessed using the AC-MT 11-14 (Cornoldi & Cazzola, 2003). We investigated their written computation ability, their comprehension and production ability, and their arithmetical reasoning ability. Table 21 presents seventh graders' performance at AC-MT 11-14.

Table 21

Descriptive Statistics for Seventh Graders AC-MT 11-14 (N = 157)

	Min	Max	*M*	*SD*	Normative scores
Written calculation	0	10	5.50	2.55	6.34
Comprehension and production	2	20	12.94	4.86	14.40
Arithmetical reasoning	5	31	16.56	5.93	19.09
AC-MT 11-14 total score	15	75	45.99	13.83	52.26

According to the performance standards established by Cornoldi and Cazzola (2003), 14% of participants need immediate intervention, 29.3% of participants need attention, 54.1% of participants performed sufficiently, and only 2.5% had excellent mathematical ability. Figure 11 shows the density distribution of the total number of correct responses (range = 0-3) for each of these four levels of performance. Overall, the four distributions overlap but those participants with an excellent mathematical ability responded correctly two or three times, substantially more often than other participants.

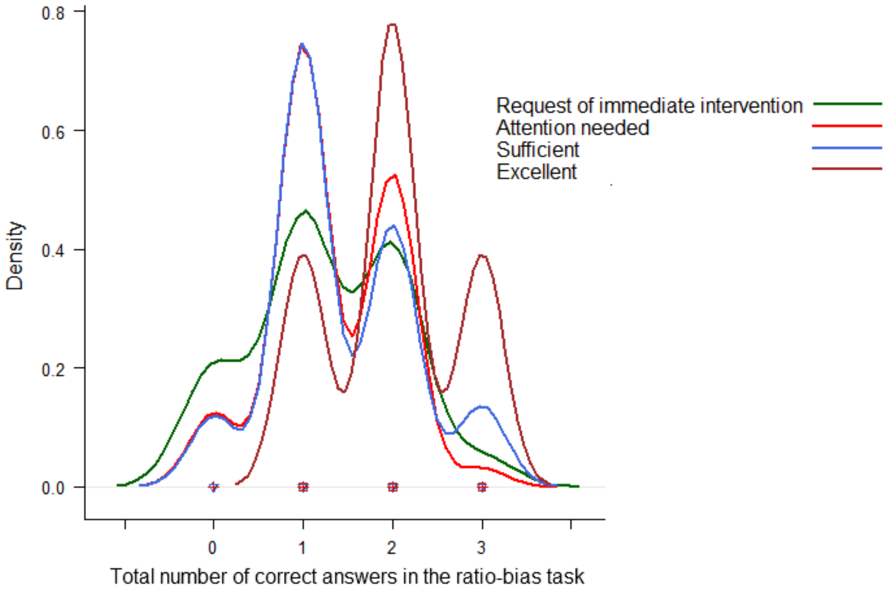

Figure 11. Density plots for correct responses grouped by seventh grade mathematical level according to the AC-MT 11-14

Table 22a and 22b show the correlations between correct responses for each trial of the ratio-bias task, total correct responses in the ratio-bias task (range: 0-3), and the AC-MT (subscales and total score) for the volleyball and vaccination problems, respectively.

Table 22a

Seventh Grade Tetrachoric, Polychoric and Pearson Correlations Between Correct Responses in the Volleyball Problem and AC-MT 11-14 (N = 86)

Variables	1	2	3	4	5	6	7	8
Ratio-bias task								
1. 9-in-10 versus 85-in-95	—							
2. 9-in-10 versus 90-in-100	.15	—						
3. 9-in-10 versus 95-in-105	-.22	-.19	—					
4. Total correct responses	.58**	.55**	.40**	—				
AC-MT 11-14								
5. Written calculation	.25*	-.06	-.14	.06	—			
6. Comprehension and production	.08	-.10	-.08	-.05	.43**	—		
7. Arithmetical reasoning	.05	.18	.14	.23*	.13	.20	—	
8. AC-MT 11-14 total score	.19	.02	-.04	.12	.78**	.71**	.63**	—

*$p < .05$; **$p < .01$

Table 22b

Seventh Grade Tetrachoric, Polychoric and Pearson Correlations Between Correct Responses in the Vaccination Problem and AC-MT 11-14 (N = 71)

Variables	1	2	3	4	5	6	7	8
Ratio-bias task								
1. 9-in-10 versus 85-in-95	—							
2. 9-in-10 versus 90-in-100	-.04	—						
3. 9-in-10 versus 95-in-105	.22	-.28*	—					
4. Total correct responses	.69**	.48**	.51**	—				
AC-MT 11-14								
5. Written calculation	.11	.04	.04	.10	—			
6. Comprehension and production	.01	.22	-.01	.12	.57**	—		
7. Arithmetical reasoning	.07	.04	-.03	.04	.32**	.41**	—	
8. AC-MT 11-14 total score	.09	.11	.01	.11	.86**	.80**	.69**	—

*$p < .05$; **$p < .01$

In the volleyball problem there were no significant correlations among the three trials of the ratio-bias task. In the vaccination problem the only significant correlation (-.28) was between trials 2 and 3. These low correlations suggest that participants may have employed somewhat different strategies or processes in different trials.

Total correct responses were positively correlated (.23) only with arithmetical reasoning in the volleyball problem and with no ability measures in the vaccination problem. These findings suggest either that the AC-MT 11-14 does not measure the mathematical ability required to respond correctly in the ratio-bias task or that mathematical ability is not an important contributor to performance in the ratio-bias task.

Middle adolescent reasoning performance

The tests used with middle adolescents were the PMA, objective numeracy, subjective numeracy, cognitive reflection ability, need for cognition and faith in intuition. Table 23 summaries descriptive statistics for each scale.

Table 23

Means, Standard Deviations and Reliabilities of the PMA, Objective Numeracy, Subjective Numeracy, Need for Cognition, Faith in Intuition, and CRT (N = 131)

	Min	Max	*M*	*SD*	Cronbach's *α*
PMA					
Spatial ability	3	52	22.76	11.06	—
Verbal ability	8	42	26.21	6.60	—
Calculation ability	7	48	21.64	9.85	—
Objective numeracy total score	0	11	7.02	2.41	.71
General numeracy	0	3	1.55	.99	.32
Expanded numeracy	0	8	5.47	1.84	.67
Subjective numeracy general mean	1.00	5.00	3.34	.65	.71
Cognitive ability	1.00	5.00	3.60	.76	.76
Preference for information display	1.00	5.00	3.07	.83	.56
Cognitive reflection task total score	0	3	0.61	0.92	—
Need for cognition general mean	1.90	4.40	3.11	0.52	.76
Rational ability	1.70	4.50	3.14	0.55	.42
Rational engagement	1.70	4.40	3.09	0.59	.75
Faith in intuition general mean	2.20	4.65	3.14	0.48	.78
Experiential ability	2.10	4.70	3.14	0.50	.60
Experiential engagement	1.60	4.80	3.16	0.56	.69

Tables 24a and 24b show the correlations for the volleyball problem and vaccination problem, respectively, between correct responses in the ratio-bias task and each scale and subscale.

Table 24a

Middle Adolescent Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Volleyball Problem (N = 62)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ratio-bias task																				
1. 9-in-10 vs. 85-in-95	—	-.42**	-.06	.52**	-.10	-.27*	.23	-.13	-.12	-.10	-.04	-.07	-.01	-.02	-.31*	-.34**	-.22	.01	-.01	.02
2. 9-in-10 vs. 90-in-100		—	.57**	.52**	.47**	.05	-.07	.41**	.46**	.29*	.13	.23	.01	.35**	.22	.25*	.16	.21	.20	.18
3. 9-in-10 vs. 95-in-105			—	.66**	.15	-.20	-.08	.31*	.33*	.21	.29*	.37**	.15	.39**	.09	.20	-.01	.15	.15	.12
4. Total correct responses (range: 0-3)				—	.12	-.23	.13	.22	.26*	.14	.13	.19	.04	.29*	-.04	-.04	-.04	.13	.11	.11
PMA																				
5. Spatial ability					—	.32**	-.12	.39**	.29*	.34**	.09	.16	.01	.25*	.06	.13	-.02	.30*	.24	.29*
6. Verbal ability						—	-.22	.27*	.21	.24	.05	-.05	.11	.09	.24*	.33**	.12	.06	.18	.08
7. Calculation ability							—	.28*	-.14	-.28*	-.24*	-.11	-.26*	.01	-.11	-.15	-.05	-.08	-.13	-.03
Objective Numeracy Scale																				
8. Objective numeracy general score								—	.68**	.92**	.29*	.25*	.22	.37**	.05	.21	-.11	.32**	.20	.36**
9. General numeracy									—	.34*	.18	.22	-.08	.40**	-.01	.09	-.09	.19	.13	.20
10. Expanded numeracy										—	.27*	.20	.24*	.27	.07	.22	-.10	.30*	.18	.35**
Subjective Numeracy Scale																				
11. Subjective numeracy general score											—	.76**	.86**	.12	.34**	.38**	.25*	.05	.04	.05
12. Cognitive ability												—	.31*	.24	.11	.22	-.02	.17	.21	.10
13. Information display													—	-.02	.41**	.38**	.38**	-.06	-.10	-.01
Cognitive Reflection Task																				
14. Cognitive reflection general score														—	-.05	.03	.12	.13	.08	.15
Need for Cognition Scale																				
15. General need for cognition score															—	.90**	.92**	-.17	.09	-.20
16. Rational ability																—	.66**	.05	.08	.02
17. Rational engagement																	—	-.34**	-.24*	-.37**
Faith in Intuition Scale																				
18. Faith in intuition general score																		—	.88**	.91**
19. Experiential ability																			—	.60**
20. Experiential engagement																				—

Note. $^*p < .05$; $^{**}p < .01$

Table 24b

Middle Adolescent Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Vaccination Problem (N = 69)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ratio-bias task																				
1. 9-in-10 vs. 85-in-95	—	-.07	.09	.60**	.08	.20	.13	-.24	-.16	.22	.08	.02	.12	.08	-.20	.07	-.28*	.12	.16	.06
2. 9-in-10 vs. 90-in-100		—	.16	.60**	-.08	-.15	-.02	.06	.20	-.02	.08	.12	-.01	.02	.14	.13	.12	.18	.14	.19
3. 9-in-10 vs. 95-in-105			—	..63**	.31**	.05	-.13	.11	.15	.06	.09	-.04	.21	.29*	.18	.29*	.05	.16	.10	.20
4. Total correct responses (range: 0-3)				—	.17	.07	.01	-.06	.07	.12	.12	.04	.16	.10	.04	.17	-.08	.21	.18	.20
PMA																				
5. Spatial ability					—	.17	-.13	.33**	.37**	.24*	.17	-.02	.30*	.45**	.22	.22	.18	.15	.13	.15
6. Verbal ability						—	-.34**	.22	.17	.20	-.08	-.08	-.05	-.10	-.18	-.27*	-.05	-.02	-.03	-.01
7. Calculation ability							—	.03	.17	-.06	.07	.05	.06	.11	.24*	-.16	-.27*	-.09	-.06	-.11
Objective Numeracy Scale																				
8. Objective numeracy general score								—	.74**	.93**	.12	.06	.14	.29*	-.01	-.07	-.10	.10	.11	.07
9. General numeracy									—	.44**	.20	.17	.16	.40**	.08	.09	.07	.02	.01	.03
10. Expanded numeracy										—	.05	-.01	.11	.16	.13	-.07	-.18	.13	.15	.09
Subjective Numeracy Scale																				
11. Subjective numeracy general score											—	.83**	.80**	.35*	.40**	.26*	.27*	.09	.21	.03
12. Cognitive ability												—	.33**	.37**	.33**	.35**	.25*	.13	.23	.02
13. Information display													—	20	.33**	.40**	.19	.02	.11	-.06
Cognitive Reflection Task																				
14. Cognitive reflection general score														—	.31**	.25*	.32**	.12	.16	.07
Need for Cognition Scale																				
15. General need for cognition score															—	.90**	.91**	.14	.27*	.01
16. Rational ability																—	.64**	.08	.26*	-.09
17. Rational engagement																	—	.17	.23	.07
Faith in Intuition Scale																				
18. Faith in intuition general score																		—	.90**	.92**
19. Experiential ability																			—	.65**
20. Experiential engagement																				—

Note. $^{*}p < .05$; $^{**}p < .01$

First, consider the correlations between accuracy in the volleyball problem and these ability measures. Accuracy in trial 1 of the ratio-bias task was correlated negatively with PMA Verbal Ability (-.27) and two subscales of Need for Cognition (-.31 and -.34). Accuracy in trial 2 was positively correlated with all subscales of Objective Numeracy (.41, .46, and .29), with the Cognitive Reflection Task (.35), and a subscale of Need for Cognition (.25). Trial 3 was positively correlated with two subscales of Objective Numeracy (.31 and .33), with two subscales of Subjective Numeracy (.29 and .37), and with the Cognitive Reflection Task (.39). The correlations with trials 2 and 3 were similar but correlations with trial 1 were strikingly different.

Second, consider the correlations between accuracy in the vaccination problem and the ability measures. Total accuracy in the vaccination problem and accuracy in trial 2 were not significantly correlated with any measure. Accuracy in trial 1 was correlated negatively with a subscale of Need for Cognition (-.28), and accuracy in trial 3 was correlated positively with the Cognitive Reflection task (.29) and a subscale of Need for Cognition (.29). There appears to be little relationship between performance in the ratio-bias task and measures of cognitive abilities and thinking dispositions.

Adult reasoning performance

The tests used with adults were the RAPM, objective numeracy, subjective numeracy, cognitive reflection ability, go/no-go task, need for cognition and faith in intuition. Table 25 summaries the descriptive statistics for each scale.

Table 25

Means, Standard Deviations and Reliabilities of the RAPM, Objective numeracy, Subjective numeracy, CRT, go/no-go task, Need for Cognition, and Faith in Intuition (N = 69)

	Min	Max	*M*	*SD*	Cronbach's α
RAPM					
Total score	2	12	9.46	2.39	—
Objective numeracy total score	6	11	9.97	1.19	.30
General numeracy	0	3	2.39	.88	.52
Expanded numeracy	6	8	7.58	.60	.23
Subjective numeracy general mean	1.88	5.00	3.78	.75	.85
Cognitive ability	1.75	5.00	3.79	.87	.85
Preference for information display	1.75	5.00	3.767	.80	.71
Cognitive reflection task total score	0	3	1.62	1.07	—
Go/no-go task					
Reaction time (ms)	228.61	499.27	306.21	46.77	—
Correct answers	56	273	236.27	44.46	—
False alarms	0	67	21.83	11.69	—
Need for cognition general mean	2.35	4.90	3.64	0.64	.92
Rational ability	2.40	5.00	3.59	0.67	.87
Rational engagement	2.10	5.00	3.69	0.74	.89
Faith in intuition general mean	2.00	4.70	3.29	0.60	.90
Experiential ability	2.00	4.70	3.29	0.60	.84
Experiential engagement	1.70	4.80	3.26	0.63	.86

Tables 26a and 26b show the correlations for the volleyball problem and vaccination problem, respectively, between correct responses in the ratio-bias task and each scale and subscale.

First, consider the correlations between accuracy in the volleyball problem and these ability measures. Total accuracy in the ratio-bias task was positively correlated with cognitive ability as measured by the Raven APM (.48) and a subscale of Need for Cognition (.40). It was also negatively correlated with a subscale of the Go/no-go task (-.38). Accuracy in trials 1 and 2 was correlated with few of the ability measures, but accuracy in trial 3 was significantly correlated with nine of them.

Table 26a

Adult Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Volleyball Problem (N = 35)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Ratio-bias task																					
1. 9-in-10 vs. 85-in-95	—	.10	.06	.54**	.24	-.19	-.09	-.24	.15	.13	.13	-.27	.06	-.48**	.03	.31	.19	.32	.22	-.42*	-.16
2. 9-in-10 vs. 90-in-100		—	.20	.69**	.33*	-.22	-.28	-.05	-.29	-.28	-.22	.12	.26	-.33*	-.35*	-.04	-.14	.06	.06	-.06	-.06
3. 9-in-10 vs. 95-in-105			—	..63**	.40*	.39*	.55**	.01	.62**	.57**	.53**	.11	.09	-.17	-.18	.51**	.62**	.23	-.12	-.12	-.33*
4. Total correct responses				—	.48**	.04	.09	-.06	.26	.22	.23	-.01	.11	-.38*	-.18	.32	.40**	.16	-.22	-.22	-.12
RAPM																					
5. RAPM total score					—	.13	.20	.02	.31	.30	.26	.11	-.05	.06	-.50**	.37*	.32	.34*	-.28	-.28	-.29
Objective Numeracy Scale																					
6. Objective numeracy general score						—	.84**	.78**	.48**	.45**	.39*	-.04	-.18	.14	.10	.26	.31	.15	.16	.16	.20
7. General numeracy							—	.25	.60**	.53**	.53**	-.03	.01	-.02	-.07	.37*	.41*	.25	.06	.06	.03
8. Expanded numeracy								—	.10	.14	.04	-.11	-.30	.29	.27	-.01	.03	.04	.21	.21	.32
Subjective Numeracy Scale																					
9. Subjective numeracy general score									—	.89**	.87**	.04	-.08	-.09	.05	.34*	.44*	.17	.12	.12	.03
10. Cognitive ability										—	.56**	-.14	-.12	-.06	.14	.38*	.45**	.22	.14	.14	-.01
11. Information display											—	.07	-.02	-.11	-.06	.22	.32	.07	.07	.07	.05
Cognitive Reflection Task																					
12. Cognitive reflection general score												—	.38*	-.21	-.34*	-.19*	-.15	-.18	-.10	-.10	-05
Go/no-go task																					
13. Reaction Time													—	-.70**	-.54**	-.10	-.04	-.14	.02	.02	.11
14. Proportion of correct responses														—	.10	-.05	-.03	-.05	.06	.06	-.21
15. Proportion of false alarm															—	-.14	-.10	-.14	.17	.17	.33*
Need for Cognition Scale																					
16. General need for cognition score																—	.90**	.88**	.07	.07	-.14
17. Rational ability																	—	.58**	.03	.03	-.28
18. Rational engagement																		—	.10	.10	.03
Faith in Intuition Scale																					
19. Faith in intuition general score																			—	.99**	.67**
20. Experiential ability																				—	.67**
21. Experiential engagement																					—

Note. *$p < .05$; **$p < .01$

Table 26b

Adult Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Vaccination Problem (N = 34)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Ratio-bias task																					
1. 9-in-10 vs. 85-in-95	—	.59**	-.06	.67**	.10	.19	.25	.01	.29*	.30*	.22	.36*	.16	-.20	-.10	.24	.19	.24	.17	.17	.44**
2. 9-in-10 vs. 90-in-100		—	.22	.54**	.28	.37*	.22	.26	.48**	.48**	.39**	.41**	-.09	.28	.01	.33*	.22	.36**	.11	.11	.24
3. 9-in-10 vs. 95-in-105			—	.59**	.32*	.08	.02	.13	.13	.19	.03	.34*	.31*	-.10	-.28	-.05	.10	-.17	-.46**	-.46**	-.56**
4. Total correct responses				—	.39*	.31*	.24	.26	.37*	.43**	.23	.58**	.26	-.23	-.23	-16	.16	.13	-.10	-.10	-.03
RAPM																					
5. RAPM total score					—	.47**	.27	.54**	.29*	.38*	.13	.43**	-.01	-.08	-.31*	.17	.16	.14	.10	.10	-.13
Objective Numeracy Scale																					
6. Objective numeracy general score						—	.88**	.69**	.62**	.56**	.55**	.63**	-.20	-.17	-.08	.32*	.29*	.29*	.19	.19	-.01
7. General numeracy							—	.26	.65**	.55**	.64**	.57**	-.26	-.12	-.06	.30*	.31*	.24	.07	.07	-.11
8. Expanded numeracy								—	.25	.30*	.14	.41*	-.01	-.15	-.07	.19	.10	.24	.29*	.29*	.14
Subjective Numeracy Scale																					
9. Subjective numeracy general score									—	.92**	.89**	.55*	-.17	-.17	-.30*	.61**	.62**	.51**	.07	.07	-.18
10. Cognitive ability										—	.64**	.57**	-.15	-.08	-.38*	.63**	.58**	.57**	.07	.07	-.13
11. Information display											—	.43**	-.16	-.23	-.16	.47**	.55**	.34*	-.07	-.07	-.19
Cognitive Reflection Task																					
12. Cognitive reflection general score												—	-.04	-.07	-.19	.44**	.32*	.47**	.01	.01	-.01
Go/no-go task																					
13. Reaction Time													—	-.45**	-.30*	-.39*	-.40**	-.32*	-.12	-.12	.09
14. Proportion of correct responses														—	.12	.01	-.01	.03	-.03	-.03	.09
15. Proportion of false alarm															—	.01	.09	-.06	.11	.11	.02
Need for Cognition Scale																					
16. General need for cognition score																—	.90**	.93**	.18	.18	-.13
17. Rational ability																	—	.67**	.03	.03	-.40*
18. Rational engagement																		—	.28	.28	.11
Faith in Intuition Scale																					
19. Faith in intuition general score																			—	.99**	.62**
20. Experiential ability																				—	.62**
21. Experiential engagement																					—

Note. $*p < .05$; $**p < .01$

Second, consider the correlations between accuracy in the vaccination problem and these ability measures. Again, total accuracy in the ratio-bias task was positively correlated with cognitive ability as measured by the Raven APM (.39). It was also positively correlated with a subscale of Object Numeracy (.31), two subscales of Subjective Numeracy (.37 and .43), the Cognitive Reflection task (.58), and a subscale of Faith in Intuition (.56). Accuracy in trials 1, 2, and 3 was also correlated with many of these measures.

