University of Trento
Doctoral School in Cognitive Sciences and Education
XXII° cycle

PhD Thesis

When 1 in 200 is higher than 5 in 1000:
The “1 in X effect” on the perceived probability of having a Down syndrome-affected child

Advisor:
Lucia Savadori

PhD candidate:
Elisa Barilli

A.Y. 2009-2010

14 December 2010

As the old joke goes, five out of four people have trouble with fractions

Ischebeck, Schocke, and Delazer (2009, p. 403)

Part of this work was supported by a grant from the European Commission (**Special Advances in Foetal and Neonatal Evaluation Network Project N8:LSHB-CT-2004-503243**).

Contents

CHAPTER 1. Field of investigation and Research problem 1

The importance of statistical information in doctor-patient communication............. 2

The subjective character of "objective" numbers: Biases in judgment and decision making 4

Communicating numerical information in health-care: The importance of studying format effects... 6

Object of investigation: A specific format effect in risk communication of prenatal genetic testing results 7

The specific case of Down syndrome 8

CHAPTER 2. Existent literature on the issue... 11

Risk as analysis and Risk as feelings... 13

What is known on comprehension of different ratio formats.............................. 15

SECTION 1. When 10 in 100 is higher than 1 in 10:

Results in favour of neglect of the denominator... 19

1.1 Explanations and experimental evidences under main theories...................... 19

1.1.1 The different number of counterfactual alternatives prompted 19

1.1.2 Experiential over rational system prevalence: Ratio-bias literature 21

1.1.3 Denominator neglect as an inclusion illusion: Fuzzy-trace theory................ 27

1.2 Other interpretations 36

1.2.1 Pseudo-multiplicity: Perceived Diversity Heuristics 36

1.2.2 Exemplar availability: Exemplar-cuing theory.. 38

1.2.3 Visualization of the numerator 40

1.2.3.1 Imaginability/images of the numerator and affect 41

1.2.3.2 Imaginability/images of the numerator, but not affectively tagged 43

1.3 Elicitation of health-state utilities, of numerical frequencies of social facts, and of the probability of harm posed by a mental patient if discharged are influenced by Ratio-bias 45

SECTION 2. When 1 in 10 is higher than 10 in 100: Results in favour of neglect of the numerator 49

The group-diffusion effect and the reference group effect 49

From group diffusion to Ratio-bias 52

CHAPTER 3. Empirical studies 55

Rationale for research studies 56

Methodological considerations 57

Preliminary brief characterization of the basic constructs investigated 57

SECTION 1, or A new bizarre systematic effect 66

Study 1.1 66

Study 1.2 75

SECTION 2, or Boundaries of the bias 81

Study 2.1 83

Study 2.2 102

Study 2.3 111

Study 2.4 116

Study 2.5 118

Study 2.6 122

The effect of individual differences on the bias under study 129

SECTION 3, or An affective explanation? 130

Study 3.1.. 131

Study 3.2.. 134

SECTION 4, or Test of two practical interventions 140

Study 4.1.. 141

Study 4.2.. 146

Study 4.3.. 150

CHAPTER 4. Conclusions and General discussion 155

BIBLIOGRAPHIC REFERENCES 167

APPENDIX. Experimental material employed in the studies.................... 185

CHAPTER 1

Field of investigation and Research problem

The importance of statistical information in doctor-patient communication

Increasing emphasis has been placed on the way doctors communicate risk to patients (Alaszewski and Horlick-Jones, 2003; Calman, Bennett, and Coles, 1999; Gigerenzer and Edwards, 2003; Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, and Woloshin, 2007; Paling, 2003). Indeed, clinicians are increasingly recommended- even mandated by law- to help patients make informed-medical decisions by paying more attention to risk counselling (Weinstein, 1999; Schwartz, Woloshin, and Welch, 1999). Since the Seventies, the paternalistic doctor-centred model of the physician-patient communication (in which the physician was chargeable with the decision of which information was relevant, and had the sole decision making responsibility) has been progressively abandoned for a more patient-oriented approach, until a deliberative model (shared-decision approach) has become customary in most of the advanced countries (Emanuel and Emanuel, 1992). In the latter model, the physician's role resembles that of a friend or teacher, whose main aim is "to help the patient determine and choose the best health-related values that can be realized in the clinical consultation. To this end, the physician must delineate information on the patient's clinical situation and then help elucidate the types of values embodied in the available options"[1] (p. 2222). In this new relationship-centred approach, patients as well are asked an active role both in the definition of the health problems they face, and in the evaluation of possible solutions.

Following a progressive acquisition of a prerogative state, the patient has become the main entitled holder of her/his wellbeing, and as such s/he ought to be provided with all necessary means to actuate a conscious decision in matters of health. For these reasons, doctors and technical health staff have the duty of ensuring that every choice is active and conscious, that informed consent is reached on all medical risks and on the selected practices to contrast these risks (Santuososso, 1996). As Hall et al. (2007, p. 564) have maintained, "This is particularly evident in the context of genetic counseling, with a summary of 51 national and international guidelines for genetic counseling emphasizing the importance of patient autonomy and non-directive information giving in this context (http://www.eurogentest.org/web/info/public/unit3/guidelineswp12.xhtm, accessed on 02.10.2006)". Moreover, this right of acquiring information has been not only identified as important to patients (e.g., people with cancer disease), but also as one of their primary unmet needs, especially concerning that of quantitative information like probabilities of risky events (Feldman-Stewart, Kocovski, McConnell, Brundage, and Mackillop, 2000). In order to take a decision concerning their health, people might need to know the frequency of occurrence of a given outcome in their population of reference (for example, how many people out of the total

[1] Not far from this idea is the concept of "libertarian paternalism" advocated by Sunstein and Thaler (2003) for private and public institutions that should affect behavior in welfare-promoting directions while also respecting freedom of choice.

of those going vacating in a tropical country each year did contract malaria despite having been vaccinated against it). In agreement with this unmet need, for instance, Feldman-Stewart et al. (2000, p. 228) reported from one of their studies that "*four of the top five items (of 59)* identified by men with early-stage prostate cancer as necessary for their treatment decisions concerned the chances of a particular event 'happening'".

The inclusion of statistical facts on the condition of interest or on the treatments available to cure it, in a risk message does not only represent an inescapable step of a medical communication carried out in accordance with the guidelines in force, but the important value of numbers in communications of statuses and actions entailing uncertainty has been also recognised by health communication and cognitive psychology scholars. It seems, that mentioning the likelihoods in a numerical format can increase trust and belief in and comfort with the risk information (Gurmankin, Baron, and Armstrong, 2004). Some studies indeed found that statistical information increases comprehension (Marteau, Saidi, Goodburn, Lawton, Michie, and Bobrow, cited in Visschers, Meertens, Passchier, and de Vries, 2009). However, it has been widely questioned whether patients really understand and ultimately use these values (see the literature on people's general innumeracy, e.g. Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, and Woloshin, 2007, or Lipkus, Samsa, and Rimer, 2001).

When they are explicitly asked, patients say that they like verbal labels more than numbers. Verbal labels (e.g., "rarely", "sometimes", "often", and so on) are viewed as more easy to use and more natural than numerical information (e.g., Brun and Teigen, 1988; Wallsten, Budescu, Zwick, and Kemp, 1993). However, due to their imprecise nature, verbal labels cause more variability in risk perception than the other type of format (i.e., numbers), both between and within the respondents (Gurmankin et al., 2004). Indeed, the interpretation of qualitative expressions of risks varies greatly, with wide ranges in the meanings or numerical values attributed to verbal descriptions of risks across individuals (e.g., Ohnishi et al., 2002) and across contexts (e.g., Mazur, 1990; Mazur and Merz, 1994). That side-effect has led some authors (e.g., Burkell, 2004, but also Edwards, Elwin, and Mulley, 2002) to urge that verbal expressions of probability should be avoided, and to state, that risk information is better imparted with numerical expressions, provided that professionals help "[...] turning raw data into information that is more helpful to the discussion than the data" (Edwards, Elwin, and Mulley, 2002, p. 827).

The subjective character of "objective" numbers: Biases in judgement and decision making

Albeit the apparent objectivity of their character, raw numbers are not exempt from issues of multiplicity of interpretation. Several numerical formats can be chosen to express a given probability of an outcome- the most common of these being percentages (e.g., 5%), single-event probabilities (e.g., 0.05), frequencies[2] (e.g., 5 in 100), and absolute frequencies (e.g., 600); selecting one of those rather than another in risk communication is not without consequences. Research by Brase (2002) in a direct comparison of these expressions found that individuals felt them different in the degree of clarity and how easy they were to be understood; in particular, frequencies[3] ("*simple frequencies*", as he termed them) were in the first position, that is, the simplest and clearest among formats judged, followed by percentages (i.e., relative frequencies), and then by absolute frequencies based on very large reference classes, while single-event probabilities were in the last place (as they were perceived as the hardest to be understood).

Format effects trace back to the pioneering work summarised in Prospect Theory (Kahneman and Tversky, 1979), demonstrating the crucial role of the superficial format in which information is presented for its assessment, and for the decisions elaborated on its basis. Through all their works on experimental observations, Kahneman and Tversky proved systematic deviations from what theoretically expected according to the "format invariance principle"[4] argued by supporters of the classical economic theory. In other words, they showed that even those changes in the options, that from the point of view of the apparently "perfectly rational agent", could be considered irrelevant would instead move her/him away from the optimal decision, namely that theoretically predicted by the normative model. For example, as "framing effect"[5] illustrated (Kahneman and Tversky, 1982), simply altering the frame (in positive/negative terms) of a scenario can change preferences for one course of action rather than another, even if the options have quantitatively identical outcomes. In the studies on the

---

[2] Frequency can be defined as "the rate at which something happens, for instance the number of times it happens in a particular period of time" (Collins COBUILD English Language Dictionary, 1996), or "the number of times an event or character occurs in a given sample." (Oxford English Dictionary, 2006). In the case of the present thesis, the sample of the last definition mostly consists of a certain group of people.

[3] These findings are consistent with studies of statistical reasoning, which indicate that frequency presentations facilitate understanding of data (Cosmides and Tooby, 1996; Hoffrage and Gigerenzer, 1998) due to the suggestion that evolution and experience have been equipping people better to understand probabilistic information expressed as frequencies in a population, rather than as probabilities for an individual. In fact, Gigerenzer and Hoffrage (1995) argued that frequencies are the natural way in which people think about probabilities. However, results of studies investigating the same issue have not been consistent, hence existent research does not establish the superiority of frequency formats.

[4] Such property implies that every economical agent behaving in accordance with axioms of the Utility Theory (Von Neumann and Morgenstern, 1947) will not modify her/his preferences on the basis of the presentation format of a choice problem.

[5] Framing itself can be defined as "presenting 'logically equivalent' information in different ways" (Wilson, Purdon, and Wallston, quoted in Edwards and Elwyn, 2001, p. i11).

Asiatic Disease Problem (Tversky and Kahneman, 1981), two groups of participants were asked to role-play to be asked to find the best solution to face a strong epidemic, and having to choose between two urgency sanitary programs. One of them guaranteed a partial albeit sure result, while the other had a given likelihood of solving completely the problem, but could also be a complete failure. The problem was presented in two versions, namely in one the outcomes were described in terms of people saved, while in the other in terms of victims. Despite according to the invariance principle no differences had to be expected in participants' choices, results showed an inversion of preferences from one version to the other. Indeed, in the version expressed in terms of saved people, the majority of participants preferred the sure option, while on the contrary, in the version where the consequences of the program were described in terms of lost lives, the majority of participants expressed a preference for the risky option. Thus, authors' work showed that individuals did not chose on the basis of a rational assessment of options -i.e., in terms of their expected value, as predicted by Utility Theory (Von Neumann and Morgenstern, 1947), but were instead prone to the influence of minor details (like the framing of options) of the problem description.

The investigation of the "*satisficing*"[6] (Simon, 1979) rather than optimal behaviour, observed in the experiments of Kahneman and Tversky has been progressively extended to judgments under uncertainty of various events in diverse areas of everyday life. The research has been re-named the "heuristic and biases" program, "a territory that Herbert A. Simon had defined and named—the psychology of bounded rationality (Simon, 1955, 1979)" (Kahneman, 2003, p. 697). The portrait of the average individual resulting from such experimental analyses is that of an agent rarely following formal statistical rules in making decisions outside the laboratory. Rather, it is that of an individual whose evaluations, instead of being the result of formal and extensive algorithmic processing, are often based on a restricted number of simplifying heuristics. These "rules of thumb" are successful most of the times, in other words they guide to an effective solution of problems by helping individuals reducing the complexity of certain issues (see, for instance, Gigerenzer, 2008). The same rules, however, can sometimes induce in errors, named *biases* by scholars. Heuristic processing, therefore, is "positive", may reveal highly adaptive, especially when decisions are being taken in situations that are changing, uncertain, and dynamic. However, a wrong interpretation of this approach has been circulating for several years, where the use of heuristics has been identified as error-prone and leading to some systematic predispositions, namely the above mentioned biases.