4.4 Discussion

In this experiment we introduced two everyday-life scenarios to understand whether contextualized situations increase the strength of heuristic representations and whether such representations develop with age. According to our predictions, everyday-life situations activate different levels of expertise and familiarity. Seventh graders did not show a clear bias toward one direction in the volleyball problem. Taken together the analysis of biased answers and the analysis of justifications, results indicate that, according to CEST, seventh graders had a preference for the small team because of the smaller number of *not* selected applicants. At the same time, a smaller proportion of seventh graders prefer the large team. In the vaccination problem, seventh graders exhibited a clear preference for the small clinic because of the absolute number of *not* effective vaccinations.

According to previous results (Alonso & Berrocal, 2003), middle adolescents exhibited a clearer preference for the small team than seventh graders in the volleyball problem. Such preference partially moves toward the more numerous alternative in the vaccination problem, suggesting that adolescents activate more than seventh graders context-related heuristics. As FTT theorists claim, such findings suggest that children change from quantitative reasoners (verbatim-based) to qualitative reasoners (gist-based) as they grow up. However, differently from FTT predictions, biased answers do not increase with age but, rather, they change according to the context and to numerical information presented.

As in Experiment 1, Cornell undergraduates gave an excellent performance suggesting that highly educated adults reason correctly on ratios and probabilities disregarding the context. In the volleyball problem adults showed no bias at all; just few adults, instead, showed a preference for the large clinic motivating that a big clinic is a guarantee of high level of experience.

With regard to individual differences, we found few positive correlations only in the condition of the volleyball problem: a) a small correlation (.25) in seventh graders between correct responses in the ratio-bias task and written calculation of the AC-MT 11-14; b) two small correlations in middle adolescents between correct responses in the ratio-bias task and general numeracy (.26), and between correct responses in the ratio-bias task and the Cognitive Reflection Task (.29).

As in Experiment 1, the total number of correct responses in the ratio-bias task for adults is negatively related to the proportion of correct answers in the go/no-go task in the volleyball problem (-.38). In the vaccination problem this correlation is negative (-.23) but not statistically significant. This finding suggests that participants might experienced the volleyball problem differently compared to the vaccination problem. This difference could be due to cultural differences. For example, in the American environment, volleyball might be less popular than in Europe. We speculate that the volleyball problem might be considered less contextualized compared to the vaccination problem. Consequently, if the vaccination problem generates more heuristic answers than the volleyball problem, then a time-paced condition discriminate less those participants who need time to decide.

Chapter 5

Experiment 3: The ratio-bias task with time pressure

5.1 Introduction

In Experiment 1 we showed that different age groups perform the ratio-bias task using different processes, and in Experiment 2 we showed that different scenarios activate different heuristics. In light of De Neys and Vanderputte (2010), these findings suggest that heuristic processes develop with age and activate context-based representations. The low correlations between accuracy in the volleyball and vaccination problems and the cognitive ability measures pose questions about the interactions between heuristic and analytic processing. According to traditional dual-process theories, rapid heuristic processes determine biased answers by default, slow and effortful analytic process override biased answers, and, if abilities related to analytic processes are strong enough, then the subject gives the correct answer. The results of Experiment 2 suggest that sometimes heuristic representations are stronger than analytic abilities or that heuristic representations work disregarding analytic abilities. Here we argue that time is a critical element for clarifying this point.

As Evans and Holmes (2005) underline, the rapid-response reasoning task is a relevant methodological innovation reported by Roberts and Newton (2002). The rationale is that constraining participants to respond within a short period of time inhibits slower analytic reasoning processes. In accordance with traditional dual-process theories, Evans and Holmes predicted that, when required to respond rapidly, participants in a belief-bias in syllogistic reasoning task would show: a) an increased level of belief bias and b) a reduced level of logical responding.

They showed that measures of belief bias, characterized as a within-participant conflict between logic-based (analytic) processes and belief-based (heuristic) processes, increase significantly when the Wason selection task is performed under time pressure. They conclude that the rapid-response reasoning methodology is a powerful technique that may be capable of shedding new theoretical light on a range of different reasoning tasks. In accordance with this view, we should expect that participants required to respond rapidly in ratio-bias tasks should be more likely to respond with biases in all scenarios and age groups.

There is, however, evidence that thinking too much can sometimes lead to poorer choices (Wilson & Schooler, 1991). For example, Dijksterhuis and Bos (2005) showed that conscious thought leads to more stereotyping than does unconscious thought. Recently, Dijksterhuis and colleagues formulated the Unconscious Thought Theory (UTT), which distinguishes between two modes of thought: unconscious and conscious (Dijksterhuis & Nordgren, 2006). This theory shares some relevant principles with FTT, such as the bottom-up direction of gist processing.

A central principle of UTT is that consciousness has limited capacity and, consequently, conscious deliberation often leads to relatively poor decisions when many factors ought to be considered and the decision is, therefore, complex. Conversely, unconscious thought, or thought without attention, can lead to good choices (Dijksterhuis, 2004; Dijksterhuis, Nordgren, & Baaren, 2006). Specifically, people make better decisions when they are distracted for a period of time than when they are asked either to think about their choice (Dijksterhuis, Nordgren, & Baaren, 2006) or to respond immediately (Dijksterhuis, 2004). The paradigm that Dijksterhuis and colleagues used is similar in different studies (Acker, 2008). Participants are presented with a number of pieces of information about a number of options to choose from. For example, four cars are described with 12 attributes each. One is the best car, with eight positive attributes and four negative attributes, one is the worst, with four positive attributes and eight negative attributes, and the other cars are intermediate, with both six positive and six negative attributes. Participants evaluate the goodness of cars on the basis of these attributes (Bos, Dijksterhuis, & van Baaren, 2008; Dijksterhuis, Nordgren, & Baaren, 2006).

These experiments clearly involve memory. Participants have to remember the attributes to judge the cars correctly. Our decision-making task, in contrast, involves capacities and heuristics that are acquired before the experiment, although the computations required to compare ratios may place demands on working memory. Therefore, Dijksterhuis and colleagues’ results cannot be directly applied to our experiments.

A key implication of both UTT and FTT, however, is that time pressure may not increase the frequency of biased answers. According to traditional dual-process theories, time pressure does not provide sufficient time for analytic processes to inhibit heuristic processes. According to UTT and FTT, to the contrary, thinking too much may generate more heuristics. If UTT and FTT are correct, then time pressure in the ratio-bias task should: a) reduce context-related differences in all age groups; b) increase reliance on comparing ratios.

5.2 Method

Participants. Participants were 92 seventh graders (50 females, 42 males) ranging in age from 12 to 14 years (mean age = 12.77 years, *SD* = 0.56 years), 98 middle adolescents (60 females, 38 males) ranging in age from 14 to 16 (mean age = 14.87 years, *SD* = 0.47 years) and 92 adults (60 females, 32 males) ranging in age from 18 to 37 (mean age = 20.93 years, *SD* = 3.18 years). Seventh graders and middle adolescents were recruited from public schools in the northeast part of Italy; all participants were typically developing children of middle socioeconomic status. Adults were undergraduate students at Cornell University (Ithaca, NY) and they were recruited in some psychology courses and via campus postings. All participants gave written informed consent and the part of the study which took place at Cornell University was approved by the Institutional Review Boards of Cornell University.

Material and procedure. This experiment was designed and run using Qualtrics.com online survey software (Qualtrics Labs Inc., Provo, UT). All respondents participated in the experiment online by accessing Qualtrics's secure survey site. Seventh graders and middle adolescents performed the experiment in their school's computer lab during lesson time. Cornell undergraduates signed up for the experiment using "Susan" (http://susan.psych.cornell.edu/) and received one class credit for completing the experiment. To protect participants' anonymity their data was identified by a sequentially generated participation number; their names were not recorded.

All participants performed the same three trials of the ratio-bias task used in Experiments 1 and 2. The presentation order of the three trials was randomized.

Participants in each age group were randomly assigned to one of three groups. One group (23 seventh graders, 35 adolescents, and 32 adults) solved the mathematical problem used in Experiment 1, the second group (40 seventh graders, 27 adolescents, and 29 adults) solved the volleyball problem of Experiment 2, and the third group (29 seventh graders, 36 adolescents, and 31 adults) solved the vaccination problem.

First participants read the problem. We controlled the reading times according to the speed-reading ability of each age group as in Experiments 1 and 2. The time available to read the problem was specified in the instructions and a countdown was visible at the bottom of the page indicating how much time participants had remaining to read and understand the problem.

Next the three options were presented and they had to choose one of them. Unlike in Experiments 1 and 2, participants had limited time to respond. The time limit was the estimated time to read the question and options plus 5 seconds. For example, the mathematical problem consists of 72 syllables, and expected reading time for seventh graders is 20 seconds. Thus, seventh graders had 25 seconds to read and respond. The time available to read and answer was specified in the instructions and a countdown was visible at the bottom of the page indicating how much time participants had remaining to read and respond to the problem.

After participants selected a response, they were asked to explain their choice. The question was: "Can you explain the basis for your response?" Finally, demographic questions requested information about Cornell undergraduates' age, gender ethnicity and student status. Seventh graders and middle adolescents were asked about age, gender, and their practice with computers or videogames. They were asked to rate their level of practice with the computer (e.g., "how often do you use your computer?"; "how often do you use videogames?").

5.2.1 Individual differences measures

Seventh graders' mathematical ability was assessed using the AC-MT 11-14 (Cornoldi & Cazzola, 2003) as in Experiments 1 and 2. Middle adolescents were administered the Primary Mental Ability test (P.M.A.; Thurstone & Thurstone, 1963). Middle adolescents and adults completed the same scales used in Experiments 1 and 2 assessing general intelligence, numeracy, thinking style and cognitive inhibition. Only adults also performed the Go/no-go task (Garavan, Ross, Murphy, Roche, & Stein, 2002) to test their ability in inhibitory control.

5.3 Results

5.3.1 Analysis of correct responses

Figure 12 presents the density plots of the total number of correct responses for each age group in the mathematical problem, volleyball problem, and vaccination problem. As Figure 12 shows, adults perform better than the other two age groups but their performance seems to decrease in the everyday-life problems compared to the mathematical problem. Seventh graders and middle adolescents had very similar performance in the mathematical problem. In the other two problems, the middle adolescents appear to perform better on average than the seventh graders, especially in the volleyball problem, where most seventh graders had just one correct response.

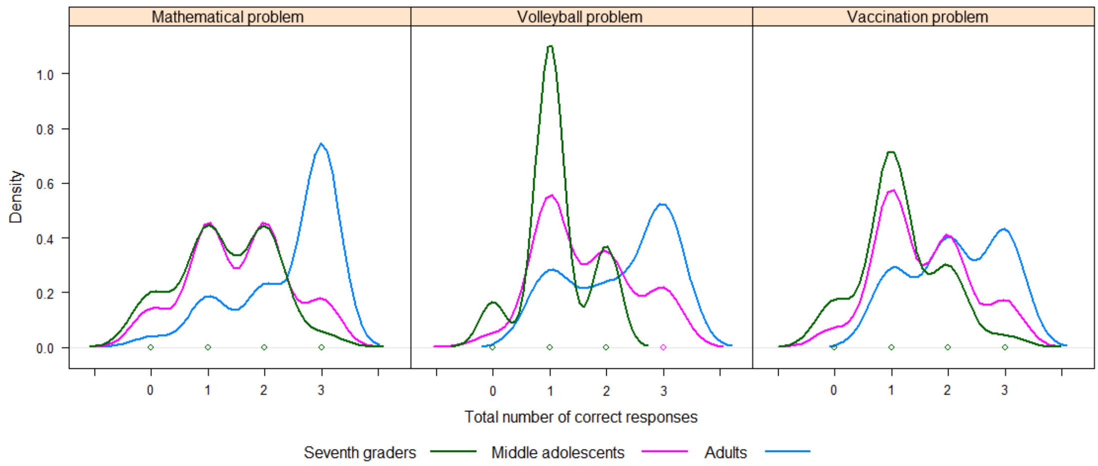

Figure 12. Density plots of the number of correct responses in each trial grouped by age for each problem

Table 27a presents the proportion of responses given by each age group to each trial in the mathematical problem; Table 27b shows these proportions for the volleyball problem; and Table 27c shows these proportions for the vaccination problem. In each problem there was a consistent increase in performance with age.

Table 27a

Proportion of Preferences (Correct Responses in Bold) in The Mathematical Problem

Age group	Trial	Preference		
		Container A (less numerous)	Container B (more numerous)	No difference
Seventh graders (*N* = 23)	9:10 vs. 85:95	**.35**	.52	.13
	9:10 vs. 90:100	.17	.26	**.57**
	9:10 vs. 95:105	.35	**.39**	.26
Middle adolescents (*N* = 35)	9:10 vs. 85:95	**.63**	.20	.17
	9:10 vs. 90:100	.29	.14	**.57**
	9:10 vs. 95:105	.43	**.34**	.23
Adults (*N* = 32)	9:10 vs. 85:95	**.88**	.09	.03
	9:10 vs. 90:100	.06	.09	**.85**
	9:10 vs. 95:105	.25	**.69**	.06

Table 27b

Proportion of Preferences (Correct Responses in Bold) in The Volleyball Problem

Age group	Trial	Preference		
		Team A (less numerous)	Team B (more numerous)	No difference
Seventh graders (*N* = 40)	9:10 vs. 85:95	**.48**	.35	.17
	9:10 vs. 90:100	.32	.37	**.30**
	9:10 vs. 95:105	.50	**.35**	.15
Middle adolescents (*N* = 27)	9:10 vs. 85:95	**.48**	.37	.15
	9:10 vs. 90:100	.19	.18	**.63**
	9:10 vs. 95:105	.33	**.52**	.15
Adults (*N* = 29)	9:10 vs. 85:95	**.72**	.10	.17
	9:10 vs. 90:100	.10	.14	**.75**
	9:10 vs. 95:105	.14	**.76**	.10

Table 27c

Proportion of Preferences (Correct Responses in Bold) in The Vaccination Problem

Age group	Trial	Preference		
		Clinic A (less numerous)	Clinic B (more numerous)	No difference
Seventh graders (*N* = 29)	9:10 vs. 85:95	**.53**	.31	.14
	9:10 vs. 90:100	.41	.24	**.35**
	9:10 vs. 95:105	.52	**.28**	.20
Middle adolescents (*N* = 36)	9:10 vs. 85:95	**.61**	.22	.17
	9:10 vs. 90:100	.22	.19	**.58**
	9:10 vs. 95:105	.36	**.36**	.28
Adults (*N* = 31)	9:10 vs. 85:95	**.74**	.16	.10
	9:10 vs. 90:100	.03	.39	**.58**
	9:10 vs. 95:105	.10	**.80**	.10

In the mathematical problem, seventh graders responded correctly to 44%, middle adolescents to 51%, and adults to 81% of all trials. In the volleyball problem, seventh graders responded correctly to 38%, middle adolescents to 54%, and adults to 74% of all trials. In the vaccination problem, seventh graders responded correctly to 39%, middle adolescents to 52%, and adults to 71% of all trials.

A Generalized Linear Mixed Model (GLMM) was computed to investigate how age, trial, and problem (mathematical, volleyball or vaccination) influence the proportion of correct responses. As in Experiments 1 and 2, the data were analyzed using a Generalized Linear Mixed Model for binomially distributed outcomes (correct response or wrong response). A Generalized Linear Mixed Model was run with age, trial, and problem as fixed effects. Participants were included in the model as a random factor to allow us to take into account the dependence between our observations due to repeated measures. The random intercept for participants allowed the intercept to vary between participants.

The Akaike information criterion AIC was employed as the model-selection method. A baseline model was constructed (with all the possible interactions and main effects), and the best-fitting model was defined as the one minimizing the AIC. The best-fit model is summarized in Table 28, including the regression coefficients, their standard errors, and the corresponding Z scores. The right side of Table 28 shows the odds ratios, their confidence intervals and the relative significance that the main effects and the interaction have in the model (Chi squared). We compared the three age levels by fixing the seventh graders as the reference and comparing middle adolescents and adults with the seventh graders. Moreover, we also compare the three trial levels by fixing the trial 9-in-10 vs. 90-in-100 as the reference, and the problem by fixing the mathematical problem as the reference.

The best-fitting model has only main effects and the only significant effects were for age. The probability that middle adolescents give a correct answer is 1.68 times higher than seventh graders ($B = .52$, $Z = 3.00$, $p < .01$), and adults are 5 times more likely to answer correctly than seventh graders ($B = 1.55$, $Z = 4.71$, $p < .001$).

Table 28

Generalized Linear Mixed Model Parameters and Effect Sizes Predicting Correct Answers in Experiment 3

Best-fit model	*B*	(*SE*)	*Z*	OR	95% CI	Chisq (*df*)
Age						74.08 (2)***
Middle adolescents	0.52	(.17)	3.00**	1.68	[1.21, 2.35]	
Adults	1.55	(.19)	8.24***	4.71	[3.25, 6.84]	
Trial						8.81 (2)*
9:10 vs. 85:95	0.19	(.18)	1.07	1.26	[.65, 2.45]	
9:10 vs. 95:105	-0.33	(.18)	-1.86	.72	[.51, 1.02]	
Problem						1.11 (2)
Volleyball problem	-0.13	(.18)	-0.71	.88	[.62, 1.25]	
Vaccination problem	-0.19	(.48)	-1.03	.83	[.58, 1.18]	

Note. *** p < 0.001, **p < .01*;* Baseline categories: seventh graders for the Age; 9-in-10 vs. 90-in-100 for Trial; and volleyball for Problem.