Even objective numbers are subjected to heuristic evaluations. As such, while is true that rules of thumb applied to their evaluation might in some cases lead to the same result of

[6] Nobel laureate Herbert Simon coined the phrase '*satisficing*' to describe a decision making process that takes the shortcut of defining what is acceptable and then settling on the first alternative that meets those minimum requirements. It is a blended word combining *satisfy* with *suffice*, meaning that you sacrifice the best alternative for one that adequately satisfies (i.e., suffices) your demands at the current time.

formal calculations (but with advantages in terms of time and resources), it is inevitable that in other cases they lead to some distortions.

Similarly to other superficial features, the numerical presentation format used in probability communication has been advocated as a determinant characteristic capable of influencing people's judgments and decisions in situations of uncertainty. In other words, it has been recognized that even numerical presentation of a probability is subjected to "framing manipulations". Its effects have been intensively studied in disciplines like decision making psychology (for a review of the effects of numerical and other formats in probability communication, see Visschers, Meertens, Passchier, and de Vries, 2009)[7].

Communicating numerical information in health-care: The importance of studying format effects

In the applied field of health-care, the issue of which numerical format is best to express a probability has a decisive importance from the point of view of informed choices, and of that of bringing about reductions in risky health behaviour, for instance. The point then is defining the term "best". Three features must be considered at least: best understood (i.e., analytically meaningful), useful for judgment and decision making (i.e., affectively meaningful), best to reduce people's reckless health behaviours.

As stated above, communicating risk is not a choice for doctors but it is a mandatory act. Many are the areas of health care in which professionals need to provide risk information to patients to enhance their decisions; consider, for instances, some of the examples reported, among others, by Burkell (2004, p. 201): "Women making decisions about hormone replacement therapy to treat menopausal symptoms", as they "must understand and weigh the reduced risk of osteoporosis, cardiovascular disease, colorectal cancer, and Alzheimer's disease against the increased risk of breast cancer, myocardial infarction, cerebrovascular disease, and thromboembolic disease"; "Men choosing among options for the treatment of localized prostate cancer", as they "want to know the likelihood of side effects associated with the treatment options before making their decision"; or, "participants in genetic counseling programs", as they "must understand the risks associated with treatment and the meaning of a positive test result to make informed decisions about genetic testing". Not only, another case in which health professionals ought to communicate probability of a certain negative health-related outcome happening to laypeople is, for instance, that of the probability that a specific pattern of behaviour (e.g., smoking) will lead to a particular health problem (e.g., lung cancer).

[7] The issues analyzed included, to quote only the most discussed, a) the difference between numerical and verbal expression of probability; b) the frequency/ percentage debate; c) the presentation of absolute or relative risk information, d) the hard comprehension of cumulative probability information; e) the effects of employing a visual aid to represent the risk on its perception.

Moreover, nowadays the number of situations in which health-care practitioners (i.e., medical doctors, nurses or medical attendants, or technical assistants like, for example, radiologists) must communicate risks has grown extraordinarily. Unbelievable advances in medical sciences, in understanding the human genome, beside the identification of the genetic errors indictable for diseases, are on the one hand making the human being feel safer, but on the other hand are leading to an increase in number for the situations endowing a certain degree of risk that ought to be communicated. Such apparent oxymoron corresponds to the idea, that if we have gained a much better knowledge on the mechanisms regulating our body thus making us able to better pursue health welfare, on the other hand the number of conditions and possible diseases we should be tested for has expanded incredibly, making us somehow potentially "more at risk". Furthermore, an additional factor adding up to the number of risky situations is represented by the increasing use of biochemical, imaging, and genetic screening tests that all provide probabilistic information.

In all those situations, determining the effects that different numerical information related to risk and presented by health care professionals have on people's understanding of probability, their judgments (e.g., Ancker, Senathirajaha, Kukafka, and Starren, 2006; Cuite, Weinstein, Emmons, and Colditz, 2008; Feldman-Stewart, Brundage, Van Manen, and Svenson, 2004; Lipkus, 2007), and even behavioural intentions (e.g., Marteau, Kidd, Cook, Michie, Johnston, Slack, et al., 1991) is of crucial importance.

Object of investigation: A specific format effect in risk communication of prenatal genetic testing results

Among features deemed responsible of biases in perception, specific object of study of the present dissertation has been the use of different ratio[8] formats in risk communication of prenatal diagnosis results. In particular, work has focused on the possible influence that those expressions of the probability related to screening[9] for Down syndrome could have on prospective parents' perceptions. While for a brief general outline of the peculiarities regarding the context of decision (i.e., antenatal screening for Down syndrome) we refer to the paragraph here following, for what concerns the elucidation of what is meant with "the use of different

[8] From the strictly mathematical point of view, the concept of "ratio" is broader than that of "fraction", as while a fraction always illustrates a "part to whole" relationship (e.g., 3/4), ratios can be used to denote a much larger set of relationships, such as part to part (e.g., 3 : 1) and whole to part (e.g., 4 : 3). Despite the present object of interest would have been more appropriately described by the term "fraction", the term "ratio" was chosen to refer to it since most of the relevant literature on the phenomenon at study appeared to have used that expression, e.g., "Ratio-bias" (Denes-Raj and Epstein, 1994). Thus, by ratio formats we meant, verbalized frequency expressions like, for instance, "1 in 5", "3 out of 10", or "2 out of every 20".
[9] Broadly talking, "Screening is a systematic attempt to identify from apparently healthy individuals, those at high enough risk of a specified disease to warrant further action. Those in the high risk group are offered interventions which are either too expensive or hazardous to be provided without such prior selection." (Green, Hewison, Bekker, Bryant, and Cuckle, 2004, p. 1)

ratio formats" the next sentences will do. When one wants to express a given probability (e.g., 10%) through a ratio format, s/he can in principle state it by means of several equivalent ratios, whose numbers at the numerator and denominator will hence be (comparatively) both smaller or larger (i.e., N in N*X, or N_1 in N_1*X, or N_2 in N_2*X ...and so on). To exemplify, one can say, that a person has a *1 in 10* probability of having a child with Down syndrome, or, for instance, that such person has a *10 in 100* probability of having a child with Down syndrome. Does choosing one expression (i.e., 1 in 10, or N in N*X) rather than the other (i.e., 10 in 100, or N_1 in N_1*X, for instance) make any difference in terms of how those values are perceived by the receiver of the communication, in the specific, the person at risk? Answering such question was the main aim of the present work.

Research originated from both a theoretical and a practical question. From a theoretical point, studying format effects is one way to validate theories on how the mind processes information about risks. These theories will be described in Chapter 2. From a practical side, format effects, as said before, have concrete consequences on judgments and decisions of individuals regarding health issues, therefore their effect could be used to promote health care.

The specific case of Down syndrome

Antenatal screening offers the possibility of preventing the birth of infants with serious congenital abnormalities, such as neural tube defects, Down syndrome, chromosome abnormalities, genetic diseases and other conditions (such as spina bifida, cleft palate, Tay Sachs disease, sickle cell anaemia, thalassemia, cystic fibrosis, and fragile x syndrome). The purposes for prenatal diagnosis are not only, as commonly thought, to give the parents the chance to abort a foetus with the diagnosed disabling condition, but also, in case they want to carry the pregnancy to full term, to enable timely medical or surgical treatment of a risky condition before or after birth; moreover, to give prospective parents the chance to "prepare" psychologically, socially, financially, and medically for a baby with a health problem or disability, or for the likelihood of a stillbirth. Also, having this information in advance of the birth means that healthcare staff has the possibility to better prepare itself for the delivery of a child with a health problem.

Among the large range of chromosomal anomalies, Down syndrome is the most common, with an estimated frequency of 1 in 600 live births (Antonarakis, Petersen, McInnis, Adelsberger, Schinzel, Binkert, et al., 1992). This disease, which often implies impairments of the cognitive ability and physical growth problems for the baby, raises many questions for prospective parents such as how to cope with the waiting, how to cope with a special needs child, what about the baby's siblings and relatives, and so on. Antenatal screening for Down syndrome consists in the estimation of the woman's risk of having a Down syndrome pregnancy. Apart from nuchal translucency, it can be calculated for each woman based on her age and any combination of maternal serum markers: she will be described as screen positive if her risk value exceeds a specified cut-off value (normally, 1 in 250/270). If an elevated risk of chromosomal or genetic abnormality is indicated by the non-invasive screening test, a more invasive but diagnostic technique may be employed to gather more information (e.g., amniocentesis or Chorionic Villus Sampling).

Despite accepting or declining prenatal screening should be the result of an informed choice, namely one based on “relevant information” and “consistent with the decision-makers’ values” (van den Berg, Timmermans, Knol, van Eijk, de Smit, van Vugt, et al., 2008), it is often not so. For instance, in their study on a sample of 1159 pregnant women offered either maternal serum screening test or the nuchal translucency measurement, van den Berg and colleagues found that only 68% of the choices could be defined actually “informed”. Indeed, women interviewed frequently showed lack of one (or even both) of the dimensions common in all definitions of informed choices, that are sufficient knowledge (i.e., of, broadly speaking, “characteristics of condition for which screening is being offered, characteristics of the screening test, and implications of the possible test results”, ibid., p. 322), and value inconsistency (i.e., “disagreement between the abstract ideals guiding their behaviour and the actual behaviour itself”, ibid.). Counseling a couple about the Down syndrome screening result in a simple yet comprehensive way is a particular challenge for many health care providers. Indeed, a critical question is whether pregnant women and her partner have an adequate grasp of the numerical information delivered, and yet several studies have demonstrated women’s poor understanding of Down syndrome screening (Eiser, 1998; Godyer, Barratt, and Irwig, 2000; Thornton, Hewison, Lilford, and Vail, 1995).

Thus, in order to favour sound decisions, it is first of all necessary to understand how future parents form a subjective probability judgment out of raw numbers. Hence, studying effects of superficial presentations in communication on patients’ assessments and intentions to act is of vital importance. Analyzing the influence of the ratio formats expressing the probability of having a Down syndrome-affected child on prospective parents’ perception of the probability can disclose effects of determinant relevance for people’s life.

CHAPTER 2

Existent literature on the issue

The initial part of this chapter will briefly review the "risk as analysis and risk as feelings" theory on how human beings comprehend risk (Slovic, Finucane, Peters, and MacGregor, 2004) on which basis the investigation was grounded. Then, the thesis will start focusing on the specific issue of research, namely how the use of superficially different but mathematically equivalent ratio formats (e.g., N in N*X or N_1 in N_1*X) affects the magnitude perception of the probability which those ratios convey.

With this aim, first of all the existent research will be briefly summarised, which focused on the impact that the type of ratio format (1 at the numerators vs. other formats) used for probability presentation has on its comprehension, particularly when related to a medical domain.

Afterwards, research will be reviewed, that analysed the effect of ratio format on perceived probability and decision making. Indeed, starting in the 1990s, various empirical studies considered whether different ratio formats denoting the same objective probability -e.g., 1 in 10 vs. 10 in 100, or more generally "1 in X" vs. "N in (N*X)"- could impact choice and subjective evaluations. These studies delivered contrasting results.

On the one hand, mainly Prof. Seymour Epstein and colleagues suggested that people tend to neglect denominators, preferring for example a lottery offering a 10 in 100 chance of winning to another offering a 1 a 10 chance of winning. That phenomenon, dubbed the "*Ratio-bias*" effect (e.g., Denes-Raj and Epstein, 1994; Denes-Raj, Epstein, and Cole, 1995) under the Cognitive-Experiential Self-Theory (CEST, Epstein, 1991, 1993) and attributed to the prevalence of the experiential system over the rational one, has been instead ascribed to a "*denominator neglect*" under Fuzzy-trace theory (Reyna, 1991). The ability of such phenomenon to influence people's behaviour in different fields of application will be illustrated in Section 1 of the present chapter.

On the other hand, Yamaguchi (1998) suggested that when people are asked to evaluate a threat, whose probability is kept constant, they tend to rate the threat as less probable when the number of individuals at threat increases, for example if it will affect 10 persons in 100, rather than 1 person in 10. That phenomenon, dubbed the "*group-diffusion*" effect, seemed to suggest that people neglect numerators rather than denominators, in single ratio presentation. Group-diffusion effect will be illustrated in Section 2 of this chapter.

No connection between these two sets of theories postulating opposite tendencies had been established until Price and Matthews's work (2009). The latter will be briefly analysed in the conclusive part of the chapter.

Risk as analysis and Risk as feelings

Recently, a comprehensive theory on how human beings perceive and evaluate risk has been formally elaborated by Slovic, Finucane, Peters, and MacGregor (2004) in the wake of dual-process approaches of thinking, knowing, and information-processing in cognitive psychology (Chaiken and Trope, 1999; Kahneman and Friederick, 2002; Sloman, 1996), and modern theories of neuroscience (e. g., Damasio, 1994). A growing volume of researches and findings is increasingly being emphasizing the existence of two different essential ways in which individuals would comprehend risk, namely an "*experiential*" way- also addressed as "System 1" by Khaneman (2003) and scholars- and an "*analytic*" way- also addressed by some as the "rational" system (e.g., Epstein, 1991) or the "deliberative" system, and by others as "System 2", see Kahneman (2003) and scholars[10]. Both systems would be fundamental for risk processing and evaluation: while the analytic system uses normative rules and formal logic, as well as algorithms and probability calculus, the experiential system uses intuition, instincts, gut feelings, and emotions (for a review of properties of the two systems, and of the different interpretation of their facets, see section 1.1.2 and beginning of section 1.1.3 of the present chapter). Both systems would be continually active, and constantly interacting in what has been characterized as "the dance of affect and reason" (Finucane, Peters, and Slovic, 2003; see also, Slovic et al., 2004). Anyway, the prevalence (for several reasons, which will be tackled throughout the present work) of one way of reasoning over the other can influence the type of "answer" given by our mind to information and problems.