5.3.2 Analysis of biased answers

The analysis of correct answers shows that accuracy increases with age and neither problems nor trials systematically influence accuracy. Analysis of correct answers does not shed light on patterns of biased responses.

To analyze biases we analyzed the actual responses (not their accuracy) to each trial using multinomial logistic regression. The comparison category for preference was 9-in-10, and preferences for both the more numerous response and for *it is the same* were compared with this category. The seventh graders were the comparison category for the three age levels, the volleyball problem was the comparison category for the three problems, and trial 1 was the comparison category for the three trials.

The Akaike information criterion AIC was employed as the model-selection method. The best-fit model took into account the main effects and the interactions between age and scenario and age and trial. The best-fit model is summarized in the left side of Table 29, including the regression coefficients, their standard errors, and the corresponding *Z* scores. The right side of Table 29 shows the odds ratios, which measure effect size, and their 95% confidence interval (CI). The odds ratios are the antilog (i.e., exponentiated values) of the model coefficients.

Table 29

Multinomial Logistic Regression Model Parameters and Effect Sizes (N = 282)

Best-fit model	*B* (SE)	*Z*	OR	95% CI
Preference for 9-in-10 vs. preference for the more numerous response				
Age (Adolescents)	-.18 (.42)	-.42	.84	[.37, 1.90]
Age (Adults)	-1.77 (.52)	-3.39**	.17	[.06, .47]
Trial (9-in-10 vs. 90-in-100)	.19 (.35)	.53	1.21	[.61, 2.40]
Trial (9-in-10 vs. 95-in-105)	-.12 (.33)	-.37	.89	[.46, 1.69]
Scenario (Mathematical)	.49 (.36)	1.38	1.63	[.81, 3.31]
Scenario (Vaccination)	-.39 (.33)	-1.20	.68	[.35, 1.29]
Age (Adolescents) X Trial (9-in-10 vs. 90-in-100)	.34 (.53)	. 63	1.40	[.50, 3.97]
Age (Adults) X Trial (9-in-10 vs. 90-in-100)	2.91 (.68)	4.27***	18.36	[4.84, 69.60]
Age (Adolescents) X Trial (9-in-10 vs. 95-in-105)	1.01 (.47)	2.15*	2.75	[1.09, 6.90]
Age (Adults) X Trial (9-in-10 vs. 95-in-105)	3.68 (.56)	6.56***	39.65	[13.23, 118.82]
Age (Adolescents) X Scenario (Mathematical)	1.26 (.52)	-2.44**	.28	[.10, .79]
Age (Adults) X Scenario (Mathematical)	-1.09 (.59)	-1.84	.34	[.11, 1.07]
Age (Adolescents) X Scenario (Vaccination)	-.12 (.49)	-.24	.89	[.34, 2.32]
Age (Adults) X Scenario (Vaccination)	1.01 (.58)	1.75	2.75	[.88, 8.56]
Preference for 9-in-10 vs. preference for the response it is the same				
Age (Adolescents)	.18 (.53)	.34	1.20	[.42, 3.38]
Age (Adults)	-.45 (.58)	-.78	.64	[.20, 1.99]
Trial (9-in-10 vs. 90-in-100)	1.33 (.40)	3.32***	3.78	[1.73, 8.28]
Trial (9-in-10 vs. 95-in-105)	.25 (.42)	.60	1.28	[.56, 2.92]
Scenario (Mathematical)	.87 (.40)	2.15*	2.39	[1.09, 5.23]
Scenario (Vaccination)	-.03 (.37)	-.08	.97	[.47, 2.00]
Age (Adolescents) X Trial (9-in-10 vs. 90-in-100)	.86 (.55)	1.57	2.36	[.80, 6.94]
Age (Adults) X Trial (9-in-10 vs. 90-in-100)	3.21 (.69)	4.65***	24.78	[6.41, 95.81]
Age (Adolescents) X Trial (9-in-10 vs. 95-in-105)	.50 (.57)	.87	1.65	[.54, 5.04]
Age (Adults) X Trial (9-in-10 vs. 95-in-105)	1.24 (.71)	1.75	3.46	[.86, 13.90]
Age (Adolescents) X Scenario (Mathematical)	-1.12 (.56)	-2.00*	.33	[.11, .98]
Age (Adults) X Scenario (Mathematical)	-1.48 (.65)	-2.27*	.23	[.06, .81]
Age (Adolescents) X Scenario (Vaccination)	-.02 (.54)	.05	.98	[.34, 2.82]
Age (Adults) X Scenario (Vaccination)	-.27 (.65)	-.42	.76	[.21, 2.73]

Note. OR = odds ratio; CI = confidence interval; *p < .05, **p <.01, ***p <.001

The top half of Table 29 compares preference for 9-in-10 to preference for the more numerous response (85-in-95, or 90-in-100, or 95-in-105). There are four significant two-way interactions. Across all problems middle adolescents were more likely than seventh graders to choose correctly the more numerous alternative (95-in-105) in trial 3 and less likely to choose it incorrectly in trial 1 ($B = 1.01$, $Z = 2.15$, $p < .01$). Adults were also more likely than seventh graders to respond correctly (95-in-105) in trial 3 and less likely to respond incorrectly in trial 1 ($B = 3.68$, $Z = 6.56$, $p < .001$). When the ratios were equal (in trial 2) adults rarely selected the less numerous or more numerous responses but more than half of seventh graders' responses were in these two categories, yielding a significant interaction ($B = 2.91$, $Z = 4.27$, $p < .001$).

In the mathematical problem seventh graders exhibited a bias for the more numerous response, selecting Container B in 39% of trials and Container A in 29%, whereas middle adolescents exhibited a bias for the less numerous response, selecting Container B in 23% of trials and Container A in 45%. In the reference volleyball problem the bias for seventh graders was reversed and adolescents showed little evidence of bias, resulting in a significant Age (Adolescents) X Scenario interaction ($B = 1.26$, $Z = -2.44$, $p < .01$). Seventh graders selected Container B in 36% of trials and Container A in 43%, whereas middle adolescents selected Container B in 36% of trials and Container A in 33%. Figure 13 shows the pattern of responses in these two problems.

In Figure 13, black lines represent the probability to give the specific response (9-in-10, the more numerous response, or it is the same) according to age group and scenario. Dashed red lines represent the confidence intervals of the odds ratios. These differences in biases between the two problems are also evident when considering the response that the two ratios are the same, as confirmed by the significant two way Age X Scenario interactions reported in the bottom half of Table 29 ($B = -1.12$, $Z = -2.00$, $p < .05$, $B = -1.48$, $Z = -2.27$, $p < .05$ with regard to middle adolescents and adults, respectively).

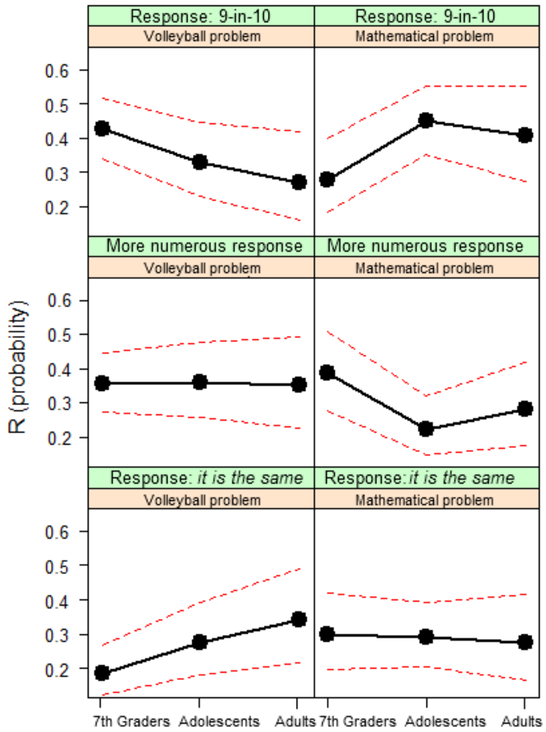

Figure 13. Graphical representation of the response choice made in each trial of both problems by all three age groups

5.3.3 Analysis of written preference explanations

As in Experiment 1 and in Experiment 2 we considered the explanations that participants gave to their responses after each trial to understand how participants reasoned. We coded their explanations in four categories:

1. Mathematical reasoning: participants who reasoned mathematically and explained their response through considerations of proportions and probabilities. For example “probabilities are the same/differ” or “because the answer I gave was correct by a mathematical point of view,” or “because of percentages”;
2. Denominator neglect: participants who stated that they gave their response according to the absolute number of marbles, applicants, or vaccinations. For example “there was about an equal chance, but since there were 85 opposed

to 9 it seemed more likely that one might pick a winning marble", or "the sheer larger number of accepted individuals made the more numerous team seem to have a higher chance at not being rejected" or "in the large clinic there are more effective vaccinations ";

3. Reasoning based on the absolute number of negative events: participants who stated that they gave their response according to smaller number of applicants/vaccinations. For example "the small container offers more chance to pick a winning marble", or "in the small clinic the probabilities that the vaccination is not effective are less than in the big clinic";
4. Intuition: participants who explained their response referring to intuition, gut feelings, sensations or they were not able to provide a formal explanation. For example, "I had an instinct that was the correct answer" or "decided by gut feeling", or "I just think that I selected the correct answer".

Table 30a, 30b, and 30c shows the proportions of correct and wrong responses associated with each category of explanation in the mathematical problem, the volleyball problem and the vaccination problem, respectively.

In the mathematical problem the proportion of seventh graders, middle adolescents and adults that justified their choice in accordance to mathematical reasoning is similar to Experiment 1 (Table 10). In trial 9-in-10 versus 90-in-100 seventh graders explained their response (correct responses) with intuition (26%) more than in Experiment 1 (6%). The explanations indicate that they probably applied an automatized mathematical reasoning that depends on their habit to simplify fractions. Interestingly, in trial 9-in-10 versus 85-in-95, seventh graders' tendency (26% in Experiment 1) to prefer 9-in-10 because of less number of losing marbles (reversed reasoning) disappears with time-pressure (4%). Thirty-five percentage of seventh graders preferred the large container because of denominator neglect, writing that "more winning marbles more winning chances". The pattern of responses in trial 9-in-10 versus 95-in-105 is similar to that found Experiment 1 in terms of the application of denominator neglect and reversed reasoning.

Table 30a

Mathematical Problem: Proportions of Correct and Wrong Responses Associated With Explanations Given by Participants of Each Age Group in Three Trials

9-in-10 versus 90-in-100	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 23)	.30	.09	—	.13	—	.04	.26	.18
Middle adolescents (*N* = 35)	.49	.06	—	.06	—	.09	.09	.23
Adults (*N* = 32)	.78	.03	—	.03	—	.03	.06	.06
9-in-10 versus 85-in-95	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 23)	.13	.09	—	.35	.04	—	.17	.22
Middle adolescents (*N* = 35)	.37	.14	—	.06	.11	—	.14	.17
Adults (*N* = 32)	.53	.06	—	.06	.03	—	.31	.03
9-in-10 versus 95-in-105	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 23)	.08	.18	.08	—	—	.13	.22	.30
Middle adolescents (*N* = 35)	.09	.17	.06	—	.03	.14	.17	.34
Adults (*N* = 32)	.53	.09	.03	—	—	.06	.12	.16

In the volleyball problem, compared to Experiment 2, the proportion of correct responses associated to mathematical reasoning decreases in all three trials and in all age groups. To note that in Experiment 2 most of middle adolescents (40%) in trial 9-in-10 vs. 95-in-105 justified the wrong answer as based on math and calculation. At the same time, only 8% of adolescents who relied on intuition gave the correct answer.

In the present experiment the number of responses justified with intuition increases compared to those in Experiment 2; in trial 9-in-10 vs. 95-in-105 the proportion of correct answers associated to intuitive explanations such as “I feel that I gave the correct answer” is high in middle adolescents and adults (37% and 31%, respectively). Contrary to Experiment 2, just few adolescents (7%) justified the wrong answer as based on mathematical reasoning and a large proportion of adolescents (31%)

who responded correctly, relied on intuition (31%). Proportion of answers associated to denominator neglect and reversed reasoning are similar to Experiment 2.

Table 30b

Volleyball Problem: Proportions of Correct and Wrong Responses Associated With Explanations Given by Participants of Each Age Group in Three Trials

9-in-10 versus 90-in-100	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (N = 40)	.22	.05	—	.18	—	.23	.07	.25
Middle adolescents (N = 27)	.48	.03	—	—	—	.07	.15	.26
Adults (N = 29)	.72	.03	—	.14	—	.07	.03	—

9-in-10 versus 85-in-95	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (N = 40)	.12	.20	—	.17	.25	—	.10	.15
Middle adolescents (N = 27)	.18	.15	—	.11	.11	—	.19	.26
Adults (N = 29)	.41	.17	—	.03	.07	—	.24	.07

9-in-10 versus 95-in-105	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (N = 40)	.05	.17	.20	.03	—	.25	.10	.20
Middle adolescents (N = 27)	.07	.04	.03	—	.03	.22	.37	.22
Adults (N = 29)	.38	.07	.07	—	—	.07	.31	.10

In the vaccination problem participants gave a low proportion of correct answers associated to mathematical reasoning. This result demonstrates the correct responses cannot always be attributed to analytic reasoning. According to FTT, slow and verbatim process is not the *sine qua non* of mathematical ability and correct response might depend also on gist processing or secondary intuition (Fischbein, 1978). With regard to the bias, almost 20% of seventh graders explained their choice with reversed reasoning, such as “less numerous clinic is equal to less probability that the vaccination is *not* effective”, in three trials. Just few middle adolescents gave explanations according to denominator neglect. Finally, it is interesting to note that, in trial 9-in-10 versus 90-in-100, the proportion of adults that applied denominator neglect (32%) is higher than in

Experiment 2 (15%). They wrote, for example, that "When more people go to a clinic, it just feels like it is a better choice even if the probability of success is lower", or "I would rather go to a clinic that vaccinates 90 people per week as opposed to 9. The more people they see, the more experience they have".

Table 30c

Vaccination Problem: Proportions of Correct and Wrong Responses Associated With Explanations Given by Participants of Each Age Group in Three Trials

9-in-10 versus 90-in-100	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 29)	.21	.21	—	.07	.07	.21	.07	.17
Middle adolescents (*N* = 36)	.50	.06	—	.11	—	.11	.03	.14
Adults (*N* = 31)	.58	.03	—	.32	—	.03	—	.03
9-in-10 versus 85-in-95	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 29)	.21	.10	.03	.14	.24	—	.07	.21
Middle adolescents (*N* = 36)	.19	.11	.03	.11	.11	—	.28	.17
Adults (*N* = 31)	.55	.06	—	.13	.10	—	.10	.06
9-in-10 versus 95-in-105	Mathematical reasoning		Denominator neglect		Reversed reasoning		Intuition	
	Correct	Wrong	Correct	Wrong	Correct	Wrong	Correct	Wrong
Seventh graders (*N* = 29)	.10	.10	.03	.03	—	.28	.14	.31
Middle adolescents (*N* = 36)	.08	.22	.11	—	—	.08	.17	.33
Adults (*N* = 31)	.61	.06	—	—	—	—	.19	.13

5.3.4 Predictors of reasoning performance

Predictors of reasoning performance in seventh graders

After participants performed the assigned problem, their mathematical ability was assessed using the AC-MT 11-14 (Cornoldi & Cazzola, 2003). We investigated their written computation ability, their comprehension and production ability, and their arithmetical reasoning ability. Table 31 presents seventh graders' performance at AC-MT 11-14.

Table 31

Descriptive Statistics for Seventh Graders AC-MT 11-14 (N = 92)

	Min	Max	*M*	*SD*	Normative ratings
Written calculation	0	10	5.48	2.57	6.34
Comprehension and production	0	20	12.64	4.86	4.90
Arithmetical reasoning	2	29	15.57	5.93	5.77
AC-MT 11-14 total rating	4	73	44.64	13.83	14.83

According to the performance standards established by Cornoldi and Cazzola (2003), 17.4% of participants need immediate intervention, 28.3% of participants need attention, 50% of participants performed sufficiently, and only 4.3% had excellent mathematical ability. Figure 14 shows the density distribution of the total number of correct responses (range = 0-3) for each of these four levels of performance.

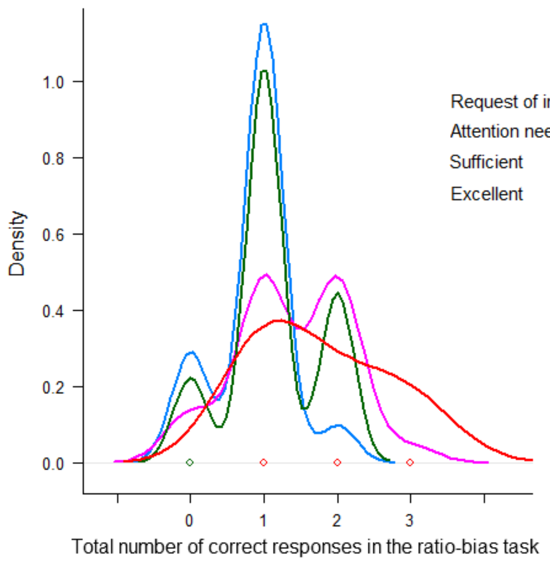

Figure 14. Density plots for correct responses grouped by seventh grade mathematical level according to the AC-MT 11-14

Overall, the four distributions overlap and those participants with an excellent mathematical ability did not respond correctly more often than other participants.

Table 32a

Seventh Grade Tetrachoric, Polychoric and Pearson Correlations Between Correct Responses in the Mathematical Problem and AC-MT 11-14 (N = 23)

Variables	1	2	3	4	5	6	7	8
Ratio-bias task								
1. 9-in-10 vs. 85-in-95	—							
2. 9-in-10 vs. 90-in-100	-.05	—						
3. 9-in-10 vs. 95-in-105	-.06	.08	—					
4. Total correct responses	.57**	.57**	.64**	—				
AC-MT 11-14								
5. Written calculation	.10	.22	.40	.36	—			
6. Comprehension and production	.28	.42*	.26	.51*	.73**	—		
7. Arithmetical reasoning	-.36	-.11	.26	-.12	.23	.24	—	
8. AC-MT 11-14 total score	.01	.22	.40	.32	.87**	.84**	.61**	—

Table 32b

Seventh Grade Tetrachoric, Polychoric and Pearson Correlations Between Correct Responses in the Volleyball Problem and AC-MT 11-14 (N = 40)

Variables	1	2	3	4	5	6	7	8
Ratio-bias task								
1. 9-in-10 vs. 85-in-95	—							
2. 9-in-10 vs. 90-in-100	.02	—						
3. 9-in-10 vs. 95-in-105	-.44**	-.50**	—					
4. Total correct responses	.51**	.43**	.13	—				
AC-MT 11-14								
5. Written calculation	.29	.49**	-.43**	.23	—			
6. Comprehension and production	.14	.31*	-.26	.12	.42**	—		
7. Arithmetical reasoning	-.22	.36*	.03	.10	.41**	.52**	—	
8. AC-MT 11-14 total score	.09	.50**	-.36*	.12	.79**	.68**	.82**	—

Table 32c

Seventh Grade Tetrachoric, Polychoric and Pearson Correlations Between Correct Responses in the Vaccinationl Problem and AC-MT 11-14 (N = 29)

Variables	1	2	3	4	5	6	7	8
Ratio-bias task								
1. 9-in-10 vs. 85-in-95	—							
2. 9-in-10 vs. 90-in-100	-.03	—						
3. 9-in-10 vs. 95-in-105	-.49**	.15	—					
4. Total correct responses	.41*	.69**	.40*	—				
AC-MT 11-14								
5. Written calculation	-.13	.49**	.01	.20	—			
6. Comprehension and production	.03	.43*	-.09	.21	.65**	—		
7. Arithmetical reasoning	.09	.61**	-.16	.31	.51**	.51**	—	
8. AC-MT 11-14 total score	-.02	.57**	-.15	.23	.88**	.80**	.81**	—

*p < .05; **p < .01

Table 32a, 32b and 32c show the correlations between responses (correct or wrong) for each trial at the ratio-bias task, the total of correct responses given by participants at the ratio-bias task (range: 0-3), and the AC-MT (subscales and total score), divided for scenario.

There were no significant correlations between ratio-bias accuracy and mathematical abilities, suggesting that either the mathematical ability measured with the AC-MT 11-14 does not measure the ability required to respond correctly in the ratio-bias task or that mathematical ability is not an important contributor to performance in the ratio-bias task.

Mathematical ability is most likely to contribute to performance in the ratio-bias task when participants attempt to determine the answer by comparing the ratios mathematically.

Predictors of reasoning performance in middle adolescents

Middle adolescents were tested in the PMA, objective numeracy, subjective numeracy, cognitive reflection ability, need for cognition and faith in intuition. Table 33 summaries descriptive statistics for each scale.

Table 33

Means, Standard Deviations and Reliabilities of the PMA, Objective Numeracy, Subjective Numeracy, Need for Cognition, Faith in Intuition, and CRT (N = 98)

	Min	Max	*M*	*SD*	Cronbach's α
PMA					
Spatial ability	1	51	21.54	11.48	—
Verbal ability	10	40	25.49	6.47	—
Calculation ability	0	46	22.44	10.35	—
Objective numeracy total score	1	11	6.09	2.33	.63
General numeracy	0	3	1.26	.93	.28
Expanded numeracy	0	8	4.84	1.79	.57
Subjective numeracy general mean	1.38	5.00	3.04	.68	.72
Cognitive ability	1.50	5.00	3.18	.87	.84
Preference for information display	1.00	5.00	2.90	.78	.44
Cognitive reflection task total score	0	3	0.38	0.73	—
Need for cognition general mean	1.90	4.35	3.11	0.53	.81
Rational ability	1.70	4.40	3.03	0.60	.74
Rational engagement	1.80	4.30	3.05	0.55	.62
Faith in intuition general mean	1.95	3.95	3.06	0.43	.70
Experiential ability	2.10	4.20	3.07	0.44	.46
Experiential engagement	1.40	4.60	3.06	0.56	.65

Tables 34a, 34b, and 34c show the correlations between the responses (correct or wrong) for each trial and each scale with the relative subscales, divided for scenario (mathematical, volleyball, and vaccination scenario, respectively).

First, consider the correlations between accuracy in the mathematical problem and these ability measures. Total accuracy in the ratio-bias task was positively correlated with PMA Verbal Ability, all three subscales of Subjective Numeracy, the Cognitive Reflection task, and all three subscales of Need for Cognition. Interestingly, most of these correlations appear to be due to correlations with performance on trial 2.

Second, consider the correlations between accuracy in the volleyball problem and these ability measures. Total accuracy in the ratio-bias task was positively correlated with five of the same nine measures as in the mathematics problem. It was correlated with subscales of Subjective Numeracy, the Cognitive Reflection task, subscales of Need for Cognition, and PMA Calculation Ability.

Finally, consider the correlations between accuracy in the vaccination problem and ability measures. Total accuracy in the ratio-bias task was positively correlated only with the Cognitive Reflection task. Although the total score was not correlated with many of the measures, accuracy in trial 2 was positively correlated with eight measures and accuracy in trial 3 was negatively correlated with seven measures. Adolescents with greater numeric and reasoning abilities were more likely to respond correctly when ratios were equal, but less likely to respond correctly in trial 3. This is consistent with the context-dependent bias toward a smaller clinic that was observed in adolescents.

Table 34a

Middle Adolescent Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Mathematical Problem (N = 35)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ratio-bias task																				
1. 9-in-10 vs. 85-in-95	—	-.03	-.05	.50**	-.05	.29*	-.28	.15	.07	.16	.07	.20	-.08	.24	.22	.20	.21	.02	-.15	.17
2. 9-in-10 vs. 90-in-100		—	.49**	.71**	.38*	.24	.33*	.33*	.28	.28	.52**	.42**	.34*	.48**	.48**	.42**	.46**	.05	.12	-.02
3. 9-in-10 vs. 95-in-105			—	.73**	.12	.23	.35*	.24	.37*	.12	.27	.11	.36*	.20	.31	.36*	.22	.05	.18	-.09
4. Total correct responses (range: 0-3)				—	.22	.35*	.18	.33*	.32	.26	.38*	.36*	.27*	.39*	.45**	.43**	.41*	.04	.01	.05
PMA																				
5. Spatial ability					—	.24	.15	.11	.20	.04	.27	.28	.18	.10	.03	.07	-.07	.02	.24	-.21
6. Verbal ability						—	.12	.19	.42**	.04	.12	.26	-.07	.19	.10	.11	.06	-.01	.04	-.05
7. Calculation ability							—	.16	.31	.05	.28	.26	.22	.23	.03	.07	-.01	.01	.15	-.12
Objective Numeracy Scale																				
8. Objective numeracy general score								—	.66**	.93**	.21	.32	.04	.21	.10	.09	.09	.39*	.39*	.27
9. General numeracy									—	.35*	.19	.16	.17	.27	.15	.12	.15	.20	.35*	.09
10. Expanded numeracy										—	.17	.33*	-.03	.13	.06	.06	.04	.39*	.33*	.32
Subjective Numeracy Scale																				
11. Subjective numeracy general score											—	.86**	.86**	.42**	.38**	.48**	.21	.20	.40**	-.04
12. Cognitive ability												—	.48**	.48**	.30	.35*	.19	.18	.33*	-.01
13. Information display													—	.25	.36*	.47**	.18	.16	.36*	-.07
Cognitive Reflection Task																				
14. Cognitive reflection general score														—	.36*	.36*	.30	.28	.46**	.05
Need for Cognition Scale																				
15. General need for cognition score															—	.93**	.92**	.15	.23	.03
16. Rational ability																—	.70**	.12	.24	-.02
17. Rational engagement																	—	.15	.18	.08
Faith in Intuition Scale																				
18. Faith in intuition general score																		—	.82**	.87**
19. Experiential ability																			—	.41*
20. Experiential engagement																				—

Note. * $p < .05$; ** $p < .01$

Table 34b

Middle Adolescent Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Volleyball Problem (N = 27)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ratio-bias task																				
1. 9-in-10 vs. 85-in-95	—	-.28	-.10	.64**	-.03	-.01	.25	-.10	.07	.02	.16	.21	.09	.25	.10	.15	.02	.05	.01	.07
2. 9-in-10 vs. 90-in-100		—	.09	.65**	.28	.14	.29	.35	.10	.43*	.53**	.27	.61**	.28	.35	.19	.49**	-.01	-.06	.02
3. 9-in-10 vs. 95-in-105			—	.47*	-.26	.-22	.33	-.03	.26	-.23	.33	.25	.27	.44*	.43*	.38*	.42*	-.35	-.42*	-.24
4. Total correct responses (range: 0-3)				—	-.17	-.03	.43*	.08	.16	.01	.51**	.27	.57**	.43*	.42*	.32	.47**	-.18	-.25	-.09
PMA																				
5. Spatial ability					—	.24	.23	.04	-.06	.11	.02	.08	-.07	.20	.01	-.03	.06	.12	.23	.02
6. Verbal ability						—	-.13	.29	.20	.26	-.12	-09	-.10	-.10	.09	-.17	.03	.15	.05	.19
7. Calculation ability							—	-.17	-.01	.24	.30	.27	.20	.54**	.37	.29	.41*	.10	-.07	.20
Objective Numeracy Scale																				
8. Objective numeracy general score								—	.77**	.88**	.43*	.29	.39*	.12	.23	.14	.31	-.04	-.26	.11
9. General numeracy									—	.37	.35	.14	.44*	.24	.13	.05	.21	-.11	-.40*	.11
10. Expanded numeracy										—	.37	.32	.25	.01	.24	.16	.29	.02	-.08	.08
Subjective Numeracy Scale																				
11. Subjective numeracy general score											—	.84**	.72**	.49**	.60**	.58**	.52*	-.06	-.13	-.01
12. Cognitive ability												—	.22	.38*	.51**	.53**	.40*	-.21	-.18	-.20
13. Information display													—	.40*	.42*	.36	.43*	.16	.01	.24
Cognitive Reflection Task																				
14. Cognitive reflection general score														—	.18	.17	.16	.05	.02	.06
Need for Cognition Scale																				
15. General need for cognition score															—	.95**	.91**	-.09	-.14	-.03
16. Rational ability																—	.73**	.09	-.06	-.09
17. Rational engagement																	—	-.07	-.21	.05
Faith in Intuition Scale																				
18. Faith in intuition general score																		—	.86**	.93**
19. Experiential ability																			—	.60**
20. Experiential engagement																				—

Note. * $p < .05$; ** $p < .01$

Table 34c

Middle Adolescent Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Vaccination Problem (N = 36)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ratio-bias task																				
1. 9-in-10 vs. 85-in-95	—	.27	-.39*	.50**	.12	-.26	.42**	.17	-.11	.27	.32	.27	.27	.22	.43**	.52**	.28	.09	-.13	.22
2. 9-in-10 vs. 90-in-100		—	-.01	.73**	.55**	-.09	.07	.74**	.57**	.69**	.28	.17	.30	.67**	.52**	.47**	.48**	.02	-.03	.04
3. 9-in-10 vs. 95-in-105			—	.40**	-.21	-.26	.24	-.49**	-.28	-.49**	-.27	-.04	-.43**	.03	-.43**	-.46**	-.33*	-.30	.13	-.53**
4. Total correct responses (range: 0-3)				—	.22	-.31	.15	.22	.08	.26	.18	.20	.09	.41**	.27	.29	.21	-.09	-.02	-.12
PMA																				
5. Spatial ability					—	.03	-.17	.53**	.26	.56**	.01	-.11	.12	.32	.31	.35*	.22	-.12	-.19	-.02
6. Verbal ability						—	.18	.16	.04	.19	-.12	-.17	-.02	-.05	-.18	-.26	-.08	.06	-.07	.14
7. Calculation ability							—	.19	.10	.19	.16	.11	.16	.20	-.01	-.06	.08	.10	.08	.09
Objective Numeracy Scale																				
8. Objective numeracy general score								—	.73**	.95**	.29	.11	.38*	.44**	.44**	.43**	.38*	.12	-.06	.21
9. General numeracy									—	.48**	.24	.17	.24	.35*	.25	.28	.18	.10	.10	.07
10. Expanded numeracy										—	.26	.07	.38*	.41*	.45**	.42*	.41*	.10	-.12	.24
Subjective Numeracy Scale																				
11. Subjective numeracy general score											—	.84**	.80**	.15	.30	.22	.34*	.17	.12	.15
12. Cognitive ability												—	.34*	.17	.21	.18	.21	.12	.18	.04
13. Information display													—	.07	.30	.18	.36*	.15	.01	.22
Cognitive Reflection Task																				
14. Cognitive reflection general score														—	.16	.11	.19	.07	.11	.02
Need for Cognition Scale																				
15. General need for cognition score															—	.92**	.93**	.05	-.12	.16
16. Rational ability																—	.71**	.02	-.18	.16
17. Rational engagement																	—	.07	-.04	.14
Faith in Intuition Scale																				
18. Faith in intuition general score																		—	.76**	.87**
19. Experiential ability																			—	.34*
20. Experiential engagement																				—

Note. * $p < .05$; ** $p < .01$

Predictors of reasoning performance in adults

Adults were tested in the RAPM, objective numeracy, subjective numeracy, cognitive reflection ability, need for cognition and faith in intuition. Table 35 summaries the descriptive statistics for each scale.

Table 35

Means, Standard Deviations and Reliabilities of the RAPM, Objective numeracy, Subjective numeracy, CRT, Go/no-go task, Need for Cognition, and Faith in Intuition (N = 92)

	Min	Max	*M*	*SD*	Cronbach's *α*
RAPM					
Total score	1	12	9.02	2.32	—
Objective numeracy total score	2	11	9.88	1.56	.68
General numeracy	0	3	2.45	.75	.42
Expanded numeracy	2	8	7.43	1.02	.57
Subjective numeracy general mean	1.63	5.00	3.67	.63	.76
Cognitive ability	1.00	5.00	3.62	.79	.80
Preference for information display	2.00	5.00	3.71	.69	.55
Cognitive reflection task total score	0	3	1.31	1.12	—
Go/no-go task					
Reaction time (ms)	228.03	498.70	317.97	51.81	—
Correct answers	56	273	238.94	43.82	—
False alarms	0	58	18.78	9.72	—
Need for cognition general mean	2.05	5.00	3.61	0.50	.88
Rational ability	2.00	5.00	3.57	0.60	.85
Rational engagement	2.10	5.00	3.66	0.52	.78
Faith in intuition general mean	1.90	4.70	3.36	0.52	.88
Experiential ability	1.90	4.70	3.36	0.55	.79
Experiential engagement	1.80	4.50	3.34	0.59	.84

Tables 36a, 36b, and 36c show the correlations between the responses at the ratio-bias task and each scale with the relative subscales.

First, consider the correlations between accuracy in the mathematical problem and these ability measures. Total accuracy in the ratio-bias task was correlated with 13 measures including the Raven APM, subscales of Objective Numeracy, one subscale of Subjective Numeracy, the Cognitive Reflection task, sub-measures of the Go/no-go task, and negatively with Faith in Intuition subscales.

Table 36a

Adult Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Mathematical Problem (N = 32)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Ratio-bias task																					
1. 9-in-10 vs. 85-in-95	—	.10	.15	.51**	.28	.17	.20	.12	.11	.22	-.04	.26	-.59**	.47**	.26	.02	.01	.03	-.02	-.02	.03
2. 9-in-10 vs. 90-in-100		—	.64**	.80**	.46**	.68**	.39*	.78**	.24	.35*	.02	.41*	-.22	.46**	.09	.07	.08	.06	-.40*	-.40*	-.37*
3. 9-in-10 vs. 95-in-105			—	.87**	.38*	.53**	.25	.65**	.26	.35*	.02	.59**	-.13	.33	.05	.21	.19	.20	-.37*	-.37*	-.34
4. Total correct responses				—	.50**	.64**	.38*	.72**	.28	.42*	.04	.59**	-.39*	.55*	.16	.15	.14	.14	-.37*	-.37*	-.33
RAPM																					
5. RAPM total score					—	.45**	.42*	.41*	.01	-.01	.01	.27	.27	-.09	.02	.18	.09	.25	-.12	-.12	-.17
Objective Numeracy Scale																					
6. Objective numeracy general score						—	.88**	.95**	.34	.35*	.21	.54**	-.04	.22	.03	.20	.19	.17	-.46**	-.46**	-.33
7. General numeracy							—	.69**	.32	.32	.21	.43*	-.07	.07	-.02	.30	.28	-.26	-.29	-.29	-.16
8. Expanded numeracy								—	.31	.33	.18	.55**	-.02	.30	.05	.11	.11	-.09	-.51**	-.51**	-.39*
Subjective Numeracy Scale																					
9. Subjective numeracy general score									—	.85**	.81**	.59**	-.36*	.30	.15	.47**	.54**	.29	.19	.19	.01
10. Cognitive ability										—	.38*	.59**	-.35*	.30	.16	.34	.45**	.15	.09	.09	-.08
11. Information display											—	.38*	-.23	.19	.09	.45**	.45**	.35*	.24	.24	.08
Cognitive Reflection Task																					
12. Cognitive reflection general score												—	-.25	.25	.08	.34	.28	.36*	-.09	-.09	.02
Go/no-go task																					
13. Reaction Time													—	-.64**	-.45**	-.14	-.18	-.07	-.08	-.08	.02
14. Proportion of correct responses														—	.01	.20	.26	.09	-.02	-.02	-.09
15. Proportion of false alarm															—	-.03	-.04	-.01	-.24	-.24	-.12
Need for Cognition Scale																					
16. General need for cognition score																—	.92**	.90**	.22	.22	.21
17. Rational ability																	—	.67**	.12	.11	.06
18. Rational engagement																		—	.29	.28	.35
Faith in Intuition Scale																					
19. Faith in intuition general score																			—	.99**	.67**
20. Experiential ability																				—	.67**
21. Experiential engagement																					—

Note. $*p < .05$; $**p < .01$

Table 36b

Adult Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Volleyball Problem (N = 29)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Ratio-bias task																					
1. 9-in-10 vs. 85-in-95	—	.09	.59**	.73**	-.42*	-.23	.02	-.43*	-.25	-.13	-.29	-.19	-.12	.09	.45**	-.32	-.19	-.37*	.22	.21	.32
2. 9-in-10 vs. 90-in-100		—	.11	.59**	.06	.20	.24	.07	-.22	.06	-.45**	.04	-.06	-.31	.03	-.29	-.39*	-.06	-.15	-.13	.33
3. 9-in-10 vs. 95-in-105			—	.78**	.07	-.14	.10	-.37*	.17	.01	.27	-.02	-.45**	.28	.25	-.02	.07	-.13	-.03	-.04	-.05
4. Total correct responses				—	-.14	-.04	.15	-.24	-.11	-.01	-.20	-.05	-.29	.05	.27	-.24	-.20	-.20	.01	.01	.28
RAPM																					
5. RAPM total score					—	.10	.18	.03	.38*	.16	.48**	.40*	-.11	.18	-.22	.37*	.32	.28	-.13	-.13	-.26
Objective Numeracy Scale																					
6. Objective numeracy general score						—	.81**	.76**	.43*	.52**	.18	.45**	-.19	.25	.03	.42*	.28	.42*	.38*	.39*	.28
7. General numeracy							—	.22	.21	.27	.08	.44*	-.34	.36	.12	.14	.09	.15	.27	.27	.24
8. Expanded numeracy								—	.48**	.56**	.21	.27	.06	.01	-.09	.54**	.37*	.53**	.32	.35	.20
Subjective Numeracy Scale																					
9. Subjective numeracy general score									—	.83**	.81**	.44*	-.09	.24	-.05	.67**	.56**	.53**	.05	.06	-.09
10. Cognitive ability										—	.35*	.29	-.03	.13	-.01	.55**	.48**	.41*	.18	.20	.01
11. Information display											—	.44*	-.12	.27	-.09	.55**	.44*	.47**	-.11	-.11	-.17
Cognitive Reflection Task																					
12. Cognitive reflection general score												—	-.07	.18	-.13	.07	-.01	.13	.25	.26	.15
Go/no-go task																					
13. Reaction Time													—	-.76**	-.62**	.03	-.04	.10	.04	.03	.18
14. Proportion of correct responses														—	.30	.11	.14	.05	.02	.01	-.13
15. Proportion of false alarm															—	-.14	.02	-.28	.03	.03	-.06
Need for Cognition Scale																					
16. General need for cognition score																—	.85**	.78**	.05	.07	-.17
17. Rational ability																	—	.34	.19	.21	-.28
18. Rational engagement																		—	-.13	-.12	.03
Faith in Intuition Scale																					
19. Faith in intuition general score																			—	.99**	.54**
20. Experiential ability																				—	.54**
21. Experiential engagement																					—