According to Slovic and colleagues, and similarly argued by Kahneman (2003), and Reyna (2004), not only the experiential system is the default system, necessary even to guide the analytic system to work properly (i.e., to "do the right thing"), but it also possess elements of rationality at same strength of the analytic system. In other words, the experiential system would be the basic fundamental way in which human beings encode and evaluate information.

This interpretation is, however, very recent. Indeed, after a period (i.e., with the neoclassic approach) in which only analytic thinking was considered legitimate object of research, exactly because of being the "epitome of rationality" (Slovic et al., 2004, p. 313), emotions[11] re-gained

[10] Otherwise differently stated, all the terms referring to one of the two systems will be used interchangeably to denote that system in the present thesis. When a specific theory will be described, instead, the terms used will be those employed by the author/s.

[11] A commonly accepted definition describes emotion as a vast disposition to answer that can include a linguistic measurable behavior, organized actions and a (somatic and visceral) physiological system of support for those events. Emotions intervene in evaluative and decisional processes, and can be broadly classified into immediate and anticipated emotions, the first ones further subdivided into integral (i.e. related to the object of evaluation/decision) and incidental (i.e. not related to it). All these types of emotions are supposed to play a role in the judgmental and choice behaviors, in other words 1) to cover an informative function (affect as information); 2) to be an instrument guiding attention ("a spotlight"); 3)

the role they should always have been entitled to, namely that of being one of the most important determinants of perception and behaviour. Previously, emotions were deemed as interfering with reason (hence somehow lowering the quality of the reasoning process).

In decision making, the revival of attention for the role of emotions approximately coincided with the conceptualization of "*the affect heuristic*", a mental shortcut consisting in reliance on the affective feelings[12] generated by a stimulus for evaluation and decision (Finucane, Alhakami, Slovic, and Johnson, 2000; Slovic, Finucane, Peters, and MacGregor, 2004). Authors incorporated the affect heuristic in their view of individuals' information-processing about risk (e.g., Slovic et al., 2004). The way in which affect heuristic works can be approximated as follows: in evaluating a situation, people would form mental images (that are influenced, sometimes even determined, by individual differences and by the type of task); such images have connotative emotions, namely conscious or unconscious tags associated with objects and events representations from the past, negativity or positivity of which the person would consult in order to gain an evaluation of the stimulus. The affect rule of thumb would substitute the systematic analysis of each of the stimulus attributes, thus increasing rapidity and automaticity of the process which, indeed, are also typical features of affective responses. As other mental shortcuts, the affect heuristic would be employed particularly in those situations where decision is very difficult (because, for instance, the individual does not know the necessary rules to address the problem considered, or does not possess the abilities to do so), where temporal limits impede a complete analysis of all the features of the situation (i.e., time constraints), or where the information available is not sufficient. As in other cognitive domains, judgments and evaluations about risk are strongly related to affect. Indeed, even in probability judgments, the affect experienced or imagined during information-processing may serve as a cue for the assessment of the probability magnitude (this point will be addressed again in Section 3 of Chapter 3). Such assessment, anyway, might be influenced even by the degree to which the ratio format is understood, an issue that will be examined in the next section.

---

to work as a guide to decision (affect as motivation and as "common currency"), see Peters, Lipkus, and Diefenbach (2006).

[12] The concept of "*affect"*- i.e., a "faint whisper of emotion" (Slovic, Finucane, Peters, and MacGregor, 2004, p. 312)- on which this heuristic is based has been explained by authors as "the specific quality of 'goodness' or 'badness' 1) experienced as a feeling state (with or without consciousness) 2) demarcating a positive or negative quality of a stimulus. Affective responses occur rapidly and automatically […]" (ibid.)

What is known on comprehension of different ratio formats

Doctors use the "1 in X" format more than "N in NX" formats

Frequency formats are extensively used in communications about risk. For instance, in a study by Michie, Lester, Pinto, and Marteau (2005) on the transcripts of 115 U.K. genetic consultations, 47% of the 492 risk expressions conveyed by practitioners were numbers (the remaining 52% were words, i.e., verbal probability statements); among those, the large majority was constituted by so- called "probabilities" (i.e., proportions or rates) rather than percentages (32% vs. 15%). Anyway, studies also reveal, that among ratio formats genetic risk is more frequently expressed as population size required for 1 expectant event (e.g., 1 in 8) than in what is considered the customary scientific format by experts in the field, namely rates of events per unit of population exposed to the risk, commonly 100 or 1000 (e.g., 12 in 100, in Miron-Shatz, Hanoch, Graef, and Sagi, 2009). Miron-Shatz and colleagues, for instance, stated that conveying prenatal screening test results in a "1-in-N" format in written communications (i.e., e-mails) to women is ordinary practice in Israel.

The preference of genetic counsellors for proportions with a numerator of 1 and shifting denominators in the expression of risks (e.g., 1 in 200 instead than 5 in 1000, for a .005 likelihood) has been found not to represent the implementation of a rule drawn from scientific literature attesting advantages of such format in any aspect supposed to improve the communication of risk. Rather, such inclination results from health practitioners' spontaneous attempt to make the population size statistics more understandable to the public (Hook, cited in Grimes and Snively, 1999). Despite health professionals' good intentions, some evidences seem to disprove their conventional wisdom that laypeople understand proportions better than rates, arguing instead for a facilitating effect of the customary scientific format over population size required for 1 expectant event. Such evidence will be briefly reviewed, and the issue of comprehension examined, in the next paragraph. Nevertheless, it is apparent that both formats are employed, as the following examples can show. For instance, in a report of the probability figures that a 37-year-old woman should weight in order to be able to make an informed selection about prenatal tests (i.e., between a screening test for chromosomal anomalies and a diagnostic test), Gates (2004) implicitly affirmed that what she named "probabilities" (i.e., "1 in X" expressions) and "frequencies" (i.e., "N in 1000" expressions) are the two normal ways of presenting statistics adopted in landmark epidemiologic studies of the prenatal diagnosis field. The same idea- leaving apart taxonomy issues (i.e., authors define "1 in X" formats as "proportions", and "N in 1000" as "rates") - is found in Grimes and Snively (1999).

Are there differences in the way individuals understand 1 in X formats and N in NX formats? This issue will be object of analysis of the next paragraph.

The advantage of rates over proportions in facilitating the performance of mathematical operations in medical decision making

Few studies have sought to determine which formats make it easier for laypeople to perform operations of the kind that might arise in medical decision making. Most of them have investigated only a single (but fundamental) operation: the ability of identifying which out of two probabilities is larger. One informative example is represented by the study of Grimes and Snively (1999), where that skill was assessed when the probabilities were expressed either as rates of disease per unit of population exposed to the threat (in that case, per 1000 people) or as proportions with 1 at the numerator and shifting denominators. To this aim, researchers presented women in numerous Obstetrics and Gynaecology outpatients clinics with a questionnaire asking, among other unrelated questions, to circle the higher out of two probabilities of having a bladder infection- both probabilities were expressed either in so-called rates (i.e., frequencies with a 1000 denominator: 2.6 in 1000 and 8.9 in 1000) or in so-called proportions (i.e., the population size required for the expected event, that is a frequency with a 1-in-n format: 1 in 112 and 1 in 384)[13]. Each individual assessed both formats, but the order in which they appeared was randomly varied to avoid sequence effects. Participants correctly identified the numerical risk expressions conveying the larger magnitude in 56% (1-in-n format) and 73% (rate format) of cases, thus showing a significant higher comprehension of rates than of proportions- the number of “don’t know” answers was always around 20% showing that many women did not understand either format. Superiority of the rate format was consistent across all primary languages, age groups, and levels of education of participants tested in the study.

The result was confirmed on a more representative sample constituted of laypeople (both women and men) by van Vliet, Grimes, Popkina, and Smith (2001). Using the same experimental design, and the same probability expressions of the previous study, but this time referred to the risk of Down syndrome, researchers replicated the finding of a superiority of the rate format on the 1-in-n format in terms of accuracy; its advantage on the other format (76.2% vs. 72.3% correct answers, respectively) was nevertheless slightly lower than that found by Grimes and Snively.

In another investigation, where percentages were tested apart from rates and proportions, the 1-in-n format confirmed its problematic character as it showed as the one creating most difficulties for the performance of mathematical operations of the types that might be encountered in discussions of risk (Cuite, Weinstein, Emmons, and Colditz, 2008). In that

[13] The risk values corresponded to the rates of Down syndrome at birth at the maternal ages of 35 and 40 years, respectively. Nevertheless, the scenario involved was that of cystitis instead of trisomy 21 to avoid generation of anxiety in respondents.

study, three waves of individuals (in total: 16,133) recruited online on a cancer-related site voluntarily took part in the research. Each wave received two experimental problems- each of which required the performance of one out of six mathematical operations (Wave1: compare and halve; Wave2: triple and add; Wave3: trade-off and sequence)- both expressed in one of the three formats under study (i.e., %, 1-in-n, and rate). All questions were presented as if they were hypothetical statements pronounced by a physician. In most cases, in risk levels used for each type of problem, probabilities were of an exactly equal size or nearly the same in the three formats (e.g., 1 in 24, 4 in 100, and 4%)[14]. A significant main effect on performance (i.e., answer accuracy) was detected for the type of format on all risk operations. Despite there was not one single format being best for all the six operations, each of them was best for at least one operation. The 1-in-n format was significantly better than the percentage format for the compare operations, but similar to the rate format (two results both in conflict with Grimes and Snively's). However, when scores were averaged across all operations, the 1-in-n format performed worse than the other two. Indeed, the mean accuracy rate for the 1-in-n format was only 45%, a significantly smaller result than the one obtained for frequency (55%), and than that obtained for percentage (57%) - the two were very similar.

Further confirmation of laypeople's difficulties with the 1-in-n format can be retrieved also in results of semi-structured interviews and correspondence with twenty women who had contacted Support after Termination for Abnormality (SAFTA), a British registered charity supporting parents who have foetal abnormalities diagnosed (Green and Statham, 1993). Among these women, who overall reported a high degree of anxiety before and after having received results of serum-screening for Down syndrome or amniocentesis results, at least eight had difficulty in applying a 1-in-n risk to their own pregnancy. The difficulty in grasping the meaning of 1-in-n expressions has been showed even in a study that reported individuals' preference for a combined percentage and frequency scale rather than a 1-in-n scale (Woloshin, Schwartz, Byram, Fischhoff, and Welch, 2000). Among available scales, the 1-in-n scale was judged as the hardest to be used. This last consideration, summed up with both the proved higher difficulty found with 1-in-n formats in performing operations normally required in health decisions, and with people's preference for other formats, made Woloshin and colleagues suggest medical personnel to avoid the 1-in-n format in risk communication.

Anyway, as documented in the first part of this section, doctors and health-care professionals appear to make large use of such format. Hence, an analysis of the effect that different ratio formats expressing a probability have on people's perceptions of the probability itself, becomes then of crucial importance. Studies performed, as well as theories that have been put forward to

[14] The risk value used in the "1-in-n" case was sometimes approximated so not to require harder calculations than the other two expressions.

account for individuals' irrational evaluations, will be analysed in the next two sections of the present chapter.

SECTION 1.

When 10 in 100 is higher than 1 in 10: Results in favour of neglect of the denominator

1.1 Explanations and experimental evidences under main theories

1.1.1 The different number of counterfactual alternatives prompted

The first research (chronologically) arguing for different evaluations of a ratio according to the magnitude of the numbers employed at its numerator and denominator can be considered that of Miller, Turnbull, and McFarland (1989). These authors observed a systematic tendency in their study participants to "[...] judge the same probability of an unlikely event as lower when the probability is presented in the form of a ratio of smaller rather than of larger numbers" (Denes-Raj and Epstein, 1994, p. 820). Their work was based on Kahneman and Miller (1986)'s Norm theory whose proposers had claimed, that the judgment of an event's "normality" (i.e., the judged degree of its occurrence in a given population) would reflect its capacity to evoke representations of similar events. In other words, according to Kahneman and Miller the easier it is for the individual to mentally simulate alternative ways to that event's occurrence (i.e., to produce post-outcome counterfactual thinking), the less normal, namely the less distributed, that event must be in the population of reference. Conversely, theorists foresaw that the more difficult it is to mentally simulate other ways in which an event could have occurred, the more normal (i.e., likely to happen) that event must be in the population of reference.