Note. *$p < .05$; **$p < .01$

Table 36c

Adult Tetrachoric, Polychoric and Pearson Correlations Between Ratio-bias Accuracy and Scales With their Subscales in the Vaccination Problem (N = 31)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Ratio-bias task																					
1. 9-in-10 vs. 85-in-95	—	-.03	-.13	.48**	-.06	.09	.16	-.02	.10	.15	-.01	.12	-.35*	.05	.66**	.20	.12	.25	.36*	.36*	.16
2. 9-in-10 vs. 90-in-100		—	.48**	.71**	.09	-.27	-.15	-.22	-.46**	-.58**	-.15	-.11	-.16	-.03	.37*	-.17	-.21	-.10	.07	.07	.17
3. 9-in-10 vs. 95-in-105			—	.69**	.35*	.17	.36*	-.15	-.22	-.20	-.18	.19	-.01	-.25	-.07	-.16	-.13	-.16	.05	.05	.19
4. Total correct responses				—	.14	-.01	.14	-.16	-.27	-.30	-.15	.07	-.26	-.09	.47**	-.05	-.11	.01	.21	.21	.21
RAPM																					
5. RAPM total score					—	.56**	.61**	.18	.18	.21	.08	.37*	.09	-.10	-.05	.22	.18	.21	-.13	-.13	.09
Objective Numeracy Scale																					
6. Objective numeracy general score						—	.68**	.71**	.35*	.32	.29	.45**	.04	.01	-.04	.25	.26	.20	-.15	-.15	.20
7. General numeracy							—	-.03	.18	.17	.14	.45**	.10	-.21	.01	.21	.22	.18	-.12	-.12	.20
8. Expanded numeracy								—	.30	.26	.27	.19	-.03	.22	-.05	.14	.16	.10	-.09	-.09	.09
Subjective Numeracy Scale																					
9. Subjective numeracy general score									—	.90**	.83**	.41*	-.13	.14	-.03	.74**	.78**	.56**	-.14	-.14	-.15
10. Cognitive ability										—	.53**	.28	-.14	.14	-.06	.68**	.73**	.51**	.01	.01	-.07
11. Information display											—	.46**	-.09	.10	.03	.60**	.62**	.47**	-.28	-.27	-.20
Cognitive Reflection Task																					
12. Cognitive reflection general score												—	.21	-.36*	-.02	.37*	.35*	.32	-.19	-.19	-.15
Go/no-go task																					
13. Reaction Time													—	-.75**	-.72**	.14	-.04	-.22	-.37*	-.38*	-.01
14. Proportion of correct responses														—	.24	.20	.09	.28	.28	.29	-.08
15. Proportion of false alarm															—	.03	-.05	.12	.25	.25	.06
Need for Cognition Scale																					
16. General need for cognition score																—	.92**	.91**	.07	.06	-.01
17. Rational ability																	—	.69**	.10	.10	.01
18. Rational engagement																		—	.02	.01	.01
Faith in Intuition Scale																					
19. Faith in intuition general score																			—	.99**	.63**
20. Experiential ability																				—	.63**
21. Experiential engagement																					—

Note. $*p < .05$; $**p < .01$

Second, consider the correlations between accuracy in the volleyball problem and these ability measures. Total accuracy in the ratio-bias task was not correlated significantly with any measure, and there were few significant correlations with accuracy in individual trials.

Finally, consider the correlations between accuracy in the vaccination problem and ability measures. Total accuracy in the ratio-bias task was positively correlated only with a sub-measure of the Go/no-go task.

5.4 Discussion

According to traditional dual-process theories, time pressure should diminish the time available for slower analytic processes to reach a conclusion and inhibit the heuristic response. Thus, when pressured for time, participants should make more frequent errors and show greater bias. According to UTT and FTT, however, time pressure should provide less time for generic or context-specific heuristic responses to emerge. Participants should exhibit less bias.

Experiment 3 was conducted using the same participant populations, methods, and scenarios as in Experiments 1 and 2 except that participants had all the time they wanted to respond in the first two experiments but were allowed only 5 seconds in Experiment 3. These similarities support cautious comparisons between experiments.

The mean proportions correct in Experiment 1 for seventh graders, adolescents, and adults were .54, .56, and .79, respectively. In the mathematics problem of Experiment 3 these proportions were .44, .51, and .81, respectively. The seventh graders were substantially less accurate, as traditional dual-process theories predict, but the differences in accuracy for the older participants were small. There was no evidence of bias in the responses of seventh graders in Experiment 1 (37% of their responses were Container A and 37% were Container B), but evidence of a bias for the more numerous alternative in Experiment 3 (29% chose Container A and 39% chose Container B). In contrast, the adolescents exhibited a strong bias toward the less numerous alternative in both Experiment 1 (41% of responses were Container A and 18% were container B) and Experiment 3 (45% of responses were Container A and 23% were container B).

The mean proportions correct in the volleyball problem of Experiment 2 for seventh graders, adolescents, and adults were .46, .51, and .75, respectively. In the volleyball problem of Experiment 3 these proportions were .38, .54, and .74, respectively. Again, the seventh graders were less accurate but the differences in accuracy for adolescents and adults were quite small. The seventh graders' responses were slightly biased toward the smaller team in Experiment 2 (37% for Team A and 31% for Team B) and in Experiment 3 (43% for Team A and 36% for Team B). Adolescents were strongly biased toward the smaller team in Experiment 2 (43% for Team A and 18% for Team B) but showed no evidence of bias in Experiment 3 (33% for Team A and 36% for Team B).

The mean proportions correct in the vaccination problem of Experiment 2 for seventh graders, adolescents, and adults were .47, .57, and .75, respectively. In the vaccination problem of Experiment 3 these proportions were .39, .52, and .71, respectively. Again, accuracy of the seventh graders was most affected by time pressure. In this problem the seventh graders were strongly biased toward the smaller clinic in both Experiment 2 (44% for Clinic A and 15% for Clinic B) and Experiment 3 (49% and 28%). Adolescents were also strongly biased toward the smaller clinic in both Experiment 2 (31% for Clinic A and 24% for Clinic B) and Experiment 3 (40% for Clinic A and 26% for Clinic B).

Taken together, these results show that heuristics involved in responding to the ratio-bias task depend on both context and age. The vaccination problem triggered a strong bias toward the smaller clinic among both seventh graders and adolescents. In the mathematics problem the seventh graders were biased toward the larger container when under time pressure but adolescents exhibited a strong bias toward the smaller container.

The effects of time pressure appear to have been age dependent, with the youngest participants showing the greatest reduction in accuracy and adults showing no systematic reduction in accuracy. The effect of time pressure on biases was inconsistent. In the mathematics problem seventh graders had no bias without time pressure but a bias toward the more numerous container with time pressure. In contrast, adolescents were strongly biased toward the smaller team in the volleyball problem without time pressure but had no bias with time pressure.

Overall, time pressure deactivates context-related heuristics that generate (as we saw in Experiment 2) different patterns of biased answers in everyday-life (contextualized) problems, as we predicted. At the same time, time pressure does not change the pattern of biased answers that we found in the mathematical (decontextualized) problem, differently from our predictions. We assumed that biased answers in all three problems are related to heuristic processes. However, results in Experiment 3 suggest that, although ratios considered are the same through three experiments, decontextualized and contextualized problems operate differently.

Specifically, it is possible that, in decontextualized problems, the heuristic processing is not involved at all. As a consequence, biased answers are related to the analytic processing and, according to Siegler (1981), they might correspond to different levels of strategies which are early-acquired during development. This should be visible in seventh graders whose heuristic processing is still developing.

Previous findings (Agnoli, Dellai, Furlan, & Stragà, 2009) showed that younger children (third graders, fourth graders, and fifth graders) have a strong bias toward the large container without time pressure. Such bias disappears in seventh graders. The pattern of correlations in Experiment 3 is the same as in Experiment 1. The bias of seventh graders for the more numerous alternative in Experiment 3 suggests that time pressure favors the application of Rule I (Siegler, 1981) as it does with younger children in the condition without time pressure.

The bias of middle adolescents toward the small container in Experiment 3 suggests that, with time pressure, participants relied on a more sophisticated formal heuristic that corresponds to Rule II (Siegler, 1981). At the same time, participants are required to give a rapid response based on ratios, rather than calculation. As a consequence, according to FTT, participants rely on gist processing. We hypothesize that participants rely on what Fischbein (1987) called secondary intuition, which is acquired through educational intervention. When participants are high in formal reasoning, numeracy and cognitive ability, they have a good performance in three problems because they correctly apply gist processing based on ratios.

This explanation is supported by the pattern of correlations of adolescents in the volleyball problem. Contrary to Experiment 1, in Experiment 3 the total number of correct responses in the ratio-bias task (mathematical scenario) is positively related to verbal ability, objective numeracy, subjective numeracy, cognitive reflection, and need for cognition. According to Wilson and Schooler (1991), thinking too much can sometimes lead to poorer choices. Instead, relying on gist processing favors correct decisions in decontextualized situations, in accordance to FTT. The same is true for adults. The total number of correct responses in the ratio-bias task is positively related to intelligence, objective numeracy, and cognitive ability. These correlations, absent in Experiment 1, seem to indicate that with time pressure several indicators of cognitive ability predict performance in the ratio-bias problem.

Interestingly, contrary to Experiment 1, the correct responses of adults in the mathematical problem are also positively related to the proportion of correct answers (.55) and negatively related to the reaction time in the go/no-go task (-.39). This finding suggests that with time-pressure those participants who are really capable to make fast decisions are the best in the ratio-bias task. These participants are capable to process information very fast.

Finally, the negative correlation between the number of correct responses in the ratio-bias task and the scores in the scale faith in intuition suggests that the view of "bad" intuition as considered by traditional dual-process theorists is different from the one held by Fischbein (1987) and by the proponents of FTT conception of "good" intuition. Negative correlations in Experiment 3 suggest that adults that correctly apply intuition based on ratios are less prone to decide on gut feelings. This result may be interpreted as evidence that time pressure does not inhibit analytic processing.

In everyday-life problems we hypothesize the heuristic process operates by default as predicted by dual-process theory. Nevertheless, we also hypothesized that time pressure inhibits the activation of context-based heuristics. Remember that participants had only few seconds to respond after they read the problem.

With time pressure we note that there is hardly any pattern of significant correlations in the vaccination problem (both in adolescents and adults). Instead, we see that the proportion of correct responses by adolescents in the ratio-bias task correlate positively with calculation ability (.43), subjective numeracy (.51), cognitive reflection (.43) and need for cognition (.42). Remember that in Experiment 2 adolescents exhibited a strong bias toward the small team; this bias disappears in Experiment 3. Time pressure condition may inhibit heuristic processing only when the contextualized representation is strong. Italian adolescents commonly play volleyball both at school and outside school.

Chapter 6

General discussion

The aim of the present dissertation was to investigate ratio biases that, as Reyna and Brainerd (1994; 2008) reviewed, are possible explanations for the great difficulty that children and adults have to understand and manage probability information. The concept of ratio is critical in everyday-life experience: medical decision-making, risk perception, risk taking, economic decisions, or gambling are all areas in which correct decisions make a difference in somebody's life. To understand how the ability to make decisions, based on ratio concepts, develops is fundamental to improve education in statistical thinking.

Markovits and Barrouillet (2004) in a special issue of *Thinking & Reasoning* explain how research whose aim is to trace the development of reasoning has become relatively rare (p. 114). Research looking at children's competence and research looking at adults' reasoning rarely interact and this is often a source of fragmentation and contradictory results. We argue that the current literature on adult decision-making often neglect developmental findings. Jacobs and Klaczynski (2002) pointed out that, recent perspectives of cognitive development (Kokis, et al., 2003; Stanovich & West, 2000) follow the core assumption that development is a progression from states of limited understanding and complexity to more advanced understanding, computational complexity, and abstraction. Consequently, a tacit assumption in decision-making research is that, for instance, if adults show a bias, then children are expected to perform worse in similar judgment tasks.

In the present work, we studied the ratio-bias phenomenon in order to create a bridge between developmental and adult research. We started from two prominent dual-process theories in the current literature on judgment and decision-making that provide different explanations for ratio biases and offer opposite predictions regarding the direction of the bias with high probability events: CEST and FTT. From a developmental perspective, while CEST claims that with adulthood the balance of influence between the two processes shifts in the direction of increased rational dominance, FTT states that intuition is the apex of development and reasoning develops from quantitative to qualitative. In our three experiments, we showed that neither CEST predictions nor FTT predictions are confirmed totally. The bias is *not* always toward the less numerous quantity or always toward the more numerous quantity. If we consider only the direction of the bias with development disregarding the difficulty of ratios and the specificity of the context, then we conclude, coherently with previous results (Agnoli, Dellai, Furlan, & Stragà, 2009), that there is a general tendency to commit denominator neglect until the age of ten. The preference for the less numerous quantity becomes stronger in seventh graders and predominant in adolescents. To summarize, the direction of the bias shifts during development.

According to CEST, we would explain this result by saying that the preference for the less numerous quantity depends on the *concretive* principle and the *experiential learning* principle. These principles develop with age; thus, the bias toward the less numerous quantity increases with age. According to FTT, children apply the gist "more is better than less" but FTT theorists also predict that adolescents and adults would have the same preference for large quantities because they neglect the denominator. In addition, adolescents and adults are more biased than children under certain conditions because children rely on quantitative information more than adolescents and adults who often rely, instead, on qualitative information.

Our experiments show that adolescents and adults do *not* commit denominator neglect in the sense of a preference for the more numerous option. However, the analysis of justifications provides additional perspectives. Remember that most participants who preferred the less numerous option explained their choice by comparing the absolute number of *negative* events. For example, they wrote that 9-in-10 was better than 85-in-95 because of the comparison between 1 losing ball and 10 losing balls. One losing ball appeared as a better option compared to 10 losing balls. We argue that this justification indicates that participants neglect the denominator while preferring the less numerous option because "less is better than more". We propose that adolescents and adults' preference for the small quantity may be explained as *reversed denominator neglect*.

We extensively showed in chapter 1 and 2 that, contrary to traditional accounts of development, arithmetical ability, mathematical competence and numeracy are necessary to process numerical information but not sufficient to lead to a correct decision. Dual-process theories on adult reasoning claim that bad decisions often origin from heuristic processing, but the development of heuristic processing has been scarcely investigated (De Neys & Vanderputte, 2010).

A critical question we tried to answer is the following: is the ratio bias irrational and related to heuristic processing as most decision-making theorists (i.e., CEST theorists) claim? Our findings suggest that the ratio biases are *not* always irrational and they origin from wrong conscious mathematical strategies that we called *formal heuristics* in chapter 3. The application of these strategies is often independent from mathematical competence as shown by the lack of relation to individual differences measures of cognitive ability, shown in Experiments 1 and 2. We also show that everyday-life problems activate context-based representations related to experience that may increase the strength of the bias with age in a time-paced condition.

For example, middle adolescents solving the volleyball problem in Experiment 2 showed a strong bias for the small team, but their choice was mathematically reasoned as seen in their justifications. According to traditional dual-process accounts, participants read the problem, activate some fast and contextualized heuristic that favors the small team and, finally, they give the correct/wrong response according to the strength of analytic processing to override the heuristic. The present results in Experiment 3 are critical because, according to traditional dual-process accounts, time pressure would inhibit analytic processing. Consequently, the bias of middle adolescents toward the small team should be stronger than in Experiment 2. Instead, our results show no evidence of bias in Experiment 3. This finding suggests that, in those problems that generate strong context-based representations, to have time to decide favors generating context-based answers.

Time pressure does not inhibit only the analytic process but, rather, both the analytic and heuristic processes. General accuracy decreases because those participants who used to rely on calculation (verbatim process according to FTT) do not have enough time to give their answer. At the same time, context-based heuristics are also inhibited. Time pressure forces to give responses based on ratios and secondary intuition (Fischbein, 1987), or gist processing (Reyna & Brainerd, 1991), are favored.

Contrary to traditional dual-process accounts, heuristic decision-making does not rely *always* on fast automatic processes, and analytic decisions do not *always* entail slower control processes when participants process numerical information. As many authors claim, dual-process theories nicely describe ‘what’ the two processes do but it is not clear ‘how’ the two processes actually operate (De Neys & Glumicic, 2008; Keren & Schul, 2009; Osman, 2004). We argue that a possible explanation of such difficulty to understand how the two processes operate, origins from the theoretical dichotomy between what is fast and what is slow, especially from a developmental perspective.

This issue is addressed in our experiments in which two kinds of information are presented at the same time: quantitative information related to numbers and qualitative information related to context. A relevant distinction should be made between what is automatic and what is automatized. Some context-based heuristics may be considered as automatic and they become more automatic with age. For example, to associate a large clinic to more experience and thus considering a large clinic better than a small clinic, is an experience-related qualitative thinking that is common and automatic in adults. In our sample of Cornell undergraduates, 10% of participants in Experiment 2 gave such explanations to their preference for the large clinic. The application of this heuristic disappears with time pressure.

Automatic heuristics differ from automatized reasoning processes on numbers and ratios. For example, in the comparison between 9-in-10 and 90-in-100, Italian middle adolescents performed quite similarly to Cornell undergraduates in all three experiments even though their mathematical competence is lower. Adolescents and adults performed better than seventh graders and we interpret this finding with an increased and automatized ability to simplify fractions.

With different ratios, the pattern of results is more complex. According to Siegler (1981), there is a hierarchy of rules (Chapter 2, p. 46) that develops with age until sufficient working memory capacity and cognitive development allow processing numerators and denominators correctly. The rule assessment method defines rule use in terms of an individual's consistent application of an algorithm across a large and diverse set of problems (Siegler & Chen, 2002, p. 448). At the same time, during development, children balance old and new rules and often apply multiple strategies. Children are especially likely to use rules or systematic goal-directed approaches when the task is unfamiliar.

Siegler (1996) proposed the overlapping waves model as a way of representing children's use of strategies and rules. The overlapping waves model's basic assumptions are: that children at any given time typically have various strategies of thinking about a problem; that strategies usually coexist and compete each other over a prolonged period of time, with the more effective ones gradually displacing the less effective ones; and that with age and experience, children increase their use of advanced approaches. For example, in the mathematical problem in a time-paced condition (Experiment 1), our sample of seventh graders did not show a clear bias, but their justifications indicate that the applications of Rule I and Rule II coexist.

We agree with Siegler's overlapping waves model and our data provide evidence that with unfamiliar tasks, such as the present task with different ratios, seventh graders and middle adolescents rely on automatized rules that are previously acquired during development. In other words, when ratios are difficult to compute, seventh graders and adolescents (who are expected to apply already Rule IV) activate a bottom-up reasoning process and retrieve automatized formal rules that simplify the solution of a problem.

We also showed that, in the mathematical problem, seventh graders exhibit a strong bias toward the large container whereas middle adolescents had a preference toward the small container stronger than in Experiment 1. These findings suggest that those participants who do not have the correct intuition on ratios, rely on formal heuristics that are automatized and related to analytic processing. Specifically, and according to Siegler with regard to the hierarchy of rules, we claim that participants retrieve different levels of algorithms according to age. Seventh graders in time pressure condition retrieve the simplest Rule I coherently to 10 year-old children do without time pressure. Middle adolescents, instead, retrieve more sophisticated rules, that are Rule II, and partially Rule III.