Interestingly, Miller, Turnbull, and McFarland demonstrated that this capacity to evoke representations of similar events is influenced by the superficial format used to describe the probability of the event. In particular, authors showed, that the magnitude of the numbers used to express the probability of the event mattered, with smaller numbers prompting a higher degree of counterfactual alternatives than larger ones in the mind of the perceiver. They found, that participants did judge the probability of a given event as more normal (therefore more probable) when conveyed through smaller absolute numbers (which as a result were judged as less normal). In the studies, participants were required to express how suspicious they were that the occurrence of the improbable event described in the scenario- each study examined a different scenario- had happened by chance. Perceived suspiciousness showed to vary on the basis of the size (smaller/larger) of the numbers used in the ratios to communicate the event probability, despite the mathematical equivalence of these expressions. More precisely, in line with the normality hypothesis, participants gave higher rates of suspiciousness for the occurrence of an improbable event when its probability had been expressed through a ratio of smaller numbers than when it had been expressed through a ratio of larger numbers (between-subjects design). For instance, in Study 1, the scenario was,

Imagine that you have a young child who loves chocolate chip cookies. Imagine further that you buy your cookies in packages that include oatmeal as well as chocolate chip cookies. Your child's practice is to go to the cookie jar and select the chocolate chip cookies, leaving the oatmeal ones to go stale. One day you think of a strategy to cope with the situation. You tell your child to close his eyes before he reaches into the jar, taking whichever cookie he grabs. He agrees to this and heads to the kitchen and the cookie jar. The jar contains 1(10) chocolate chip cookie(s) and 19(190) oatmeal cookies. Shortly, he comes back, exclaiming that he did just what you said and he selected a chocolate chip cookie. (Miller et al., Study 1, p. 583)

Participants expressed higher judgments of suspiciousness that the child in the scenario had peaked when he could successfully draw one of the favourite cookies from a jar if the latter had been described as containing 1 of the favourite cookies (and 19 of the non-preferred type) rather than 10 of the favourite cookies (and 190 of the non-preferred type), despite the equivalence of the proportions of the preferred biscuits in the two jars. Results can be read as showing, that people thought that the child's chance to have extracted one of the desired cookies without peeking was larger in the case of the urn containing 10 desired cookies out of 200 than in that of the urn containing 1 desired biscuit out of 20. Nevertheless, looking in detail at the experimental material, it must be noticed that participants' judgment did not actually refer to a ratio expression (e.g., 1 out of 20) where one of the two terms (i.e., the numerator) conveyed the instances of the event occurring (1, in the example) and the other (i.e., the denominator) the total number of possible events (e.g., 20). Rather, such evaluation had stemmed from an odd evaluation[15]. Actually, out of the five studies, only three (Study 3, 4, and 5) employed experiential stimuli asking participants to evaluate a ratio, while in the case of Study 1 and 2, the comparison between the absolute instances of the event occurring (e.g., 1) and those of the event not occurring (i.e., 19) was made explicitly salient. Thus, it could have been the case that the mental representations elicited in the two sub-groups of studies in Miller, Turnbull, and McFarland differed- three reasons supporting this affirmation are considered here below.

First of all, the mental operation normally prompted when evaluating an odd could be a comparison between favourable and unfavourable chances (e.g., 1 against 19); instead, that prompted in the case of a ratio could reasonably be the comparison between the positive chance of the event occurring and the total of cases (e.g., 1 out of 20). Secondly, the two messages might differ in terms of the effort required to the reader to build up a comprehensive picture of

[15] "1 chocolate chip cookie and 19 oatmeal cookies" is an odd expression.

the possible occurring outcomes and their respective chances. Indeed, while in the "odd-communication" version of the problem, people are explicitly given descriptions of both sides of the coin[16], in the "ratio-communication" version only one of those descriptions is provided[17]. What these observations point at is, as Denes-Raj, Epstein, and Cole (1995) have observed in their article, that only the type of stimuli used in Study 3, 4, and 5 of Miller and colleagues' work can be considered as properly in line with later research on the Ratio-bias phenomenon (see 1.1.2). Thirdly, some other factors could intervene to influence the degree of suspiciousness for the unlikely event occurring that not necessarily would be involved in the probability assessment of the magnitude of the same probability.

1.1.2 Experiential over rational system prevalence: Ratio-bias literature (CEST Theory)

Taking a start from the aim of proving experimentally the dual nature of their new information-processing system theory (Epstein and Pacini, 1999), research on the so-called "*Ratio-bias effect*" (Denes- Raj and Epstein, 1994) has been flourishing during all the nineties even independently from its original aims, up to the point to acquire a reason on its own. Indeed, the expression "*Ratio-bias"* was coined by Denes-Raj and Epstein (1994, p. 820) to refer to the systematic tendency to "[...] judge the same probability of an unlikely event as lower when the probability is presented in the form of a ratio of smaller rather than of larger numbers" (for the first time described by Miller, Turnbull, and McFarland, 1989). The appearance of such systematic tendency has later been attested with written vignettes illustrating improbable outcomes (e.g., Alonso and Fernández-Berrocal, 2003; Denes-Raj, Epstein, and Cole, 1995), but mainly through a game of chance created for the specific purposes of the research (e.g., Denes-Raj et al., 1995; Kirkpatrick and Epstein, 1992; Pacini and Epstein, 1999-a). Through the game, participants' choice between two options was demanded. Each option consisted in an urn containing balls in two colours (e.g., some white and some red), and described[18] as offering a given numerical probability of winning (or losing) associated with the withdrawal of a ball of

---

[16] As it will be highlighted throughout the next paragraph (i.e., 1.1.2), this happens in many stimuli employed in empirical studies under CEST approach.

[17] An additional operation is required to those people who want to figure out the other "side of the coin"; in other words, imagining the complementary event to that described requires the person to compute the probability that the event will not occur, i.e. the difference between the total possibilities and those favourable to the event whose chances had been originally described. The requirement of additional mental operations could have significant consequences on the way the message is elaborated (depending on, for instance, participant's degree of motivation, her/his abilities to compute, or the amount of available time) hence this operation could be performed or not. This, in turn, will likely influence the answer given to the task.

[18] If we exclude the only case where participants took part to the game in the laboratory (i.e. Kirkpatrick and Epstein, 1992, Study 3), in all other cases participants read a fictitious scenario describing the two options in written format.

one specific of the two colours (e.g., the red one). Alike in those other studies that have been classified as "heuristic" problems (Reyna and Brainerd, 2008; for a distinction between "heuristic" and "non-optimal" problems, see the same source), urns offered the same proportion of balls of either colour, but they differed in the total number of balls of each colour they contained, with one urn (described as the "large" one, in some studies) displaying a larger number of balls than the other (the "small" one). As an example of the typical game of chance proposed to participants, the following scenario illustrates:

> *Imagine that someone is 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. The person must draw one ticket (without peeking, of course) from either bowl: if he/she draws a ticket marked "winner" he/she wins $8.00, otherwise he/she wins nothing and the game is over.*
> *Even though the odds are identical for the two bowls, research shows that many people have a distinct preference as to which of these bowls they would rather draw from. Which bowl do you think most people choose in this situation?*
> (Kirkpatrick and Epstein, 1992: Exp. 2, p. 539)

As it is even openly remarked in the scenario, options in the problem offered an identical chance (.1) of withdrawing the target item. The rational answer would have been indifference toward the two urns, but that answer was generally achieved only when participants answered from their own point of view (i.e., "self-perspective"), or from the point of view of a "completely logical person" (see Amsel, Close, Sadler, and Klaczynski, 2009; Epstein and Pacini, 2000-2001). Instead, when answering from the point of view of "the average person" (e.g., Kirkpatrick and Epstein, 1992) or in other words, of "most people" (e.g., Denes-Raj, Epstein, and Cole, 1995; Pacini and Epstein, 1999-a), individuals tended to express a preference for the bowl offering a larger absolute number of winning tickets- in the example above, the 10/100 one[19]. Overall, findings supported Epstein and colleagues' predictions, as the self and logical perspectives always resulted in prompting a similar, mostly unbiased, answer (i.e., no preference between the two urns) while the others-perspective made the bias occurr (see Alonso and Fernández-Berrocal, 2003- for the heuristic pair; Denes-Raj et al., 1995- Study 3; Epstein

[19] The device of eliciting people's choices according to multiple views had been adopted in this and other studies with the explicit aim of making the dissonance among people's thoughts evident, with choices made under the prevalence of the rational self mirrored in personal and logical person's choices, and those made under the prevalence of the actual self mirrored in other people's choices (because people would not feel the exigency of presenting themselves as rational, in the latter cases).

and Pacini, 2000-2001, for standard instructions- i.e., textual without visualisation ; Kirkpatrick and Epstein, 1992; Pacini and Epstein, 1999-a except the lose condition in self perspective).

The interesting point is, why would people manifest a systematic irrational tendency to prefer the option whose probability of success is stated in ratios of larger numbers, to another option whose probability of success is, instead, stated in ratios of smaller numbers, despite having understood the two options mathematical equivalence? The explanation offered by scholars referred to the CEST (Cognitive-Experiential Self-Theory, Epstein, 1991, 1993, 2003) relating such bias to a prevalence of "feelings" over "reasoning". On the basis of study participants' feedback after the task completion, authors described the bias as resulting from the rational understanding of the mathematic equivalence of the options in the scenarios, but at the same time the subjective (experiential) feeling that the odds described by the "large" urn were more favourable than those described by the "small" one[20](e.g., Kirkpatrick and Epstein, 1992). CEST is the "dual process" theory comprehensively developed by Epstein, such theory, in line with the other so-called "dual system" approaches in judgement and decision making (e.g., Chaiken and Trope, 1999; Kahneman and Friederick, 2002; Sloman, 1996) assumed the existence of two alternating modes of thinking, an experiential and a rational one[21]. The attributes of the two system, listed in Table 1 as reported by Epstein (2003), can be summarised in author's words as: operating "in a manner that is preconscious, automatic, rapid, effortless, holistic, concrete, associative, primarily nonverbal, and minimally demanding of cognitive resources" (Epstein, 2003, p. 5) for what concerns the experiential system, and operating "in a manner that is conscious, analytical, effortful, relatively slow, affect-free, and highly demanding of cognitive resources" for what concerns the rational system (Epstein, 2003, p. 6). The two systems, operating in parallel and sometimes interacting, would not always be synchronous, but occasionally would let their inherent qualities become apparent to people, who instead normally are only aware of what appears to them to be a single process. According to authors, when the experiential system becomes robustly engaged and prevails on the rational one (either because the latter did not intervene to correct possible dissonant tendencies, or because, despite intervening, the appropriate rule to the situation had been bypassed, or instead could not be retrieved), Ratio-bias would occur. A more detailed description of the principles determining the irrational tendency is contained in the section above.

[20] Indeed, authors (e.g., Pacini and Epstein, 1999-a) considered that commonly experienced contradictory sensation whose arousal is spontaneous and hard to contrast as a clear evidence for the existence of the two distinct ways of information- processing inside the same person hypothesised under CEST (i.e., experiential and rational ways of information-processing).

[21] Contrary to the mainstream trend in dual theories considering the rational system as more advanced than the experiential one, and similarly to Fuzzy-Trace Theory (see, e.g., Reyna and Brainerd, 2008), under CEST the experiential system had been described as having a much longer evolutionary history than the rational system, and as such as not essentially flawed (see, instead, e.g., Chaiken and Trope, 1999, Kahneman and Frederick, 2002; Sloman, 1996).

Table 1
Comparison of the experiential and rational system attributes (Epstein, 2003)

Experiential system	Rational system
1. Holistic responding	1. Analytic responding
2. Automatic, effortless processing	2. Intentional, effortful processing
3. Affective processing: Pleasure- or pain-oriented (what feels good or bad)	3. Logical processing: Reason-oriented (what is rational)
4. Associative connections	4. Logical connections
5. Encoding of reality in concrete images, metaphors, and narratives	5. Encoding of reality in abstract symbols, words, and numbers
6. More rapid processing: Oriented toward immediate action	6. Slower processing: Oriented toward delayed action
7. Slower, more difficult changes: Changes with repetitive or intense experience	7. More rapid, easier changes: Changes with strength of argument and new evidence
8. More crudely differentiated constructs: Broad generalization gradient, stereotypical thinking	8. More highly differentiated constructs
9. More crudely integrated and less coherent networks: Dissociative, emotional complexes; context-specific processing	9. More highly integrated and coherent networks: Context-general principles
10. Passive and preconscious experience of events: We are seized by our emotions	10. Active and conscious experience of events: We are in control of our thoughts
11. Self-evident validity: "Experiencing is believing"	11. Need of justification via logic and evidence

Features of the experiential system causing Ratio-bias

Denes-Raj, Epstein, and Cole (1995) explained the bias as due to two attributes of the experiential system, i.e., the "concretive principle", and the "small-numbers effect"; a third facet, the "affirmative-representation principle", was further added to their theoretical explanation to account for people's choice behaviour when a negative outcome was involved (Pacini and Epstein, 1999-a). For what concerns the first principle, in authors' words (p. 307), "[…] the experiential system encodes and better comprehends numerosity than ratios because single numbers are more concrete than relations between numbers". Thus, the concretive principle would be responsible for people's strong tendency to be overly influenced by absolute numbers rather than by ratios, because of their more concretive nature (congruent with the concrete way in which experiential system would represent information). This facet is in line with what foreseen by the numerosity heuristic (Pelham, Sumarta, and Myakovsky, 1994), according to which people attribute a judgment of quantity (or probability) to a stimulus on the basis of the number of units into which the stimulus is divided, without completely taking into account determinant variables like the size of the units. In the case of Ratio-bias, the tendency to focus on absolute numbers rather than ratios could have lead participants to perceive the probability as higher when presented in larger rather than smaller numbers, as individuals would concentrate mainly on numerators (10 > 1) without taking enough account of the reference population out of which their value is expressed (100, 10).