Taken together the application of formal heuristics and the strength of the context-representation to activate experience-based heuristics, our results suggest that, according to Stanovich (1999), the ability to decontextualize is relevant. If a problem is decontextualized by default, as the mathematical problem is, then participants' errors depend on formal heuristics when the information is unfamiliar or difficult to process. If a problem is contextualized, as the volleyball and the vaccination problems are, we claim that experience-based heuristics operate in parallel with the analytic processing of numerical information. Experience-based heuristics' strength increase coherently to the level of participants' personal involvement in the context considered. For example, middle adolescents' positive correlation (.29) in Experiment 2 (volleyball problem) between accuracy in the ratio-bias task and the Cognitive Reflection Task indicates that adolescents activated a decontextualization. In the vaccination problem, this correlation is not significant denoting that the vaccination problem probably does not activate strong context-based representation.

Another critical point we developed in this dissertation is timing. What happens when participants are asked to decide in just few seconds? According to traditional dual-process accounts, participants would decrease in accuracy because the slow analytic processing is inhibited (Evans & Holmes, 2005). Our data confirm this prediction. At the same time, according to traditional dual-process accounts, biased answers based on rapid heuristic processing would increase, because the analytic processing is inhibited and the heuristic processing is "free to decide". Our findings show a more complicated pattern of results. Again, we try to interpret such results in light of developmental and adults decision-making research together.

The influence of time pressure on responses is to inhibit both the heuristic and the analytic processing with regard to time-consuming processes belonging to each process. In other words, time pressure inhibits those context-based heuristics that are *not* already automatic and strongly rooted in everyday-life experience, and inhibits what FTT calls verbatim process that is the slow and sequential arithmetical calculation.

When the scenario is decontextualized, time pressure operates only on the analytic processing because there are not context-based representations to inhibit. Time pressure has two effects: on the one hand, as we discussed in Experiment 3 with regard to individual differences, those participants who have the ability to process information very fast respond correctly. FTT defines this ability as being high in the gist processing. Other authors as Fischbein would define this ability as secondary intuition. In any case, we conclude that time pressure erases time-related additional and wrong considerations that are independent from strictly mathematical considerations.

When the scenario is contextualized, results are less clear with regard to the relative contribution of context-based heuristics and formal heuristics. The justifications showed that seventh graders and adolescents relied more on formal heuristics than on context-based heuristics. The interaction between these two kinds of heuristics will be object of future investigation.

References

Acker, F. (2008). New findings on unconscious versus conscious thought in decision making: additional empirical data and meta-analysis. *Judgment and Decision Making*, *3*, 292-303.

Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's ability to make probability estimates: Skills revealed through application of Anderson's functional measurement methodology. *Child Development, 60*, 933-945.

Adelson, E. H. (1995). *Checker-shadow illusion.* http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html.

Agnoli F., Dellai, V., Furlan S., & Stragà C. (2009). *The Ratio-Bias Phenomenon: Different Intuitions at Different Ages*. Society for Research in Child Development, Denver, CO.

Agnoli, F. (1991). Development of judgmental heuristics and logical reasoning: training counteracts the representativeness heuristic. *Cognitive Development*, *6*, 195-217.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, *21*, 515−550.

Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2nd ed.). New York, NY: Wiley.

Aizpurua, A., & Koutstaal, W. (2010). Aging and flexible remembering: Contributions of conceptual span, fluid intelligence, and frontal functioning. *Psychology and Aging, 25*, 193-207.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.

Alonso, D., & Fernandez-Borrocal, P. (2003). Irrational decisions: attending to numbers rather than ratios. *Personality and Individual Differences*, *35*, 1537-1547.

Amir, G., & Williams, J. (1999). Cultural influences on children's probabilistic thinking. *Journal of Mathematical Behavior*, *18,* 85-107.

Amsel, E., Close, J., Sandler, E., & Klaczynski, P. A.(2009). College students' awareness of irrational judgments on gambling tasks: a dual-process account. *The Journal of Psychology*, *143*, 293-317.

Amsel, E., Klaczynski, P. A., Johnston, A., Bench, S., Close, J., Sadler, E., & Walker, R. (2008). A dual-process account of the development of scientific reasoning: the nature and development of metacognitive intercession skills. *Cognitive Development*, *23,* 452-471.

Anderson, N. H., & Schlottmann, A. (1991). Developmental study of personal probability. In N. H. Anderson (Ed.), Contributions *to information integration theory*. Vol. III: Developmental (pp. 110-134). Hillsdale, NJ: Erlbaum.

Arkes, H. R., & Ayton, P. (1999). The sunk cost and concorde effects: are humans less rational than lower animals? *Psychological Bulletin, 125*, 591–600.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*, 124–140.

Arthur, W., Jr., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, *54*, 394–403.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Baayen, R. H., Tweedie, F. J., & Schreuder, R. (2002). The subjects as a simple random effect fallacy: subject variability and morphological family effects in the mental lexicon. *Brain and Language, 81*, 55–65.

Babai, R., Brecher, T., Stavy, R., & Tirosh, D. (2006). Intuitive interference in probabilistic reasoning. *International Journal of Science and Mathematics Education*, *4*, 627-639.

Baddeley, A. D. (1976). *The psychology of memory.* New York: Basic Books.

Baddeley, A. D. (1986).*Working memory.* Oxford: Oxford University Press.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, *49*A, 5-28.

Baron, J. (2008). *Thinking and deciding* (4th ed.). New York: Cambridge University Press.

Barrouillet, P. (1996). Transitive inferences from set-inclusion relations and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1408-1422.

Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, 5, 27-30, URL http://CRAN.R-project.org/doc/Rnews/.

Bates, D. M. (2007). Linear mixed model implementation in lme4. Ms., University of Wisconsin, Madison, August 2007.

Bates, D. M., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes. R package version 0.9975-12.

Bjorklund, D. F. (1989). *Children's thinking*. Forest Grove, CA: Brooks/Cole.

Björklund, F., & Bäckström, M. (2008). Individual differences in processing styles: validity of the Rational–Experiential Inventory. *Scandinavian Journal of Psychology, 49*, 439-446.

Bonner, C., & Newell, B. R. (2008). How to make a risk seem riskier: the ratio bias versus construal level theory. *Judgment and Decision Making, 3*, 411-416.

Bos, M.W., Dijksterhuis, A., & van Baaren, R. B. (2008). On the goal-dependency of unconscious thought. *Journal of Experimental Social Psychology, 44*, 1114-1120.

Boyd, R., & Richerson, P.J. (2005). *The origin and evolution of cultures*. New York: Oxford University Press.

Boyer, T. W. (2007). Decision-making processes: Sensitivity to sequentially experienced outcome probabilities. *Journal of Experimental Child Psychology, 97*, 28–43

Braine, M. D. S. (1959). The ontogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. *Psychological Monographs, 73*, 1-43.

Brainerd, C. J. (1979). Symposium on the future of Piagetian psychology. *Genetic Epistemologist*.

Brainerd, C. J. (1981). Working memory and the developmental analysis of probability judgment. *Psychological Review, 88*, 463-502.

Brainerd, C. J. (1983). Young children's mental arithmetic errors: A working-memory analysis. *Child Development, 54*, 812-830.

Brainerd, C. J. (1985). Model-based approaches to storage and retrieval development. In C. J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development* (pp. 143–208). New York: Springer-Verlag.

Brainerd, C. J. (2004). Dropping the other U: an alternative approach to U-shaped developmental functions. *Journal of Cognition & Development, 5*, 81-88.

Brainerd, C. J., & Kingma, J. (1984). Do children have to remember to reason? A fuzzy-trace theory of transitivity development. *Developmental Review, 4*, 311-377.

Brainerd, C. J., & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology, 17*, 210–247.

Brainerd, C. J., & Reyna, V. F. (1988). Generic resources, reconstructive processing, and children's mental arithmetic, *Developmental Psychology, 24*, 324-334.

Brainerd, C. J., & Reyna, V. F. (1990a). Gist is the grist: Fuzzy-Trace Theory and the new intuitionism. *Developmental Review, 10*, 3-47.

Brainerd, C. J., & Reyna, V. F. (1990b). Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. *Developmental Review, 10*, 365-403.

Brainerd, C. J., & Reyna, V. F. (1992). Explaining “memory free” reasoning. *Psychological Science, 3*, 332-339.

Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review*, *100,* 42– 67.

Brainerd, C. J., & Reyna, V. F. (1995). Autosuggestibility in memory development. *Cognitive Psychology*, *28*, 65−101.

Brainerd, C. J., & Reyna, V. F. (1998a). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71*, 1-78.

Brainerd, C. J., & Reyna, V. F. (1998b). When things that never happened are easier to remember than things that did. *Psychological Science*, *9,* 484-489.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. In H. W. Reese & R. Kail (Eds.), *Advances in child development and behavior* (Vol. 28, pp. 41–100). San Diego: Academic Press.

Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review*, *24,* 396-439.

Brainerd, C. J., Reyna, V. F., & Poole, D. A. (2000). Fuzzy-trace theory and false memory: Memory theory in the courtroom. In D. F. Bjorklund (Ed.), *False memory creation in children and adults* (pp. 93-128). Mahwah, NJ: Erlbaum.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Society*, *88*, 9–25.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children’s mathematical ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*, 273–293.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). New York, NY: Springer.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116-131.

Cacioppo, J. T., Petty, R. E., Feinstein, J., & Jarvis, W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119,* 197–253.

Callahan, P. (1989). *Learning and development of probability concepts: Effects of computer assisted instruction and diagnosis.* Unpublished doctoral dissertation, University of Arizona, College of Education, Tucson. Mentioned in Reyna & Brainerd (1994).

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, *3*, 182–191.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representations of faces. *Science*, *195*, 312-314.

Carlson, J. (1970). The development of probabilistic thinking in children: a comparison of two methods of assessment. *Journal of Genetic Psychology, 116*, 263-269.

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Case, R. (1992). The role of the frontal lobes in the regulation of cognitive development. *Brain and Cognition*, *20*, 51–73.

Chaiken, S., & Trope, Y. (1999). *Dual process theories in social psychology*. New York, NY: The Gilford Press.

Chapman, M., & Lindenberger, U. (1988). Functions, operations, and decalage in the development of transitivity. *Developmental Psychology*, *24*, 542–551.

Chapman, R. H. (1975). The development of children's understanding of proportions. *Child Development*, *46*, 141-148.

Cohen, A. R., Scotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *Journal of Abnormal and Social Psychology*, *51*, 291-294.

Cornoldi, C., & Cazzola, C. (2003). *AC-MT 11-14 - Test di valutazione delle abilità di calcolo e problem solving dagli 11 ai 14 anni*. Trento, Erickson.

Cornoldi, C., & Colpo, G. (1995). *Nuove prove di lettura MT per la scuola media inferiore*, Organizzazioni Speciali, Firenze.

Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harper.

Dale, D., Rudski, J., Schwarz, A., & Smith, E. (2007). Innumeracy and incentives: a ratio bias experiment. *Judgment and Decision Making*, *2*, 243-250.

Daniel, D. B., & Klaczynski, P. A. (2006). Developmental and individual differences in conditional reasoning: Effects of logic instructions and alternative antecedents. *Child Development*, *77*, 339-354.

Davidson, D. (1995). The representativeness heuristic and conjunction fallacy effect in children's decision-making. *Merrill-Palmer Quarterly*, *41*, 328-346.

Davidson, D., Suppes, P., & Siegel, S. (1957). *Decision making: an experimental approach.* Stanford: Stanford University Press.

Davies, R. (2006). *Decision making and alcohol use in adolescents: A dual process approach*. Unpublished Master thesis, University of New England, Armidale, New South Wales, Australia. In Marks, Hine, Blore, & Phillips, W. J. (2008).

De Neys, W. (2006a). Dual processing in reasoning—two systems but one reasoner. Psychological *Science*, *17*, 428–433.

De Neys, W. (2006b). Automatic-heuristic and executive – analytic processing during reasoning: chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology, 59*, 1070-1100.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theory of thinking. *Cognition*, *106*, 1248-1299.

De Neys, W., & Vanderputte, K. (in press). When less is not always more: Stereotype knowledge and reasoning development. *Developmental Psychology*.

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, and Behavioral Neuroscience, 10,* 208-216.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and counterexample retrieval for causal conditionals. *Thinking and Reasoning, 11,* 123-150.

DebRoy, S., & Bates, D. M. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis*, *91*, 1–17.

Dellai, V. (2007). *Sviluppo del ragionamento analitico e dell'intuizione: Il caso del Ratio Bias* [The development of analytic thinking and intuition: the case of the ratio-bias]. Unpublished Thesis, University of Padua, Italy.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica*, *101*, 395-416.

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.

Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, *21*, 1083–1092.

Dijksterhuis, A. (2004). Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology, 87*, 586–598.

Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1*, 95–109.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On Making the Right Choice: the deliberation-without-attention effect. *Science*, *311*, 1005–1007.

Dixon, J. A., & Moore, C. F. (1996). The developmental role of intuitive principles in choosing mathematical strategies. *Developmental Psychology, 32*, 241-253.

Donovan, S., & Epstein, S. (1997). The difficulty of the Linda conjunction problem can be attributed to its simultaneous concrete and unnatural presentation, and not to conversational implicature. *Journal of Experimental Social Psychology, 33,* 1–20.

Epstein, S. (1973). The self-concept revisited, or a theory of a theory. *American Psychologist, 28*, 404-416.

Epstein, S. (1990). Cognitive-Experiential Self-Theory, in L. Pervin (Ed.), *Handbook of Personality Theory and Research* (pp. 165-192). New York: Press.

Epstein, S. (1991). Cognitive-experiential self-theory: an integrative theory of personality. In R. Curtis (Ed.), *The relational self: convergences in psychoanalysis and social psychology* (pp. 111–137). New York: Guilford.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49,* 709-724.

Epstein, S. (2003). *Cognitive-experiential self theory of personality*. In T. Millon & M. J. Lerner (Eds.), Handbook of psychology: Vol. 5. Personality and social psychology (pp. 159–184). Hoboken, NJ: Wiley.

Epstein, S., & Pacini, R. (1999). Some basic issues regarding dual-process theories from the perspective of cognitive-experiential self-theory. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 462–482). New York, NY: The Gilford Press.

Epstein, S., & Pacini, R. (2001). A comparison of the influence of imagined and unimagined verbal information on intuitive and analytical information processing. *Imagination, Cognition, and Personality*, *20*, 195–216.

Epstein, S., Denes-Raj, V., & Pacini, R. (1995). The Linda problem revisited from the perspective of Cognitive-Experiential Self-Theory. *Personality and Social Psychology Bulletin*, *21*, 1124-1138.

Epstein, S., Donovan, S., & Denes-Raj, V. (1999). The missing link in the paradox of the Linda conjunction problem: Beyond knowing and thinking of the conjunction rule, the intrinsic appeal of heuristic processing. *Personality and Social Psychology Bulletin*, *25*, 204–214.

Epstein, S., Lipson, A., Holstein, C., & Huh, E. (1992). Irrational reactions of negative outcomes: evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *62*, 328-339.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71,* 390-405.

Estrada, C., Barnes, V., Collins, C., & Byrd, J. C. (1999). Health literacy numeracy. *Journal of the American Medical Association, 282*, 527.

Evans J. St. B. T., Newstead, S. N., & Byrne, R. (1993). *Human reasoning: the psychology of deduction*. Lawrence Erlbaum Associates, Hove, UK.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* London: Erlbaum Associates.

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. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*, 378-395.

Evans, J. St. B. T. (2007). *Thinking: Dual Processes in Reasoning and Judgment*. Hove: Psychology Press.

Evans, J. St. B. T. (2008). Dual processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278.

Evans, J. St. B. T. (2009). How many dual process theories do we need: One, two or many? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford, UK: Oxford University Press.

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*, 382-389.

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

Fagerlin, A., Ubel, P. A., Smith, D. M., & Zikmund-Fisher, B. J. (2007). Making numbers matter: Present and future research in risk communication. *American Journal of Health Behavior*, *31*, S47–S56.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*, 672-680.

Falk, R., & Wilkening, F. (1998). Children’s construction of fair chances: adjusting probabilities. *Developmental Psychology*, *34*, 1240–1357.

Faraway, J. J. (2006). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Boca Raton, FL: Chapman and Hall.

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.

Fiedler, K. (1988) The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, *50*, 123–29.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. The Netherlands: Reidel, Dordrecht.

Fischbein, E. (1987). *Intuition in science and mathematics*. The Netherlands: Reidel, Dordrecht.

Fischbein, E. (1990). Intuition and information processing in mathematical activity. *International Journal of Educational Research, 14,* 31-50.

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

Fischbein, E., Pampu, I., & Manzat, I. (1970). Comparison of ratios and the chance concept in children. *Child Development*, *41*, 377-389.

Fisk, J. E., Bury, A. S., & Holden, R. (2006). Reasoning about complex probabilistic concepts in childhood. *Scandinavian Journal of Psychology*, *47*, 497–504.

Flavell, J., Miller, P., & Miller, S. (2002). *Cognitive development* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Fodor J. (1983). *The Modularity of Mind*. Scranton, PA: Crowell.

Fodor J. (2001). *The mind doesn't work that way*. Cambridge, Mass.: MIT Press.

Frank, M. J., Cohen, M. X., & Sanfey, A. G. (2009). Multiple systems in decision making: A neurocomputational perspective. *Current Directions in Psychological Science*, *18*, 73-77.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25-42.

Freud, S. (1953). The interpretation of dreams. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vols. 4 and 5). London: Hogarth. (Original work published 1900)

Freud, S. (1959). *Beyond the pleasure principle.* New York: Norton. (Original work published 1920)

Fudenberg, D., & Levine, D. K. (2006). Superstition and Rational Learning. *American Economic Review*, *96*, 630-651.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: Inhibition, error detection, and correction. *NeuroImage*, *17*, 1820-1829

Gigerenzer, G. (2008). Why heuristics work. *Perspective on Psychological Science, 3*, 20-29.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103,* 650–669.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple Heuristics that Make Us Smart* (pp. 75–95). New York: Oxford University Press.

Girotto, V., & Gonzales, M. (2008). Children's understanding of posterior probability. *Cognition*, *106*, 325-344.

Goldberg, S. (1966). Probability judgments of preschool children: Task conditions and performance. *Child Development*, *37*, 158-167

Gopnik, A. (1996). The Post-Piaget era. *Psychological Science*, *7*, 221-225.

Hammond, K. R. (1996). *Human judgment and social policy*. Oxford University Press.

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

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: why small samples? *Cognition*, *115*, 225-237.

Hoemann, H. W., & Ross, B. M. (1971). Children's understanding of probability concepts. *Child Development*, *42*, 221-236.

Holyoak, K. J. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology*, *10*, 203-24

Houdé, O. (2007). First insights on neuropedagogy of reasoning. *Thinking & Reasoning*, *13*, 81-89.

Houdè, O., & Guichart, E. (2001). Negative priming effect after inhibition of number/ length interference in a Piaget-like task. *Developmental Science*, *4*, 119-223.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Inhelder, B., & Piaget, J. (1964). *The Early Growth of Logic in the Child*. New York: Routlege and Kegan (original in French, 1959).

Jacobs, J. E., & Klaczynski, P. A. (2002). The development of judgment and decision making during childhood and adolescence. *Current Directions in Psychological Science*, *11*, 145-149.

Jacobs, J. E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development*, *62*, 166-178.

Jaeger, F. T. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-446.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology*, *35*, 904–911.