In the second principle posited by scholars (i.e., the small-numbers effect) it is assumed, that the experiential system comprehends smaller numbers better than larger ones, because the former are more concrete than the latter, in the sense they are easier to visualize, a conventional test of concreteness (Paivio, 1991). Advocates of the CEST explanation (e.g., Pinto-Prades, Martinez-Perez, and Abellán-Perpiñán, 2006) had been again called Pelham and others (1994)' findings upon as proofs in favour of the fact that probability conveyed in (comparatively) smaller numbers would be easier to interpret than that in (comparatively) larger numbers. In such an experiment, participants were required to express their preferences for one out of two lotteries: one where all individuals received one ticket or one in which all individuals received 10 tickets. Results indicated that people had a preference for the "large" lottery (the one where they received 10 tickets) if they were told the number of participants was 1 million, but they did not show preferences for either lottery if the number of participants they were communicated was only two people. It appeared as when the lottery was described through lower numbers "they (could) better realize that the chances are the same in both lotteries" (Pinto-Prades et al., 2006, p. 120).

For what concerned the third principle of the experiential system responsible for Ratio-bias, instead, some actual evidences exist in support of their authors' explanation. That principle would explicate people's behaviour in those situations where a negative event is involved (e.g., the possibility to lose some money in association to the withdrawal of a red jellybean, in the case of the game of chance). In such instances, authors affirmed, that people would normally reverse their focus of attention compared with those situations involving a positive outcome- in the study of Kirkpatrick and Epstein (1992), a large number of participants reading a scenario entailing losses, indeed, reported to have focused on the desirable white jellybeans rather than on the undesirable red ones during the decisional process. Authors claimed this shift of attention would happen for in general affirmation is more concrete than negation: "[…] the experiential system can more easily encode positive representations (e.g. drawing a desirable white jellybean) than negative ones (e.g., not drawing an undesirable red jellybean)" (Pacini and Epstein, 1999-a, p. 310).

Furthermore, Pacini and Epstein also suggested and tested a supplementary effect (i.e., "the experiential-learning principle") which induces people to judge the numerical probability

of events occurrence from the representations encoded in the experiential system, namely generalisations from past emotionally significant experiences (Kirkpatrick and Epstein, 1992). For instance, authors argued, that because life is full of experiences in which, when facing the probability of “1 in a large number” of a given event, that event does not occur in fact (e.g., statistics on winnings at the lottery), people might have learned to classify the corresponding expressions as conveying the probability of an event that rarely takes place. Hence, people would judge the probability in an “1 in X” format (where X is a sufficiently large number) as smaller than the equivalent one conveyed in ratios of larger numbers (i.e., a format not making use of small numbers like 1 in the numerator). This principle is of particular relevance for the empirical work that will be illustrated in the present dissertation, hence it will receive further attention, especially in the conclusive chapter.
The mechanisms through which the three principles illustrated would interact to produce the Ratio-bias effect have been elucidated with clarity under the CEST theory, and will be reviewed in the following section.

Outcome valence, probability value, and effect intensity

The two main latent facets of the experiential system (i.e., the concretive principle and the small-number effect) have been described by authors as always operating conjunctively, but at the same time their visible responses are depicted as determined by their net effect (Pacini and Epstein, 1999-a). In particular, when related to a positive outcome (e.g., winning something), for low probabilities (e.g., 10%) both principles would work in the same direction inducing people to favour the large urn over the small one, as in the former target items have a numerosity advantage in the large urn- i.e., a greater number of winning balls is contained- plus, coherently with the small number effect, the associated probability format would convey less clearly for the experiential system the idea of a low probability than the one associated to the small urn. On the contrary, for high probabilities (e.g., 90%), the two facets would pull in opposite directions, with the large urn favoured by the numerosity principle, while the description of the small urn more clearly transmitting the idea of a high probability of the event occurrence (Pacini and Epstein, 1999-a). When related to a negative outcome (e.g. losing something), the effect would reverse, in line with the shifted focus of attention (i.e., the risk of drawing a jellybean of the undesired colour equals to the complementary probability of drawing a jellybean of the desired colour), thus determining a weak bias in favour of selecting the large urn when the event probability is low (e.g., 10%), and a strong Ratio-bias effect when the event probability is high (e.g., 90%). These assumptions on people’s tendencies were tested in an experimental study (Pacini and Epstein, 1999-a) that adopted the urn and balls game of chance, with participants assigned either to a positive (win) or to a negative (lose) condition, and the probability of withdrawing a red ball (10-30-50-70-90%) was varied within-subjects. As an example, the 30% condition read: "Consider a condition in which there are 30% red jellybeans in both trays. That is, there are 3 red jellybeans and 7 white jellybeans in the small tray, and there are 30 red jellybeans and 70 white jellybeans in the large tray" (Pacini and Epstein, 1999-a, p. 314). Participants asked from which tray they (and most people) would prefer to select a jellybean (small/ large/ no preference) generally confirmed the bias in others-perspective, with the relationship between probability magnitude and size of the effect represented by a negative linear trend in win conditions. Ratio-bias effect was significant for probability values of 10, 30 and 50%. A positive linear trend was found, instead, in lose conditions. Ratio-bias effect was significant for probability values of 50, 70, and 90%. Despite from a self-perspective the effect showed the same trend, differences in people's declared choices did not reach significance for most probability levels in win conditions, but they did reach it in lose conditions, with a significant preference for the option stated through ratios of large numbers for probability values of 50, 70, and 90% (Pacini and Epstein, 1999-a).

1.1.3 Denominator neglect as an inclusion illusion: Fuzzy-trace theory

The paternity of the discovery of those tendencies that like Ratio-bias are due to a neglect of the denominator, has been openly claimed by theorists of the "Fuzzy-trace" approach: "The ratio-bias phenomenon is a rediscovery of the same phenomenon that occupied researchers in the probability judgment literature beginning in the 1970's (for the most recent review of that literature, see Reyna and Brainerd, 1994)" (Reyna and Brainerd, 2008, p. 95). That phenomenon, generally dubbed "*denominator neglect*" [22], refers to, in Okan, Garcia-Retamero, Cokely, and Maldonado (in press, p. 6)'s words, "people's tendency to pay too much attention to numerators in ratios (i.e. the number of times a target event has happened) and insufficient attention to denominators (i.e. the overall opportunities for it to happen; Reyna, 2004; Reyna & Brainerd, 2008)".

Similarly to CEST (and other dual-process theories), Fuzzy-Trace Theory of reasoning, or the "New Intuitionism" as its founders have sometimes defined it (e.g., Reyna, 2004; Reyna and Brainerd, 1995) postulated the existence of two parallel modes of representing information[23] about the world in memory, namely, a vague qualitative intuitive gist-based one and a detailed quantitative verbatim one. Under this approach, though, it is further maintained that people have a preference for reasoning and performing decisions on the basis of the vague gist of information even in those cases in which verbatim (e.g., quantitative) detail of the it is

[22] Or "numerosity effect" as dubbed by Reyna and Brainerd (2008, p. 96).

[23] In opposition with most dual theories of reasoning and decision making, but similarly to CEST theory, intuitive thinking is generally considered a) more advanced than rational thinking under this approach, and b) more typical of adult age than of childhood, and c) characteristic of experts rather than novices (Reyna and Brainerd, 1994).

remembered. More precisely, individuals would normally favour to operate as closely as possible on categorical gist, namely, on the least precise representation that can be used to accomplish the task at hand. The logic of this preference would reside in its "economical" nature "[…] from an evolutionary perspective, in that many routine tasks requires only a fuzzy representation" (Wolfe, 1995, p. 86). Indeed, as Wolfe précised, "Just because subjects can discriminate differing quantities, and can act on those discriminations, it does not follow that problems are invariably solved by processing information at the highest possible resolution" (Wolfe, 1995, p. 86). According to authors under this approach, the explanation of Ratio-bias would reside exactly on the fuzzy processing preference just described. Ratio-bias would be nothing else than a reasoning error occurring from the wrong comparison of numerical parts to numerical wholes (Brainerd and Reyna, 1990; Reyna, 1991), likewise the other problems defined as "inclusion illusions" (e.g., base-rate neglect, the conjunction and disjunction fallacy, or overestimation of small probabilities). Indeed, Reyna (quoted in Barbey and Sloman, 2007, p. 258) observed that "problems in the inclusion illusions family have two-dimensional structures, with one dimension (the subset-subset) being salient and easy to process and the other (the subset-superordinate set), which is crucial to solution, being obscure". Denominator neglect would be a product of the structure of information, as "processing focuses on the subset mentioned in the question, the superordinate set recedes, and the question appears to involve nothing more than...a subset-subset comparison." Because of the nature of the structure, indeed, "Subsets disappear whenever the mind focuses on the superordinate set and the superordinate set disappears whenever the mind focuses on the subsets" (Reyna, 1991, p. 325).

In the original version of the prototypical task used to study "inclusion illusions" (i.e., Piaget's class-inclusion problem, see Piaget and Inhelder, 1951/1975), children until the age of ten (and adults in some cases) offered a display of seven cows and three horses and asked to report whether there were more cows or more animals in the display, generally erroneously indicated cows as more numerous. Those authors supporting the Fuzzy-Trace Theory of reasoning maintained the effect being created by overlapping classes as a "minor mental book-keeping confusion rather than a fundamental flaw in reasoning and memory" (Reyna and Brainerd, 2008, p. 95). This new view contrasted with both what theorist of information-processing had been arguing (i.e., that people's difficulty resides in general human working memory limitations), and with what Piaget and neo-Piagetians had been claiming about children's reasoning skills on the task (i.e., this type of errors would be due to a lack of logical competence). Indeed, findings have been brought which demonstrated, that children are able of grasping both the roles of numerator and denominator, and can perform a combination of both information in probability judgments (e.g., Acredolo, O' Connor, Banks, and Horobin, 1989; Offenbach, Gruen, and Caskey, 1984, cited in Reyna and Brainerd, 2008); therefore, the logical deficit hypothesis could be discharged. On the other hand, proofs of an absolute independency of the occurrence of the bias from memory capacity both in adults and in children (Brainerd and Knigma, 1985) made it possible to reject the memory dimension claim.

Some empirical evidences in Fuzzy-Trace theory: Developmental studies

Fuzzy-Trace Theory was informed to a great extent by developmental studies on probability judgment. Three of such studies will be described here.

In the study by Fischbein, Pampu, and Manzat (1970), six out of the eighteen trials employed involved the comparison between options whose probabilities were equivalent despite being expressed by means of numerators and denominators of different magnitudes. Participants (i.e., preschooler children) were presented with two sets ("boxes") of marbles in two colours (target and non-target). The proportions of target and non-target balls was the same for both sets, but the overall quantity of balls differed in the two cases, with one set having a higher numerosity than the other. This game of chance is very similar to CEST game of chance. Participants had to indicate which option (set) they believed offered the greatest chance of drawing a marble of the target colour. Sets could either be comparatively "small" (i.e., 2/4 vs. 1/2; 3/1 vs. 6/2; 4/2 vs. 2/1) or "large" (4/8 vs. 2/4; 6/2 vs. 12/4, 8/4 vs. 4/2). Authors could compare participants' performance in Ratio-bias-type pairs (same probability in the options) with that in the other pairs (different probability in the options): results indicated a higher number of uncorrected answers (i.e., below average) in the Ratio-bias-type pairs than in the others. Such higher degree of errors was explained in terms of the perceptual difference between the two boxes in the Ratio-bias-type pairs, one of which offered more target instances than the other (in absolute terms)[24]. Probably, participants instinctively draw a representation of the problem akin to the following: "urn A contains more winning balls than urn B", which would then translate in a choice of the larger urn (A), in line with the Ratio-bias effect.

The second study informing Fuzzy-Trace Theory that is described here (i.e., Acredolo, O' Connor, Banks, and Horobin, 1989), originated in the consideration that Hoemann and Ross (1982)' attribution to children of a lack of those abilities necessary to perform correct fraction calculations before the formal operation period could instead be the result of the methodology used by authors to assess those abilities. Such methodology, i.e., Piaget and Inhelder "choice paradigm" (1951/1975), consisted in asking children to choose between one out of two jars containing a mixture of target and nontarget items on the basis of the best chance of getting a target item on a random draw. In Acredolo and colleagues' view, such procedure was either incapable of detecting children's accurate employment of those skills, or could have

[24] Nevertheless, results of this study should be taken with caution remembering that they were obtained under three rather different conditions. For instance, in one case the three possible answers, i.e. left box, right box, or same chance in either boxes, were three and the participant's performance was reinforced after each trial; in another case, the necessity to estimate the chance of winning by relying on something else than the number of target cases was emphasized.

discouraged the use of such abilities while instead prompting alternative decision rules, like simply comparing relative numerators. When a different methodology was adopted, i.e., Anderson's Functional measurement procedure or information integration technique (Anderson, 1980), Acredolo and colleagues could show that children in fact possess the capacity to integrate two relevant stimulus dimensions in order to evaluate a third one. That methodology, originally used by its developer for other purposes, was adapted to demonstrate that even concrete operational children have an excellent quantitative grasp of the roles of numerators and denominators (and the way they should combine) in probability judgments. More precisely, in the first study presented by Acredolo and colleagues, children had to estimate on a simple visual non-numeric scale the probability of drawing a target jellybean from a bag containing either 1, 2, or 3 jellybeans of that colour and a total of 6, 8 or 10 jellybeans (they were exposed to each of the combinations of numerator and denominator across three replica). Results returned correctly ordered distribution of estimates by participants, demonstrating that children took into account both variations in the numerators and in the denominators: "with denominators held constant, higher estimates were assigned as the number of target items increased, and with numerators held constant, lower estimates were assigned as the total number of items increased. An appropriate multiplicative integration of cues was observed in the experiment" (Acredolo et al., 1989, pp. 936-937). Nevertheless, children demonstrated to be more influenced by variations in numerator than by variations in denominator. Such result could be nonetheless attributed to the nature of the task, given that numerator changes had been made very salient relative to changes in denominator. Therefore, a second study was ideated to correct for possible problems, by 1) giving equal salience to numerators and denominators, i.e., making them vary randomly across trials; 2) offering a larger selection of numerators and denominators values so that more critical cases were present in which the ratio with larger numerator had actually a lower probability (in this cases, the absence of errors would have confirmed that children were influenced by both numerator and denominator). In the computer-administered task, displays of planters were presented containing (2, 3, 4, or 5) flowerpots and(6, 8, or 10) pots in total; children had to assess the probability of a bug falling on a pot containing a flower. The evaluated probability was expressed on a continuous visual scale similarly to the first study. Results confirmed that children responded to changes in both numerator and denominator, showing even a more appropriate weight of both values than in the previous experiment: the majority of participants integrated those quantities at least additively, with a large number of them even multiplicatively[25].

---

[25] Analyzing children's performance on those displays presenting fractions (3/6, 4/8, 5/10) whose true probabilities were equivalent (= .5) shows higher probability in two out of the three cases, i.e. rather for 5/10 than for 3/6 ($t(59) = 2.35$, $p < .05$), and for 4/8 rather than for 3/6 ($t(59) = 2.60$, $p < .05$), but not for 5/10 than for 4/8 (*n.s.*). In these types of displays, whose paired comparisons are similar to the between

A further demonstration that children hold the capacities to assess probabilities considering both values at the numerator and at the denominator in an exact way (despite sometimes not showing those capacities in their performance) has been provided in the unpublished study by Callahan (1989). Author employed grouped trials of two-choice problems presented in a computer task: young participants had to decide from which of two buckets holding different quantities of two visually different types of balls (i.e., target and nontarget) they wanted to select in order to obtain the best chance of drawing a target (i.e., winning) ball. Author proved that the majority of younger children used an “Only Winner” strategy, in other words either a “Most Winners” strategy (i.e., they considered only the winner balls in each bucket, and selected the bucket offering the higher number of those independently from the proportion of winner to loser balls) or a “Fewest Winners” strategy (they considered only the winner balls in each bucket and selected the one offering the lower number of those, a result difficult to explain). Therefore, children tended to focus on what could be considered the relevant part of the information displayed to them, i.e., the value at the numerator of the ratio. Nevertheless, the strategy sophistication was related to age with older children displaying less confusion and using (correct) proportional reasoning to a greater degree than younger ones. This finding about children’s tendency to focus on the value at the numerator is completely in line with the concept of denominator neglect mentioned above for Fuzzy-Trace Theory.

Robust denominator neglect in the evaluation of health-related outcomes (Yamagishi, 1997-a)

Fuzzy-Trace Theory has been argued as explaining also Yamagishi (1997-a)’s results. Yamagishi studied probability perception of health outcomes expressed through frequencies with the purpose of investigating an implication from the results of other two studies he performed in 1994 (Yamagishi, 1994-a, -b). In the experiment of 1997-a, participants were asked to evaluate the mortality rates of well-known causes of death (see Yamagishi, 1994-a, -b) expressed in frequencies, varying both the magnitude of the population frame (Range: Wide, out of 10,000; or Narrow, out of 100) and the percentage incidence rate (Frequency: Smaller or Larger) within-subjects. The four conditions resulting from the experimental design- all 11 negative events expressed in SW (Smaller frequency Wide range, e.g., 1,286 in 10,000), LW (Larger frequency Wide Range, e.g., 2,414 per 10,000 people), SN (Smaller frequency Narrow range, e.g., 12, 86 in 100), and LN (Larger frequency Narrow Range, e.g. 24, 14 in 100) formats- were accessed in four separate sessions 7 days apart one from the other.

Results showed that ratings of risk (“How risky this appear to you?”) on a 26-point Likert scale whose extremes were labelled as “no risk at all” and “maximal possible risk” were

presentation of Ratio-Bias problems in single evaluation, reliance on the magnitude of numerators to estimate probabilities is detected as more prevalent than in others displays.

systematically greater in LW than in SW conditions, that in turn were greater than those in LN conditions, to end with the lowest evaluations for ratios of SN conditions. The phenomenon appeared to be robust as was present across mortality causes ($p < 0.05$, for 7 out of 11 causes) - see Table 2. When performing separate pair wise comparisons, in many cases the irrationality of participants' judgments resulted surprising when considered from the pure statistical point of view. For example, it appeared that cancer was rated as riskier when described as "killing 1,286 out of 10,000 people" than as "killing 24.14 out of 100 people", despite the first probability is in fact exactly half as big as the second one.

Yamagishi' s findings, in line with a total neglect of denominator, have been explained by the author with the combination of two cognitive mechanisms, i.e., the concurrent sensitivity to the rote frequency and the insensitivity to the total number expressed in the magnitude of the population frame, as manifestations of "anchoring and adjustment" and "base-rate neglect", respectively. As a consequence of the first tendency, people would start their assessment with a given reference point (i.e., anchor) suggested by the formulation of the problem (i.e., plausibly, the integer number of deaths), and would then adjust its evaluation until a plausible estimate is reached. However, the adjustment would tend to be insufficient (namely, not to take enough count of the denominator) because of its effortful nature, thus would stop once a plausible solution is found (Epley and Gilovich, 2006). Instead, as a consequence of the second tendency, i.e., base-rate neglect (Khaneman and Tversky, 1973), people would "[…] underutilize relevant information about population statistics [reference statistics] and instead over utilize other salient information [the number of deaths per year] (Kahneman and Tversky, 1973)" (Yamagishi, 1997-a, p. 497; information in squared brackets by the writer).

As claimed by Price and Matthews (2009, p. 445) these results could be also ascribed to "the amount of attention drawn to the denominator versus the numerator of the relevant ratio" (p. 445). Indeed it has to be noted that in Yamagishi' s study, participants were informed about the denominator (or base rate) at the beginning of the session only. The instructions read as follows, "Shown below is a list of causes of death. For each cause, the number of people who die of the particular cause is estimated. The estimation is the number of deaths per 100 people in the public every year. For each cause, please rate how risky it appears to you." As suggested by Reyna and Brainerd (e.g., 2008), when including the members of a smaller category (e.g., people expected to die) in a larger category (e.g., people exposed to a disease) is inherently difficult, the numerator would be easily remembered and processed, while the denominator would be neglected. The lack of repetition of the denominator in Yamagishi' s study may have affected participants' judgments, giving rise to the denominator neglect effect.

Finally, the last feature of the study making an extension of its results hard for the case at study in the present thesis (i.e., Down syndrome), is the absence of a weight for the possible influence of the severity of the cause on probability evaluation, a feature that will instead be controlled in the experimental studies presented in the present work. Because of that, in Yamagishi' study it is not possible to completely rule out an explanation in terms of diverse weight attributed by participants to the different lethal events.

Table 2

Mean perceived probability for the 11 death causes according to each of the 4 within-subjects exp. conditions in Yamagishi (1997-a)

	Ratio condition				
Death causes	**LW**	**SW**	**LN**	**SN**	**F**
Asthma	459[/10,000]	87[/10,000]	4.59[/100]	0.87[/100]	
	5.62	3.95	3.00	3.13	12.37***
Bronchitis	524[/10,000]	107[/10,000]	5.24[/100]	1.07[/100]	
	6.52	4.90	3.63	3.60	14.04***
Cancer	2,414[/10,000]	1,286[/10,000]	24.14[/100]	12.86[/100]	
	12.21	10.79	8.69	8.19	14.66**
Heart diseases	2,394[/10,000]	1,512[/10,000]	23.94[/100]	15.12[/100]	
	11.75	11.35	8.33	8.31	13.90***
HIV	1,255[/10,000]	735[/10,000]	12.55[/100]	7.35[/100]	
	10.02	9.35	7.33	6.73	13.12***
Homicide	1,373[/10,000]	487[/10,000]	13.73[/100]	4.87[/100]	
	10.23	7.25	6.69	6.13	18.80***
Influenza	585[/10,000]	141[/10,000]	5.85[/100]	1.41[/100]	
	6.02	4.77	3.44	3.33	10.38***
MVA	1,798[/10,000]	893[/10,000]	17.98[/100]	8.93[/100]	
	10.81	8.37	7.88	7.81	8.69***
Pneumonia	755[/10,000]	196[/10,000]	7.55[/100]	1.96[/100]	
	6.44	4.33	3.90	3.85	14.17***
Suicide	930[/10,000]	376[/10,000]	9.30[/100]	3.76[/100]	
	7.15	5.65	4.85	4.61	10.19***
Tuberculosis	590[/10,000]	157[/10,000]	5.90[/100]	1.57[/100]	
	5.60	4.12	3.58	3.77	8.07***

*** p=.001, **p=.05

Denominator neglect in the evaluation of safety in oral contraceptive use (Halpern, Blackman, and Salzman, 1989)

A further study supporting denominator-neglect is that of Halpern, Blackman, and Salzman (1989). They investigated probability perception of a side-effect associated with oral contraceptive use (i.e., death due to a circulatory disorder- “abnormal blood clotting, heart attack, and stroke due to haemorrhage”, ibid., p. 255). The numerical risk was presented in six different formats between-subjects: four base-rate information formats- i) frequencies (“1 in 12,000 die”), (ii) natural frequencies framed in a positive format (“99,991.7 out of 100,000 will not die”), (iii) natural frequencies framed in a negative format (“8,3 out of 100,000 die”), (iv) percentages (“.0083% probability of dying”)- and two relative information formats - (v) relative ratio (“4.15 times greater risk of death”) and (vi) relative percentages (“415% greater risk of death”). All formats conveyed an equivalent numerical information. Formats differed, apart from in being either positively or negatively framed, also in that the last two (compared to the other four) did not convey base-rate information. In other words, for the last two it was not possible to calculate the expected frequency of occurrence of the event described. Participants were asked to express the probability of the side-effect occurring relative to the probability of 11 other events (e.g., dying of the flu, getting divorced, bus collision with a train, having an appendectomy) on 7-point Likert scales whose extremes where “death due to circulatory disorder much less likely than this one” and “death due to circulatory disorder much more likely than this one”. Researchers supposed that people would experience difficulties both in 1) evaluating a base-rate information format, given that it “conveys little that is meaningful and useful because consumers have no first-hand experience with base-rates of this magnitude as they cannot detect the incidence of such low-probability events” (Halpern, Blackman, and Salzman, 1989, p. 253), and 2) in assessing a relative information format, due to its intrinsically ambiguous nature. Due to those reasons, they predicted that “oral contraceptive safety assessments would be determined by the absolute size of the number presented- the only interpretable information left when you disregard the information format” (Halpern, Blackman, and Salzman, 1989, p. 253). In other words, they expected that, not differently from what had been described in Yamagishi (1997-a), people would have put in execution a rule of thumb according to which 1) they would have ignored the format in which the numerator was embedded, and 2) they would have simply estimated its magnitude in a fuzzy way (i.e., a classification of the kind of, for instance, “small” for numbers lower than 10 and “large” for numbers higher than several hundred). Results corroborated authors’ hypothesis in that 1) when the base-rate information was given, events were systematically perceived as less probable than when such information was not given (i.e., in the case of a relative format); moreover, 2) people ignored the difference between the two relative formats, focusing as expected on the absolute magnitude of the numbers (415 > 4,15). Both findings went in the direction of a neglect of the information contained (explicitly or implicitly) in the denominator. Nevertheless, the specific hypothesis formulated on the evaluations of two of the formats, i.e., i) and iii) that are expressed in ratio between smaller and larger numbers (the hypothesis read, "Because differences between very large numbers are difficult to comprehend, respondents attend to the more meaningful and smaller foreground numbers in assessing safety. Thus, when confronted with risk information in the format "X out of Y occurrences", respondents will focus on the smaller, more meaningful, and more salient X on foreground number and will tend to ignore Y, the large background number"- Halpern et al., 1989, p. 254) was not confirmed. No significant difference in the perception was detected for 1 in 12,000 and 8, 3 in 100,000. Thus, there was no specific confirmation of the denominator neglect in these two experimental conditions.