Jones, G. A., & Thornton, C. A. (2005). An overview of research into the teaching and learning of probability. In G. A. Jones (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning*, (pp. 66-92). New York: Springer.

Kahneman, D. (2003). A perspective on judgment and choice. Mapping bounded rationality. *American Psychologist, 58*, 697-720.

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

Kahneman, D., & Miller, D. T. (1986). Norm theory: comparing reality to its alternatives. *Psychological Review*, *93*, 136-153.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430-454.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*, 111-132.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic & Tversky (Eds.), *Judgment under certainty*: *Heuristics and biases* (pp. 201–208). New York, NY: Cambridge University Press.

Kahneman, D., & Tversky, A. (1986). On the reality of cognitive illusions. *Psychological Review*, *103*, 582-591.

Kemmelmeier, M. (2010). Authoritarianism and its relationship with intuitive-experiential cognitive style and heuristic processing. *Personality and Individual Differences*, *48*, 44-48.

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

Kieren, T. E. (1988). Personal knowledge of rational numbers: its intuitive and formal development. In J. Hiebert, & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 162-181). Reston, VA: NCTM; Hillsdale, NJ: Lawrence Erlbaum.

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. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, *71*, 1347-1366.

Klaczynski, P. A. (2001a). Analytic and heuristic processing influences on adolescent reasoning and decision-making. *Child Development*, *72*, 844-861.

Klaczynski, P. A. (2001b). 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* (pp. 73–123). San Diego, CA: Academic Press.

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

Klaczynski, P. A., & Cottrell, J. E. (2004). A dual-process approach to cognitive development: the case of children's understanding of sunk cost decisions. *Thinking & Reasoning*, *10*, 147-174.

Klaczynski, P. A., & Narasimham, G. (1998). Development of scientific reasoning biases: cognitive versus ego-protective explanations. *Developmental Psychology*, *34*, 175–187.

Klaczynski, P. A., Fauth, J. M., & Swanger, A. (1998). Adolescent identity: rational vs. experiential processing, formal operations, and critical thinking beliefs. *Journal of Youth and Adolescence*, *27*, 185-207.

Klaczynski, P. A., Gordon, D.H., & Fauth, J. (1997). Goal-oriented critical thinking biases and individual differences in reasoning biases. *Journal of Educational Psychology*, *89*, 470-485.

Kline, R. B. (2004). *Beyond significance testing*. *Reforming data analysis methods in behavioural research*. Washington, DC: American Psychological Association.

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.

Kühberger, A., & Tanner, C. (2009). Risky choice framing: task versions and a comparison of prospect-theory and fuzzy-trace theory. *Journal of Behavioral Decision Making*, *23*, 314-329.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*, 178–181.

Kuhn, D. (2009). Adolescent thinking. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., Vol. 1, pp. 152–186). New York: Wiley.

Kuhn, D., & Pearsall, S. (2000). Development of origins of scientific thinking. *Journal of Cognition and Development*, *1*, 113–129.

Kwon, Y. J., Lawson, A. E., Chung, W. H., & Kim, Y. S. (2000). Effect on development of proportional reasoning skill of physical experience and cognitive abilities associated with prefrontal lobe activity. *Journal of Research in Science Teaching*, *37*, 1171–1182.

Kyllonen, P. C., & Christal, R. E. (1990) Reasoning ability is (little more than) working memory capacity?! *Intelligence*, *14*, 389–433.

Liben, L. S., & Posnansky, C. J. (1977). Inferences on inference: the effects of age, transitive ability, memory load, and lexical factors. *Child Development*, *48*, 1490-1497.

Lieberman, M. D. (2003). Reflective and reflexive judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. R. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 44-67). New York: Cambridge University Press.

Lindeman, M., & Aarnio, K. (2006). Paranormal beliefs: Their dimensionality and correlates. *European Journal of Personality*, *20*, 585–602.

Lipkus, I. M., & Peters, E. (2009). Understanding the role of numeracy in health: proposed theoretical framework and practical insights. *Health Education and Behavior*, *36*, 1065-1081.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37–44.

Mack, N. K. (1990). Learning fractions with understanding: building on informal knowledge. *Journal for Research in Mathematics Education*, *21*, 16-32.

Macpherson, R. (2001). *Analytic and heuristic processing in development of statistical reasoning in children.* Master Thesis, University of Toronto.

Markovits, H., & Barrouillet, P. (2004). Introduction: Why is understanding the development of reasoning important? *Thinking & Reasoning*, *10*, 113-121.

Markovits, H., & Dumas, C. (1999). Developmental patterns in the understanding of social and physical transitivity. *Journal of Experimental Child Psychology*, *73*, 95-114.

Markovits, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking and Reasoning*, *8*, 97 – 107.

Marks, A. D. G., Hine, D. W., Blore, L. R., & Phillips, W. J. (2008). Assessing individual differences in adolescents' preference for rational and experiential cognition. *Personality and Individual Differences*, *44*, 42-52.

McCullagh P., & Nelder, J. A. (1989). *Generalized linear models*. 2nd ed., Chapman & Hall Ltd., London.

Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science, 4*, 379-383.

Miller G. A. (1956). The magical number seven plus or minus two. *Psychological Review, 63,* 81-97.

Miller, D. T, Turnbull, W, & McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology*, *57*, 581-589.

Moore, C. F., Dixon, J. A., & Haines, B. A. (1991). Components of understanding in proportional reasoning: A fuzzy set representation of developmental progressions. *Child Development*, *62*, 441-459.

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.

Myung, I. J., Forster, M. R., & Browne, M. W. (Eds.) (2000). Model selection [Special section]. *Journal of Mathematical Psychology*, *44*, 1-2.

Nelder, J. A., & Wedderburn, R.W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, *135*, 370-384.

Nelson, W., Reyna, V. F., Fagerlin, A., Lipkus, I., & Peters, E. (2008). Clinical implications of numeracy: theory and practice. *Annals of Behavioral Medicine*, *35*, 261-274.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review*, *108*, 291-310.

Novak, T. P., & Hoffman, D. L. (2008). The fit of thinking style and situation: new measures of situation-specific experiential and rational cognition. *Journal of Consumer Research*, *36*, 56-72.

Nunes, T., Schliemann, A.D., & Carraher, D.W. (1993). *Mathematics in the Streets and in Schools.* Cambridge, U.K: Cambridge University Press.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior and Organization*, *29*, 147-152.

Offenbach, S. I., Gruen, G. E., & Caskey, B. J. (1984). Development of proportional response strategies. *Child Development*, *55*, 965-972.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*, 988-1010.

Osman, M., & Stavy, R. (2006). Development of intuitive rules: evaluating the application of the dual-system framework to understanding children's intuitive reasoning. *Psychonomic Bulletin & Review*, *13*, 935-953.

Pacini, R., & Epstein, S. (1999a). The interaction of three facets of concrete thinking in a game of chance. *Thinking and Reasoning, 5*, 303-325.

Pacini, R., & Epstein, S. (1999b). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology***,** *76,* 972-987.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*, 255–287.

Pavan, A. (2005). *Il conflitto tra istruzione formale e risposte intuitive: Il caso del Ratio Bias* [Conflict between formal instruction and intuitive answers: the ratio-bias phenomenon]. Unpublished thesis, University of Padua, Italy.

Pennington, N. (1990). The complexity of human responses to (apparently) simple judgment problems*. Contemporary Psychology*, *35*, 32-33.

Perner, J., & Mansbridge, D. G. (1983). Developmental differences in encoding. *Child Development*, *54,* 710–719.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*, 407–413.

Peterson, J. B., Pihl, R. O., Higgins, D. M., Seguin, J. R., & Tremblay, R. E. (2003). Neuropsychological performance, IQ, personality, and grades in a longitudinal grade-school male sample. *Individual Differences Research*, *3*, 159–172.

Piaget, J. & Inhelder, B. (1975). *The Origins of the Idea of Chance in Children*, W. W. Norton, New York. (Original work published 1951)

Piaget, J. (1930). *The child's conception of causality*. London: Kegan Paul.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International University Press. (Original work published 1936)

Piaget, J., & Inhelder, B. (1973). *Memory and intelligence*. New York: Basic Books.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.

Price, P. C., & Matthews, T. V. (2009). From group diffusion to ratio bias: effects of denominator and numerator salience on intuitive risk and likelihood judgments. *Judgment and Decision Making, 4*, 436-446.

Quené, H., & Van den Bergh, H. (2004). On Multi-Level Modeling of data from repeated measures designs: a tutorial. *Speech Communication*, *43,* 103-12.

Raidl, M.H., & Lubart, T. I. (2000-2001). An empirical study of intuition and creativity. *Imagination, Cognition, and Personality*, *20*, 217–230.

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. New York: Oxford University Press.

Reeves, T., & Lockhart, R. S. (1993). Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology: General*, *122*, 207–226.

Reyna, V. F. (1991) Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review*, *11,* 317–36.

Reyna, V. F. (1995). Interference effects in memory and reasoning: A fuzzy-trace theory analysis. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 29–59). San Diego, CA: Academic Press.

Reyna, V. F. (1996). Conceptions of memory development, with implications for reasoning and decision making. *Annals of Child Development, 12,* 87–118.

Reyna, V. F. (2004). How people make decisions that involve risk: a dual-processes approach. *Current Directions in Psychological Science*, *13*, 60-66.

Reyna, V. F. (2008). A theory of medical decision making and health: fuzzy trace theory. *Medical Decision Making*, *28,* 850-865.

Reyna, V. F., & Adam M. B. (2003). Fuzzy-Trace Theory, risk communication, and product labelling in sexually transmitted diseases. *Risk Analysis, 23,* 325-342.

Reyna, V. F., & Brainerd, C. J. (1990). Fuzzy processing in transitivity development. *Annals of Operations Research*, *23,* 37-63.

Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice. Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making, 4*, 249−262.

Reyna, V. F., & Brainerd, C. J. (1992). A fuzzy-trace theory of reasoning and remembering: Paradoxes, patterns, and parallelism. In N. Hearst, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (pp. 235−259). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Reyna, V. F., & Brainerd, C. J. (1993). Fuzzy Memory and mathematics in the classroom. In G. M. Davies & R. H. Logies (Eds.), *Memory in everyday life* (pp.91–134). Elsevier Science Publishers B.V.

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: a review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability*. New York: Wiley.

Reyna, V. F., & Brainerd, C. J. (1995a). Fuzzy-trace theory: an interim synthesis. *Learning and Individual Differences, 7*, 1-75.

Reyna, V. F., & Brainerd, C. J. (1995b). Fuzzy-trace theory: Some foundational issues. *Learning & Individual Differences*, *7,* 145−162.

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences*, *17*, 147-159.

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*, 89-107.

Reyna, V. F., & Ellis, S. C. (1994). Fuzzy-Trace Theory and framing effects in children's risky decision making. *Psychological Science, 5*, 275-279.

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.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30,* 178−191.

Reyna, V. F., & Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor and Symbolic Activity*, *10*, 309–331.

Reyna, V. F., & Lloyd, F. J. (2006). Physician decision-making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing. *Journal of Experimental Psychology*: *Applied*, *12*, 179-195.

Reyna, V. F., & Narter, C. (1991). *Theoretical implications of children's decision making*. Presented at the 32nd Annual Meeting of the Psychonomic Society, San Francisco, CA.

Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2003). Memory, development, and rationality: An integrative theory of judgment and decision making. In S.L. Schneider & J. Shanteau (Eds.), Emerging *perspectives on judgment and decision research* (pp. 201–245). New York: Cambridge University Press.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*, 943-973.

Reyna, V.F., & Mills, B. A. (2007). Converging evidence supports fuzzy-trace theory's nested sets hypothesis (but not the frequency hypothesis). *Behavioral and Brain Sciences, 30*, 278-280.

Roberts, M. J., & Newton, E. J. (2002). Inspection times, the change task, and the rapid response selection task. *Quarterly Journal of Experimental Psychology, 54A*, 1031-1048.

Rottenstreich, Y., & Kivetz, R. (2006). On decision making without likelihood judgment. *Organizational Behavior and Human Decision Processes, 101*, 74-88.

Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development, 72*, 103–122.

Schlottmann, A., & Anderson, N. H. (1994). Children's judgements of expected value. *Developmental Psychology, 30*, 56-66.

Schlottmann, A., & Christoforou, M. (2005). *Why are young children so good at expected value judgment?* Biennial Meetings of the Society for Research in Child Development. Atlanta, Georgia, USA.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127*, 966-972.

Seligman, M. E. P., & Kahana, M. (2009). Unpacking intuition: a conjecture. Perspectives on *Psychological Science, 4*, 399–402.

Shallice, T., & Burgess, P. W. (1993). Supervisory control of thought and action. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp. 171 – 187). Oxford: Oxford University Press.

Sheridan, S. L., & Pignone, M. (2002). Numeracy and the medical students' ability to interpret data. *Effective Clinical Practice, 5*, 35–40.

Shiffrin, R. M., & Schneider, W. (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–90.

Shiloh, S., Salton, E., & Sharabi, D. (2002). Individual differences in rational and intuitive thinking styles as predictors of heuristic responses and framing effects. *Personality and Individual Differences, 32*, 415-429.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Society for Research in Child Development Monographs, 46*.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S., & Chen, Z. (2002). Development of rules and strategies: balancing the old and the new. *Journal of Experimental Child Psychology*, *81*, 446-457.

Singer, J. A., Kohn, A. S., & Resnick, L. B. (1997). Knowing about proportions in different contexts. In: T. Nunes, & P. Bryant. (Eds.). *Learning and teaching mathematics*. East Sussex, UK: Psychology Press, 115-132.

Sladek, R. M., Bond, M. J., & Phillips, P. A. (2010). Age and gender differences in preferences for rational and experiential thinking. *Personality and Individual Differences*, *49*, 907-911.

Sloman S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22.

Smith, E., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality & Social Psychology Review*, 4, 108-131.

Smith, S. M., & Levin, I. P. (1996). Need for cognition and choice framing effects. *Journal of Behavioral Decision Making*, *9*, 283–290.

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. (1999). *Who is rational? Studies of individual differences in reasoning*. Hillsdale, NJ: Erlbaum.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago, IL: University of Chicago Press.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*, 342-357.

Stanovich, K. E., & West, R. F. (1998a). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*, 161-188.

Stanovich, K. E., & West, R. F. (1998b). Individual differences in framing and conjunction effects. *Thinking and Reasoning*, *4*, 289–317.

Stanovich, K. E., & West, R. F. (1999) Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology 38*, 349–85.

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. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*, 225-247.

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*, 251-285.

Stavy, R., & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: Intuitive rules.* New York: Teachers College Press.

Stavy, R., Goel, V., Critchley, H., & Dolan, R. (2006). Intuitive interference in quantitative reasoning. *Brain Research, 1073-1074*, 383-388.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, *28*, 78–106.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personal & Social Psychology Review*, *8*, 220-247.

Subbotsky, E. V. (1990). Phenomenal and rational perception of some object relations by preschoolers. *Soviet Psychology*, *28*, 5–24.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decision. *Psychological Science in the Public Interest*, *1*, 1-26.

Tirosh, D., & Stavy, R. (1999). Intuitive rules: a way to explain and predict students' reasoning. *Educational Studies in Mathematics, 38*, 51-66.

Toates, F. (2006). A model of the hierarchy of behaviour, cognition and consciousness. *Conscious Cognition*, 15, 75-118.

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. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

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

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

Tversky, A., & Kahneman, D. (1983). Extension versus intuititve reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293-315.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*, 252-278.

Wagenmakers, E.J., & Waldorp, L. (Eds.) (2006). Model selection: theoretical developments and applications [Special section]. *Journal of Mathematical Psychology*, *50*, 99-100.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), New horizons in psychology (pp. 135–151). Harmonsworth, England: Penguin.

Wason, P. C., & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition*, *3*, 141–154.

Weber, E. U., & Johnson, E. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*, 53-85.

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.

Wheeler, P., & Hyland, M. E. (2005). Cognitive style predicts use of CAM and attitude to CAM. *Focus on Alternative and Complementary Therapies*, *10* (Suppl. 1), 56.

Wilkening, E, & Anderson, N. H. (1991). Representation and diagnosis of knowledge structures in developmental psychology. In N. H. Anderson (Ed.), *Contributions to information integration theory*. Vol. III: Developmental (pp. 46-80). Hillsdale, NJ: Erlbaum.

Wilkinson, A. C. (1982). Theoretical and methodological analysis of partial knowledge. *Developmental Review*, *2*, 274-304.

Wilson, T. D. (2002). *Strangers to ourselves*. Cambridge, Mass.: Belknap Press.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, *60*, 181–192.

Winer, G. A., Craig, R. K., & Weinbaum, E. (1992) Adults failure on misleading weight-conservation tests – A developmental analysis. *Developmental Psychology*, *28*, 109–20.

Wolfe, C. R., & Reyna, V. F. (2009). Semantic coherence and fallacies in estimating joint probabilities. *Journal of Behavioral Decision Making*, *23*, 203-223.

Yamigishi, K. (1997). When a 12.86% mortality is more dangerous than 24.14%: Implications for risk communication. *Applied Cognitive Psychology*, *11*, 495–506.

Yee, T. W., & Mackenzie, M. (2002). Vector generalized additive models in plant ecology. *Ecological Modeling*, *157*, 141-146.

Yost, P. A., Siegel, A. E., & Andrews, J. M. (1962). Nonverbal probability judgments by young children. *Child Development*, *33*, 769-780.

Zikmund-Fisher, B. J., Smith, D. M., Ubel, P. A., & Fagerlin, A. (2007). Validation of the Subjective Numeracy Scale (SNS): effects of low numeracy on comprehension of risk communications and utility elicitations. *Medical Decision Making*, *27*, 663-671.

Appendix A

ITALIAN VERSION: SEVENTH GRADERS AND MIDDLE ADOLESCENTS

PROBLEMA MATEMATICO:

Ora ti saranno presentati 3 problemi.

Avrai a disposizione un tempo massimo per leggere e capire il problema. Poi ti apparirà automaticamente una pagina in cui dovrai rispondere a delle domande. Rispondi con serietà e per conto tuo. Ognuno di voi ha una versione diversa.

Clicca "AVANTI" per proseguire.

Hai a disposizione 60 secondi (40 secondi per gli adolescenti) per leggere il testo. Dopo 60 secondi (40 secondi per gli adolescenti) passerai automaticamente alla pagina successiva.

Due contenitori, che chiameremo Contenitore A e Contenitore B, sono riempiti da biglie di due colori: rosso e blu.

Il contenitore A contiene 10 biglie: 9 biglie sono rosse e 1 biglia e' blu.

Il contenitore B, invece, contiene 100 [95, 105] biglie: 90 [85, 95] biglie sono rosse e 10 biglie sono blu.

Devi estrarre una biglia (ovviamente senza guardare) da uno dei due contenitori.
Se peschi una biglia rossa vinci, altrimenti non succede nulla e il gioco finisce.

Ti chiediamo ora di dare lo stesso una risposta, anche se non ti senti sicuro e ti stai basando solo sulle tue sensazioni.

Quale contenitore ti offre maggiori probabilità di vincita? Scegli una delle tre alternative.

a. Contenitore A (9:10)
b. Contenitore B (90:100, 85:95, 95:195)
c. Non fa nessuna differenza: le probabilità di vincere sono le stesse

PROBLEMA DELLA PALLAVOLO:

Ora ti saranno presentati 3 problemi.

Avrai a disposizione un tempo massimo per leggere e capire il problema. Poi ti apparirà automaticamente una pagina in cui dovrai rispondere a delle domande. Rispondi con serietà e per conto tuo. Ognuno di voi ha una versione diversa.