Unfortunately, results are not directly applicable to the context examined in the present thesis, as, indeed, participants were not required an absolute evaluation of the probability of the side-effect occurrence, rather an assessment relative to *their* evaluation of other (eleven) death causes. This meaning, that since the evaluation of such other death causes had not been assessed (and indeed it is expected to show wide variance among participants), data did not allow a direct comparison between individuals' assessments of target risks of similar magnitude , as in fact each of them refers to a different (not weighted) comparison.

1.2 Other Interpretations

1.2.1 Pseudo-multiplicity: Perceived Diversity Heuristics

A different explanation of people's preference (in positive domain) for the option whose probability is stated by means of a ratio between larger rather than smaller numbers, has been given under a theoretical framework called Perceived Diversity Heuristics (PDH, Ayal and Zakay, 2009). The primary interest of such an approach was that of offering an explanation for the advantageous naïf behaviour showed by individuals when evaluating the risk of a pool of goods. Indeed, in most of the cases, they prove to "know" (despite having no conscience of it) a specific normative way to decrease the risk associated with groups of events, thus revealing an intuitive understanding of "portfolio theory" (Coombs, 1975; Markowitz, 1952). Portfolio theory is a formal account advising investors to reduce the risk connected with their action by means of diversifying the sources of investment trough the combination of different stocks in the same portfolio. In such a way, indeed, this way a poor performance in one field can be offset by better performance in others.

Ayal and Zakay argued, that in their actual behaviour people attempt to implement portfolio theory. However, beside these so called "normative paths" corresponding to situations in which the activation of the Perceived Diversity Heuristics would be justified because it corresponds to a real diversity, pseudo-paths can also be present. While the first group of paths (the normative ones) leads to an adaptive judgment of risk reduction, the second group (pseudo-paths) leads to a pseudo- diversification bias (see below). Pseudo-paths are created by the identification of dimensions (i.e., pseudo-distinctiveness and/ or pseudo-multiplicity) which do not affect the measures of distributions of prospective outcomes that decision-makers should compare, i.e., the actual variance or the expected utility of the pool. Perceived distinctiveness (which can be influenced by multiple dimensions, e. g., the degree of similarity or the physical distance among the sources), regards the degree to which goods in the pool are perceived to be different one form another, while perceived multiplicity (which can be manipulated quantitatively), regards the number of sources of the pool.

Under PDH, pseudo-diversification is deemed as the cause of Ratio-bias (see the parallelism with the concept of numerosity heuristic in Pelham, Sumarta, and Myakovsky, 1994). In games of chance, people would choose the larger of two urns offering the same probability of success because the numerator in the ratio would be erroneously perceived as a source of greater multiplicity (i.e., a range of pseudo-diverse sources) when such probability is expressed in (comparatively) larger rather than smaller numbers. Due to this reason, "the Ratio-bias could therefore be reinterpreted as an illusion of risk reduction that leads to pseudo-diversity bias" (Ayal and Zakay, 2009, p. 560).

Study 1C is the only study in the paper employing a problem that is somehow comparable to the classic game of chance of CEST. Three scenarios described a highly, a moderate, and a non-diversified lottery offering possibility/ies of winning a single car or more than one car (Ayal and Zakay, 2009). Lottery A (the "highly diversified lottery") was described as follows, "10 prize cars (4 Volvo S-80s, 3 Toyota Corollas, 3 Honda Civics) will be raffled among 100,000 costumers"; Lottery B (the "moderately diversified lottery") was described as follows, "10 prize cars (10 Volvo S-80s) will be raffled among 100,000 costumers"; and description of Lottery C (the "nondiversified lottery") read, "1 prize cars (Volvo S-80s) will be raffled among 10,000 costumers." Despite two of the three lotteries (i.e., B and C) were equivalent from a normative point of view because they offered the same outcome (i.e., the possibility of winning a Volvo S-80 car) with the same probability (0.1), they differed in the superficial presentation of the probability in two respects, namely 1) the format of the ratio expressing it, and 2) the degree of elicited diversity.

For what concerned the first point, one probability was stated using a ratio employing larger numbers (i.e., 10 in 100,000, or "large lottery" hereafter) while the other using a ratio employing smaller numbers (i.e., 1 in 10,000, or "small lottery" hereafter); regarding the second point, the large lottery offered 10 possible prizes (10 Volvo V-80) while the small one only 1 (pseudo-multiplicity), despite in fact from the scenario description it seemed that the draw was a one-off event instead than a multiple event.

Two conditions were tested, a within-subjects condition (253 participants) and a between-subjects condition (143 participants). In the within-subjects condition, each participant read the description of two out of three lotteries, and had to decide how to allocate 100 points between them. In the "multiplicity version", participants had to split the points they owned between the moderately diversified pool and the nondiversified pool (i.e., Lottery B vs. Lottery C). In the "distinctiveness version", participants had to split the points between the highly diversified pool and the moderately diversified pool (i.e., Lottery A vs. Lottery B). Finally, in the "multiplicity plus distinctiveness version", participants had to divide the points between the highly diversified pool and the nondiversified pool (i.e., Lottery A vs. Lottery C). Instead, in the between condition people saw only one of the three lottery descriptions (i.e., Lottery A, B, or C), and were asked to state the maximum amount of money that they would agree to pay as a registration fee for the lottery[26].

Results of the within-subjects condition analysis on 232 participants proved that 58% of those participants assigned to the questionnaire including B and C lotteries showed indifference between them, as it should be expected if people behave rationally. Notwithstanding, when

[26] As a control measure, participants were then asked to rank order the three cars of the highly diversified lottery (Volvo S-80, Toyota Corolla, Honda Civic) according to their price, from the most expensive to the least expensive, both in within- and between-subjects conditions.

analyzing the remaining participants' answers (i.e., those expressing a preference for one of the two lotteries) there was a significant difference between a majority (33.3%) preferring the lottery offering the greater absolute number of winning instances (i.e., the B Lottery), and a small percentage of them (9%) preferring the nondiversified C Lottery. The difference between the two preference groups was significant in the PDH predicted direction, χ^2 (1) = 10.94, $p <$ .01.

Results of this version and the magnitude of the diversification preferences presented directly replicated the Ratio-bias reported by Epstein and colleagues (Kirkpatrick and Epstein, 1992), because people attributed higher points to the lottery offering a winning probability expressed with larger numbers- i.e., Lottery B, offering 10 prizes in 100,000, than to that expressed with lower numbers- i.e., Lottery C, offering 1 prize in 10,000. Nevertheless, an alternative possibility could be, that the higher-numbered option was interpreted as actually offering a higher number of prizes than the smaller-numbered option.

For what concerned the between-subjects data, analysis on the 131 participants' measures of the Willingness To Take part (WTT) on the lottery showed the highest price for the highly diversified Lottery A, followed by Lottery B and then by Lottery C, in line with the degree of diversification[27].

Through their studies, authors demonstrated that people do exhibit a diversity preference in gain conditions but a diversity aversion in conditions of loss. In other words, diversification would operate in synchrony with the framing effect. More precisely, under conditions of gain, people would tend to diversify because it reduces the risk of outcomes and ensures that at least some resources will be fruitful (i.e., increasing the probability of at least one gain, see also Kahn and Lehmann, 1991). On the other hand, under conditions of loss, people would prefer to avoid diversification, as they believe that such a strategy would enable them to escape losing at all. On the contrary, we know that in fact diversification can actually protect from extreme losses.

1.2.2 Exemplar availability: Exemplar-cuing theory

Exemplar-cuing theory can […] *account for ratio-bias, though it does not require a comparison between options that differ in the absolute number of ways a favourable outcome can occur. Instead, exemplar cuing requires only consideration of whether examples of other 'winners' are cognitively available, regardless of the actual number of ways that winners might appear.*

(Koehler and Macchi, 2004, p. 541)

[27] A limitation of these measurements consisted in the large deviations from the mean (SD_1 = 11.38; SD_2 = 5.64; SD_3 = 8.25) obtained on the WTT that are due to task assignments- experimenters did not fix a minimum-maximum amount of money that could be offered by participants, thus fostering largely different sums. Also, since demographic variables like income are not specified in the study, it is not possible to get an adjustment of those WTT measures according to the value attributed to money by the single persons, thus making it hard for that measure to return actual participants' behavioral intentions.

Despite Exemplar-cuing theory (Koehler, 1997, 2001-a, 2001-b; Koehler and Macchi, 2004) has been proposed in the contest of legal decision making to explain when and why legal jurors are persuaded by DNA match statistics, its authors claimed it could be applied to other domains to explain how people think about low-probability events. For instance, to give an account of the Ratio-bias phenomenon (Koehler and Macchi, 2004). The theory predicts that the description of an unlikely coincidental outcome will be valued differently depending on the description's ability to cue similar examples. "The perceived probative value of a statistical DNA match (and, by extension, other forensic match evidence) depends on the ease with which triers of fact can imagine examples of others who would also match the DNA profile. When triers of fact find it hard to imagine examples of others who might match by chance, the evidence will be treated as compelling proof that the matching suspect is the source of the recovered DNA evidence. But when such matches are easier to imagine, the evidence will seem less compelling." (Koehler, 2001-a, p. 8).

Authors developed precise mathematical rules governing their forecast for people's evaluative behaviour. "The weight decision makers attach to low-probability events is, in part, a function of whether they can easily generate or imagine exemplars for the event" (Koelher and Macchi, 2004, p. 540), namely, a function of "exemplar availability". The latter would work in an inversely proportional way to the convincing value of the evidence, thus it would stand as an indicator of the subjective weight that people attach to low probability events. Exemplar availability would depend on both the reference-class size (the target, either single or multiple) and the rareness of the event under exam, which interact in a significant way[28]; two mechanisms, i.e., the "multiplicative" and "numerator" mechanism, describe this interaction. According to the first one, when the product of reference-class size and its incidence is greater than 1, exemplars would be very likely to be cued, therefore people will give more weight to the possibility that the unlikely coincidental outcome will occur. However, when the product of the two factors be lower than 1, exemplars would be less likely to be cued, and people give little weight to the possibility of the unlikely event.

[28] Despite being based on the same principle governing the availability heuristic (Tversky and Kahneman, 1973) - i.e., the fact that people form their probability evaluations about an event on the ease with which similar instances come to their mind- Exemplar-cuing differs in the explanation on the reason why events would be mentally available. According to Koehler (2001-b) and Koehler and Macchi (2004), the mental presence of an exemplar does not have necessarily to be related to the prevalence of that event in the environment– as instead the literature on heuristics and biases affirms- rather it could be based on vividness (Nisbett and Ross, 1980), i.e., the ability of information to catch and hold people's attention, which in turn is influenced by how much unusual, important, interesting, personally relevant, publicized etc. that information is. According to Exemplar-cuing theory, once into mind that information would become crucial as it has the power of changing the personal evaluation of the event, and therefore, of influencing her/his choice. Authors specify that the latter affirmation would be especially true for low-probability events which would be ignored otherwise.

The second mechanism, which operates only on information presented in frequencies, states that exemplars are cued when the numerator of the incidence rate is equivalent or bigger than 1; thus, decimals (e.g., 0.1 out of 10) are not expected to prompt examples. Please consider as illustration, the following two descriptions of the chance of winning the daily three-digit New York Lottery Numbers game employed in Koehler and Macchi (2004, p. 540),

a) "There is a 0.1% chance that a given ticket will win";
b) "One in every 1,000 tickets out of the 500,000 tickets that are sold each day will win."

In a), since the numerator of the incidence rate is smaller than 1, that expression is not expected to prompt examples. The contrary happens in b), because 1) the incidence of the phenomenon (1 in 1000) is expressed by means of a 1; and 2) it is provided in combination with a large reference class (e.g., 500,000 tickets sold); therefore, winning-ticket exemplars are likely to be cued (approximately 1/1000 x 500,000= 500 tickets in this lottery will be winners).
Despite in both experiments included in Koehler and Macchi (2004) data were consistent with the two mechanisms claimed to govern the cue of exemplars, it has to be considered that those are the only evidences available for the multiplicative mechanism. For the strong predictions made on the basis of the two principles (the multiplicative one in particular) to be claimed as valid, it would be advisable to test a broader range of values as incidence rates and reference classes, and also to test them in different contexts from that of DNA.
Implication for the Ratio-bias effect stemming from Exemplar-cuing Theory will be described in the following section.