Clicca "AVANTI" per proseguire.

Hai a disposizione 80 (54 secondi per gli adolescenti) secondi per leggere il testo. Dopo 80 (54 secondi per gli adolescenti) secondi passerai automaticamente alla pagina successiva.

Diverse squadre di pallavolo stanno selezionando nuovi atleti da mettere in squadra e tu vieni invitato a partecipare alla prima fase delle selezioni.
Tra le varie squadre, ce ne sono due che ti piacciono particolarmente ma che comunque sono a pari livello nelle classifiche. Con le tue capacità potresti essere adatto ad entrare in ciascuna delle due, ma sei costretto a esprimere la tua preferenza soltanto per una.

Nella Squadra A ci sono 10 atleti (di cui uno saresti tu) che vogliono entrare e 9 di loro supereranno la prima fase delle selezioni.

Nella Squadra B ci sono 100 atleti (di cui uno saresti tu) che vogliono entrare e 90 di loro supereranno la prima fase delle selezioni.

Ti chiediamo ora di dare lo stesso una risposta, anche se non ti senti sicuro e ti stai basando solo sulle tue sensazioni.

Quale squadra ti offre maggiori possibilità di superare la prima fase delle selezioni? Scegli una delle tre alternative.

a. Squadra A (9:10)
b. Squadra B (90:100, 85:95, 95:195)
c. Non fa nessuna differenza: le probabilità di passare la prima fase delle selezioni sono le stesse

PROBLEMA DELLA VACCINAZIONE:

Ora ti saranno presentati 3 problemi.

Avrai a disposizione un tempo massimo per leggere e capire il problema. Poi ti apparirà automaticamente una pagina in cui dovrai rispondere a delle domande. Rispondi con serietà e per conto tuo. Ognuno di voi ha una versione diversa.

Clicca "AVANTI" per proseguire.

Hai a disposizione 68 secondi (42 secondi per gli adolescenti) per leggere il testo. Dopo 68 secondi (42 secondi per gli adolescenti) passerai automaticamente alla pagina successiva.

Vieni chiamato perché devi sottoporti a una vaccinazione. Ci sono due ambulatori di pari competenza dove puoi andare a farti fare il vaccino e devi sceglierne uno.

Ti viene detto che nell'ambulatorio A sono in programma 10 vaccinazioni a settimana (una delle quali sarebbe la tua), e 9 avranno esito positivo, cioè saranno efficaci.

Nell'ambulatorio B, invece, sono in programma 100 vaccinazioni a settimana (una delle quali sarebbe la tua), e 90 avranno esito positivo, cioè saranno efficaci.

Il tempo di attesa per farsi il vaccino è lo stesso in entrambi gli ambulatori.

Ti chiediamo ora di dare lo stesso una risposta, anche se non ti senti sicuro e ti stai basando solo sulle tue sensazioni.

Quale squadra ti offre maggiori possibilità che il vaccino sia efficace? Scegli una delle tre alternative.

a. Ambulatorio A (9:10)
b. Ambulatorio B (90:100, 85:95, 95:195)
c. Non fa nessuna differenza: le probabilità che il vaccino sia efficace sono le stesse

ENGLISH VERSION: CORNELL UNDERGRADUATES

MATHEMATICAL PROBLEM

Now you have 40 seconds to read carefully the problem below. After 40 seconds you will be automatically moved to the next page.

Two containers, labeled A and B, are filled with red and blue marbles in the following quantities. Container A contains 10 marbles, 9 blue and 1 black. Container B contains 100 [95, 105] marbles, 90 [85, 95] blue and 10 black. You must draw a marble (without looking, of course) after choosing one of the containers. If you draw a blue marble, you win, otherwise you win nothing and the game is over.

Which container gives you a better chance of winning? Choose one of the three answers.

a. Container A (9:10)

b. Container B (90:100/85:95/95:105)

c. It would not matter to me: chances are the same

VOLLEYBALL PROBLEM

Now you have 40 seconds to read carefully the problem below. After 40 seconds you will be automatically moved to the next page.

A Selection Committee invited you to participate in a first-round volleyball recruitment. You like two teams of equal skill. Your volleyball skills are equally good for each team, but you can only apply for one. For Team A there are 10 applicants (one of them would be you), and 9 of whom will be selected. For Team B, there are 100 [95, 105] applicants (one of them would be you) and 90 [85, 95] of whom will be selected.

Which team gives you a better chance of being selected? Choose one of the four answers.

a. Team A (9:10)

b. Team B (90:100/85:95/95:105)

c. It would not matter to me: chances are the same

VACCINATION PROBLEM

Now you have 50 seconds to read carefully the problem below. After 50 seconds you will be automatically moved to the next page.

You are called because you must get a vaccine. There are two clinics of equal competence where you can go to get vaccinated and you must select one of them. You are told that there are 10 people vaccinated per week (including you) at Clinic A, and 9 vaccines are expected to be effective. At Clinic B, instead, there are 100 [95, 105] people vaccinated per week (including you), and 90 [85, 95] vaccines are expected to be effective. The waiting time to get the vaccine is the same at both clinics.

Which clinic gives you a better chance having a positive vaccine? Choose one of the three answers.

a. Clinic A (9:10)

b. Clinic B (90:100/85:95/95:105)

c. It would not matter to me: chances are the same

Appendix B

ITALIAN VERSION

Objective numeracy scale (Lipkus, Samsa & Rimer, 2001)

General numeracy items

1. Immagina di lanciare 1000 volte un dado non truccato a sei facce. Su questi 1000 lanci, quante volte pensi che uscirà un numero pari (cioè 2, 4, o 6)? ______________
2. In una lotteria la probabilità di vincere 10 Euro è l'1%. Se 1000 persone comprano ciascuna un solo biglietto di questa lotteria, quante di queste 1000 persone vinceranno il premio da 10 Euro?___________
3. Immagina che in un concorso a premi la probabilità di vincere un'automobile sia 1 su 1000. Quale sarà la percentuale di biglietti vincenti?________________

Risky numeracy items

4. Quale delle seguenti proporzioni rappresenta il rischio maggiore di contrarre una malattia? _____1 su100, _____1 su 1000, _____1 su 10.
5. Quale delle seguenti percentuali rappresenta il rischio maggiore di contrarre una malattia? 1%, 10%, 5%?
6. Se la probabilità di contrarre una malattia è il 10%, quante persone su 100 ci si aspetta che la contraggano? ________________
7. Se la probabilità di contrarre una malattia è il 10%, quante persone su 1000 ci si aspetta che la contraggano? ___________________
8. Se ci sono 20 possibilità su 100 di contrarre una malattia, questo significa che la probabilità di contrarre la malattia è pari al _________ %.
9. Se per una persona X il rischio di contrarre una malattia è pari all'1% in dieci anni e per una persona Y il rischio di contrarre la medesima malattia è il doppio della persona X, quale è il rischio che corre la persona Y?__________ %.
10. Se per una persona X le possibilità di contrarre una malattia sono 1 su 100 in dieci anni e per una persona Y il rischio di contrarre la medesima malattia è il doppio della persona X, quale è il rischio che corre la persona Y? _______________
11. La probabilità di contrarre un'infezione virale è pari a .0005. Su un campione di 10.000 persone, quante di loro ti aspetti che contrarranno l'infezione? ____________________

Subjective numeracy scale (Fagerlin, Zikmund-Fisher, Ubel, Jankovic & Smith, 2007)

Cognitive abilities

Per ciascuna delle seguenti domande metti una crocetta sul valore che meglio rappresenta quanto sei capace di fare le seguenti cose:

1. Quanto sei bravo con le frazioni? (1 = *per niente*, 5 = *molto bravo*)
2. Quanto sei bravo con le percentuali? (1 = *per niente*, 5 = *molto bravo*)
3. Quanto sei bravo a calcolare uno sconto del 15% (1 = *per niente*, 5 = *molto bravo*)
4. Quanto sei bravo a capire quanto costa una maglietta scontata del 25%? (1 = *per niente*, 5 = *molto bravo*)

Preference for display of numerical information

Per ciascuna delle seguenti domande metti una crocetta sul valore che meglio ti rappresenta.

5. Quando leggi, quanto ritieni utile che figure e tabelle siano riportate come informazione aggiunta al testo? (1 = *per niente*, 5 = *totalmente*)
6. Quando le persone ti parlano della probabilità che accada qualcosa, preferisci che usino *parole* (come "accade raramente") o *numeri* (come "c'è l'1% di probabilità che accada")? (1 = *totalmente parole*, 5 = *totalmente numeri*)
7. Quando ascolti le previsioni meteorologiche, preferisci che le previsioni ti vengano date in percentuale (ad esempio "domani la probabilità che piova è del 20%") oppure solo a parole (ad esempio "la probabilità che piova domani è bassa")? (1 = *totalmente in percentuale*, 5 = *totalmente in parole*; reversed item)
8. Quanto spesso ti capita di pensare che un'informazione numerica sia utile? (1 = *mai*, 5 = *sempre*)

Rational Experiential Inventory (Pacini & Epstein, 1999)

Rationality scale (Need for Cognition)

Istruzioni: Per ciascuna delle seguenti domande, Scegli uno dei valori tra 1 e 5 dove 1 rappresenta la risposta "*non mi descrive affatto*" fino a 5 che, invece, rappresenta la risposta "*mi descrive completamente*".

1 = non mi descrive affatto
2 = mi descrive poco
3 = non so
4 = mi descrive abbastanza
5 = mi descrive completamente

1. Tendo a evitare le situazioni in cui devo ragionare molto su qualcosa. (re-)[1]
2. Non sono un gran che a risolvere problemi complicati. (ra-)
3. Le sfide intellettuali mi stimolano e mi divertono. (re)
4. Non sono molto bravo a risolvere problemi che richiedono una analisi logica e attenta (ra)
5. Non mi piace ragionare troppo (re-)
6. Mi piacciono i problemi che richiedono molto ragionamento per essere risolti. (re)
7. Ragionare non corrisponde alla mia idea di divertimento. (re-)
8. Non ho un modo di ragionare molto analitico. (ra-)
9. Ragionare sulle cose con attenzione non è uno dei miei punti di forza. (ra-)
10. Preferisco i problemi semplici a quelli complessi. (re)
11. Provo poca soddisfazione nel ragionare intensamente su qualcosa per molto tempo. (re-)
12. Non riesco a ragionare se mi sento sotto pressione. (ra-)
13. Sono molto più capace delle altre persone a usare la logica per capire le cose. (ra)
14. Ho una mente logica. (ra)
15. Mi piace ragionare in astratto. (re)
16. Non ho nessun problema del ragionare sulle cose in modo attento (ra)
17. Usare la logica funziona molto bene quando devo risolvere problemi nella mia vita. (ra)
18. Sapere la risposta senza dover per forza capire il ragionamento sottostante è più che sufficiente per me. (re-)
19. Di solito le mie decisioni si basano su ragioni chiare e motivate (ra)
20. Trovo che sia stimolante imparare nuovi modi di ragionare. (re)

[1] *re* corrisponde a *rational engagement*, *ra* corrisponde a *rational ability*; *re-* and *ra-* indicano gli item reversed

Experientiality scale (Faith in Intuition)

Istruzioni: Per ciascuna delle seguenti domande, Scegli uno dei valori tra 1 e 5 dove 1 rappresenta la risposta "*non mi descrive affatto*" fino a 5 che, invece, rappresenta la risposta "*mi descrive completamente*".
1 = non mi descrive affatto
2 = mi descrive poco
3 = non so
4 = mi descrive abbastanza
5 = mi descrive completamente

1. Amo contare sul mio intuito. (ee)[2]
2. Non ho un buon intuito. (ea-)
3. Quando devo risolvere problematiche nella mia vita solitamente funziona affidarmi alle mie sensazioni. (ea)
4. Ho fiducia nelle mie sensazioni. (ea)
5. L'intuito può essere uno strumento molto utile per risolvere problemi. (ee)
6. Spesso faccio affidamento all'istinto quando devo decidere di intraprendere qualcosa. (ee)
7. Credo nelle mie prime impressioni sulle persone. (ea)
8. Solitamente mi affido alle mie sensazioni per dare o non dare fiducia alle persone. (ea)
9. Se mi fossi basato sulle mie sensazioni avrei commesso errori frequentemente. (ea-)
10. Non mi piacciono le situazioni in cui devo affidarmi all'intuito. (ee-)
11. Credo che ci siano circostanze in cui una persona dovrebbe affidarsi al proprio intuito. (ee)
12. Penso che sia sciocco prendere decisioni importanti sulla base delle proprie sensazioni. (ee-)
13. Non credo sia una buona idea affidarsi all'intuito nel momento in cui si devono prendere decisioni importanti. (ee-)
14. Generalmente le mie sensazioni non mi aiutano a prendere decisioni. (ee-)
15. Molto difficilmente sbaglio nel trovare una risposta quando ascolto le mie sensazioni.(ea)
16. L'ultima cosa che vorrei sarebbe dipendere da qualcuno che si descrive come una persona che nella vita si basa sull'intuito. (ee-)
17. Quando esprimo un giudizio di getto, probabilmente non sono bravo come la maggior parte delle altre persone. (ea-)
18. Tendo a usare il cuore come guida per le mie azioni. (ee)
19. Sono capace di percepire se una persona è buona o cattiva anche se non sono in grado di spiegare come lo so. (ea)
20. Credo che le mie impressioni siano spesso precise tanto quanto imprecise. (ea-)

[2] *ee* corrisponde a experiential engagement, *ea* corrisponde a experiential ability; *ee-* and *ea-* indicano gli item reversed

Cognitive Reflection task (Frederick, 2005)

Rispondi alle seguenti domande.

1. Supponi che una mazza e una palla costino 1.10 Euro in tutto. La mazza costa 1 Euro in più della palla. Quanto costa la palla? ________________

2. Se 5 macchinari impiegano 5 minuti per produrre 5 oggetti, di quanti minuti hanno bisogno 100 macchinari per produrre 100 oggetti? ______________

3. In un lago si è formata un'area di foglie di ninfee. Ogni giorno quest'area diventa due volte più grande. Se quest'area costituita da foglie impiega 48 giorni per espandersi fino coprire l'intera superficie del lago, quanti giorni sono necessari per coprire metà della superficie del lago? ___________________

ENGLISH VERSION: CORNELL UNDERGRADUATES

Objective numeracy scale (Lipkus, Samsa & Rimer, 2001)

General numeracy items

1. Imagine that we role a fair, six-sided die 1,000 times. Out of 1,000 roles, how many times do you think the die would come up even (2, 4, or 6)?_______
2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each by a single ticket to BIG BUCKS?________________
3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car? ____________

Risky numeracy items

4. Which of the following numbers represents the biggest risk of getting a disease? _____1 in 100, _____1 in 1000, _____1 in 10.
5. Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%? _________
6. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100? ___________
7. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000? ____________
8. If the chance of getting a disease is 20 out of 100, this would be the same of having a _______% chance of getting the disease.
9. If person A's risk of getting a disease is 1% in ten years, and person B's risk is double than of A's, what is B's risk? _____________
10. If person A's chance of getting a disease is 1 in 100 in ten years, and person B's risk is double than of A's, what is B's risk? ____________
11. The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected? _____________

Subjective numeracy scale (Fagerlin, Zikmund-Fisher, Ubel, Jankovic & Smith, 2007)

Cognitive abilities

For each of the following questions, please check the box that best reflects how good you are at doing the following things:

1. How good are you at working with fractions?
2. How good are you at working with percentages?
3. How good are you at calculating a 15% tip?
4. How good are you at figuring out how much a shirt will cost if it is 25% off?

These responses are rated from 1 (*not at all good*) to 5 (*extremely good*)

Preference for display of numerical information

For each of the following questions, please check the box that best reflects your answer:

5. When reading the newspaper, how helpful do you find tables and graphs that are parts of a story? (1 = *not at all*, 5 = *extremely*)
6. When people tell you the chance of something happening, do you prefer that they use *words* ("it rarely happens") or *numbers* ("there's a 1% chance")? (1 = *always prefer words*, 5 = *always prefer numbers*)
7. When you hear a weather forecast, do you prefer predictions using percentages (e.g., “there will be a 20% chance of rain today”) or predictions using only words (e.g., “there is a small chance of rain today”)? (1 = *always prefer percentages*, 5 = *always prefer words*, reverse coded)
8. How often do you find numerical information to be useful? (1 = *never*, 5 = *always*)

Rational Experiential Inventory (Pacini & Epstein, 1999)

Rationality scale (Need for Cognition)

1. I try to avoid situations that require thinking in depth about something. (re-)[3]
2. I'm not that good at figuring out complicated problems. (ra-)
3. I enjoy intellectual challenges. (re)
4. I am not very good at solving problems that require careful logical analysis. (ra)
5. I don't like to do a lot of thinking. (re-)
6. I enjoy solving problems that require hard thinking. (re)
7. Thinking is not my idea of an enjoyable activity. (re-)
8. I am not a very analytical thinker. (ra-)
9. Reasoning things out carefully is not one of my strong points. (ra-)
10. I prefer complex problems to simple problems. (re)
11. Thinking hard and for a long time about something gives me little satisfaction.(re-)
12. I don't reason under pressure. (ra-)
13. I am much better at figuring things out logically than most people. (ra)
14. I have a logical mind. (ra)
15. I enjoy thinking in abstract terms. (re)
16. I have no problem thinking things through carefully. (ra)
17. Using logic usually works well for me figuring out problems in my life. (ra)
18. Knowing the answer without having to understand the reasoning behind it is good enough for me. (re-)
19. I usually have clear, explainable reasons for my decisions. (ra)
20. Learning new ways to think would be very appealing to me. (re)

These responses are rated from 1 (*definitely not true of myself*) to 5 (*definitely true of myself*)

[3] *re* means rational engagement, *ra* means rational ability; *re-* and *ra-* are the reversed items

Experientiality scale (Faith in Intuition)

1. I like to rely on my intuitive impressions. (ee)[4]
2. I don't have a very good sense of intuition.(ea-)
3. Using my gut feelings usually works well for me in figuring out problems in my life. (ea)
4. I believe in trusting my hunches. (ea)
5. Intuition can be a very useful way to solve problems. (ee)
6. I often go by my instincts when deciding on a course of action. (ee)
7. I trust my initial feelings about people. (ea)
8. When it comes to trusting people, I can usually rely on my gut feelings. (ea)
9. If I were to rely on my gut feelings, I would often make mistakes. (ea-)
10. I don't like situations in which I have to rely on intuition. (ee-)
11. I think there are times when one should rely on one's intuition. (ee)
12. I think it is foolish to make important decisions based on feelings. (ee-)
13. I don't think it is a good idea to rely on one's intuition for important decisions. (ee-)
14. I generally don't depend on my feelings to help me make decisions. (ee-)
15. I hardly ever go wrong when I listen to my deepest gut feelings to find an answer.(ea)
16. I would not want to depend on anyone who described himself or herself as intuitive. (ee-)
17. My snap judgments are probably not as good as most people's. (ea-)
18. I tend to use my heart as a guide for my actions. (ee)
19. I can usually feel when a person is right or wrong, even if I can't explain how I know. (ea)
20. I suspect my hunches are inaccurate as often they are accurate. (ea-)

These responses are rated from 1 (*definitely not true of myself*) to 5 (*definitely true of myself*)

[4] *ee* means experiential engagement, *ea* means experiential ability; *ee*- and *ea*- are the reversed items

Cognitive Reflection task (Frederick, 2005)

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? ___cents
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ________ minutes
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ______ days