1.2.3 Visualization of the numerator

In the domain of judgment and decision making research, the shared acceptation of the term "imaginability" (in the wake of Tversky and Kahneman, 1973) [29] is, the ease with which associations or instances of the probability to be judged do come to mind. Some research (e.g., Carrol, cited in MacInnis and Price, 1987) suggested that imaginability would affect perceived probability directly, by calling upon the availability heuristic[30] (Tversky and Kahneman, 1973). In particular, it was argued that the experience of imagining an outcome would increase its perceived probability by making it more salient and easier to recall. In their work, Price and Mattews (2009) explicitly referred to imaginability as a factor playing a role in, and by some

[29] The term has been used by, among others, Newell, Mitchell, and Hayes (2008), and does not correspond to imagery, in that it is a property of the stimulus and not an individual's characteristic. In some of the literature of the field, there appears to be confusion on the distinction of the two terms.
[30] According to Availability heuristic, people would predict the frequency of an event, or a proportion within a population, based on how easily an example can be brought to mind.

authors even deemed as a cause of, “the ratio bias and similar effects” (p. 445). These contributions will be described in the two sections here above.

1.2.3.1 Imaginability/images of the numerator and affect

Some authors have explicitly alleged, that the ease of imaginability is influenced by the statistical format used to convey the probability, with frequencies producing higher probability evaluations than equivalent percentage presentations (see Slovic, Monahan, and MacGregor, 2000, Study 3, also described in the last part of this section). Slovic and colleagues, urged that even Ratio-bias results obtained under CEST studies (e.g., Kirkpatrick and Epstein, 1992; Denes-Raj and Epstein, 1994) could be explained by imaginability. According to them, in those experiments, imagined or actually seen images of winning beans would have conveyed positive affect, which in turn would have motivated the choice of the urn containing the larger absolute number of red beans. Similar studies under the CEST (i.e., Slovic, Monahan, and MacGregor, 2000- Study 3) demonstrated furthermore in one of their investigations, that some measures could be affected by the format used to describe the harm probability (“N out of 10” rather than “10N out of 100”) of a psychiatric patient, and such tendency was related to the degree of imaginability evoked in each case. Participants (members of the American Psychology-Law Society) were sent one of seven version of a one-page questionnaire describing the short assessment report of a hypothetical patient with mental disorder (Mr. Jones) about to be discharged. The versions varied only in the way the patient’s probability of harming someone other than himself was expressed. In particular, conditions number 3- 4 and 6-7 are those of interest to the present case, where the magnitude of the numbers employed at the numerator and denominator of the ratios in each pair was either (comparatively) smaller or larger:

Of every 10 [100] patients similar to Mr. Jones, 1 [10] is [are] estimated to commit an act of violence to others during the first several months after discharge.
Condition 3 (in squared brackets the expressions used in condition 4)

Of every 10 [100] patients similar to Mr. Jones, 2 [20] are estimated to commit an act of violence to others during the first several months after discharge.
Condition 6 (in squared brackets the expressions used in condition 7)

Role-playing the supervisor of the mental health facility, participants had to evaluate the degree of risk posed by Mr. Jones (low/medium/high), they had to express an opinion on the option of discharging the patient (discharge/ not discharge now/ obtain second opinion), and finally they had to state the degree of closeness suggested for monitoring to the patient when in the community (not/ somewhat/ very closely). There did not seem[31] to be a difference due to the magnitude of the numerator and denominator in the ratios employed to convey the risk neither in the probability judgments, nor in the opinion on the degree of closeness suggested for monitoring. Nevertheless, there was a difference in the behavioural intentions according to the perceived degree of risk showed an influence (in Ratio-bias direction) of the magnitude of the numbers used. Indeed, while in those receiving a smaller-numbered ratio condition (i.e., 3 and 6) the percentages of participants refusing to discharge the patient (“do not discharge now”) were 21.2% and 20.3% (for the 1/10 and for the 2/10 probability, respectively), in those receiving a larger-numbered ratio condition (i.e., 4 and 7) such percentages jumped to 21.0% and 40.6% , respectively. In sum, the superficial difference among the two types of formats had been able to double the number of people taking a hypothetical decision, even though from the pure rational point of view no reason could justify the fact that the same probability elicited different answers. Moreover, in the prior studies of the same article (Studies 1 and 2), authors showed that a frequency presentation of the same probability scenario (20 out of 100) generated higher risk judgments that a percentage presentation (20%), and they ascribed the effect obtained to the capacity of frequencies to generate a “terrifying image” of the recidivist in the mind of the clinicians involved in their study. Similarly to other studies (e.g., Finucane, Alkhami, Slovic, and Johnson, 2000; Finucane, Peters, and Slovic, 2003; Loewenstein, Weber, Hsee, and Welch, 2001; Slovic, Finucane, Peters, and MacGregor, 2004), since the higher probability evaluation had been detected in conjunction with extreme feelings appraisals, the affective richness of the imagery evoked by the statistical format had been attributed the power to mediate that relationship. Indeed, the explanation given to the more extreme behavioural intentions under the 20 in 100 rather than the 2 in 10 in Study 3 referred once again to the higher degree of imaginability evoked under the larger rather than smaller frequency presentation. It has to be noticed tough, that authors’ claim of a higher degree of affective images elicited for larger numbers compared to smaller ones is in fact a speculation[32]. Indeed, it was not provided an assessment of the capacity of the two formats to prompt affective imagery.

In a similar vein to Slovic, Monahan, and MacGregor’ s study (2000), Peters, Vastfjall, Slovic, Mertz, Mazzocco, and Dickert (2006) hypothesized a mechanism based on the affect generated by the evoked images to explain results of their fourth study. That study examined the effect of numeracy on various decision making tasks, among which that involving the ball-and-

[31] Because of not being the main object of interest of authors, no results of statistical tests are presented in the article for evaluations under the smaller and larger numbers ratio conditions. Nevertheless, it is still possible to observe some tendencies.

[32] “A[*n*] […] issue, is the extent to which different formats do indeed affect imaginability. In most studies these effects are simply inferred from the responses on other dependent measures (e.g. willingness to take risks) (Slovic et al., 2002). Some work has been done examining the relation between imagery and decision-making (e.g. Slovic, Layman, Kraus, Flynn, Chalmers, and Gesell, 1991) but evidence of the direct link between frequency formats and enhanced imaginability is currently lacking in the literature” (Newell, Mitchell, and Hayes, 2008, p. 319).

jar game of chance is reported here. As in Denes- Raj and Epstein (1994), the small urn of Peters et al.'s scenario (described as the one "having 10% colored jellybeans") offered a 1 in 10 probability of drawing a colored jellybean, while the large (labelled as the one "having 9% colored jellybeans") offered a 9 in 100 chance of drawing a colored jellybean. As expected, the majority of participants irrationally chose the large urn, despite being the sub-optimal option, thus showing a Ratio-bias. Authors hypothesized a leading role of people's affective response in those situations. They indicated the cause of the bias in the images mediating the affective response, namely "participants' affective images of 9 winning balls in the large urn dominated the image of 1 winning ball in the small urn" (Peters et al., 2006, p. 410), even though it must be noticed that there are not evidences for this explanation in their study.

In their view, in such a game of chance two conflicting types of affective reactions would be experienced, one stemming from the mental (or real, in the case where the game of chance had been played in reality) image of the balls and the urn, and the other due to a conscious and logic thinking process. A dual-system theory of information-processing view is implicit in this interpretation, but differently from most dual process theories where only the experiential system is deeply influenced by affective information, in the model proposed by Peters and colleagues affect has the capacity to influence both the experiential and the rational systems. Indeed, authors stated, "Affect can be a direct "hit" from an object"- Zajonc's (1980) notion that affect comes before conscious deliberation- "or it can be the result of prior deliberation."(Peters et al., 2006, p. 410). The comparison among the two urns made under a leading role of the experiential system would generate a more positive (or less negative) feeling of affect for the larger urn. On the other hand, the affect derived from thinking deliberately at the stated probabilities applying mathematical norms would make the small urn look advantageous as offering an objectively greater probability than that of the other urn. Nevertheless, this supposed double influence of affect was not in fact tested in the study, as only a single affective rating was assessed. Indeed, participants' degree of affective feelings was assessed by asking them to indicate "how good or bad Bowl's A 9% chance" made her/him feel, on a 7-point scale ranging from -3 (very bad) to +3 (very good). Also, mean evaluations according to the choice made (large or small urn) were not reported in the study, thus it is not possible to test if there was a different degree of feeling for the two options in study participants.

1.2.3.2 Imaginability/images of the numerator, but not necessarily affectively-tagged

Two views will be presented here that offered an opposite (in terms of direction of causality in respect to Slovic and colleagues' study, described in 1.2.3.1) interpretation of the relationship between the format of probability and images generated.

The first view is that of Reyna and Brainerd (2008). According to them, images are not the cause of the higher probability judgments given under the frequency format if compared with the probability format, rather, in the example, they are a side-effect of the categorical judgment of violence (in the example) elicited by the frequency format but not under the probability one. In fact, according to the Fuzzy-Trace theory (see 1.1.2), “20 people out of 100” would be evaluated as all categorically violent, while “20%” -referred to a single person- would be evaluated as a relatively small tendency (or even an inexistent one) to engage in a violent behaviour. Reyna and Brainerd used the same logic to give account of the higher number of participants declaring they would not discharge Mr. Jones in the 20 out of 100 (and 10 in 100) condition than in the 2 in 10 (and 1 in 10) one.

The second view in contrast to Slovic and colleagues about the role of imaginability in affecting probability judgments of ratios expressed with smaller or larger numbers is that of Koelher and Macchi (2004). According to researchers, first and most importantly of all, to get an impact on people’s evaluations there would be no need for the images generated by stimuli to be affectively rich (likewise, instead Slovic, Monahan, and MacGregor, 2000, and Tversky and Kahneman’ s availabity heuristic, 1973). Instead “the weight decision makers attach to low-probability events is, in part, a function of whether they can easily generate or imagine exemplars for the event” (Koelher and Macchi, 2004, p. 540). Hence, in their Exemplar-Cuing Theory (see 1.2.2), differently from affect-laden theories (see preceding section, i.e. 1.2.3.1), for an effect to be produced it would suffice that the statistics evoke thoughts about examples of the target event. A second difference between Exemplar-Cuing Theory and other explanations, no privileged role of the frequency format is claimed over the percentage one by Koelher and Macchi, as the generation of exemplar cues depends on the reference class that is responsible for identifying the sample space and not on the numerical expression used to convey probability (see Macchi, 2000). Indeed, as Newell, Mitchell, and Hayes (2008) critically observed in a study examining positive and negative low-probability events, no evidences exist of a direct link between enhanced imaginability and the format used to describe probability, in particular frequency formats. Newell and others affirmed that in most studies these effects were postulated indirectly from results on other dependent measures, like Willingness to Take Risks (see for instance Slovic et al., 2000). A common critical point in all the above-presented approaches trying to explain the reasons why ratios in larger number formats would elicit higher probability evaluations than ratios in lower number formats is, then, that of how imaginability could be properly assessed. That issue will be analysed further in a study expressly devoted to the measurement of imaginability (see Study 3.2).

1.3 Elicitation of health state utilities, of numerical frequencies of social facts, and of the probability of harm posed by a mental patient if discharged are influenced by Ratio-bias

The Ratio bias has also been shown to influence the elicitation of health-state utilities[33] (Pinto-Prades, Martinez-Perez, and Abellán-Perpiñán, 2006) by means of two different methods, i.e., the Standard Gamble (SG), and a method that authors have dubbed the "double lottery" method (ibid.).

SG is very common in research on individual decision making, but is employed in social decision making, too. Not only, in fact utilities can be used to estimate the expected utility of a health treatment, in order to help decision on the treatments available, but also they can serve to gain estimation on the cost-effectiveness ratio of medical treatments, in order to take resource allocation decisions.[34] In the SG technique for chronic health states, usually participants are asked to choose between suffering a given condition for the rest of their life, and a medical treatment that can return them to full health, considering anyway that as a side-effect it could also fail and cause immediate death. The formulation of questions employed in Pinto-Prades and colleagues' Study 1 will illustrate:

Suppose that you are experiencing health state X. If you do not receive treatment you will remain in X for the rest of your life. However, you can receive a medical treatment (treatment ALFA) that if successful, will result in return to normal health. Nevertheless, treatment ALFA can also fail and in this case you will die. We are going to show you different probabilities of success and failure and you will tell us if you think you would chose treatment ALFA or no in each case.

(Pinto-Prades et al., 2006, p.124)

Authors were interested in studying whether the way in which the probabilities of success and failure of treatment ALFA for four diverse health states (X, W, Z, and Y) expressed in frequencies of smaller or larger numbers, could produce different subjective utilities. In particular, by employing a ping-pong search procedure (i.e., first presenting a 5% risk of death, then a 90%, then 10%, then 80%, 20%, 70%, 30%, 60% and so on) they aimed at testing

---

[33] By health utilities it is meant how a person thinks having a particular condition would compare to the best option (being in perfect health, at one extreme) or to the worst option (being dead, at the other).

[34] In the first case, for chronic health states usually the SG technique asks people to choose between suffering a given condition for the rest of her/his life, and a medical treatment that can return him/her to full health, but could also fail and cause her/his immediate death. In the other case, the cost- effectiveness of the introduction of a medical intervention can be calculated as the difference between the utility of patient health state before and after treatment.

whether the indifference point between the gamble offering a certain probability of death and the complementary probability of complete recovering, and the actual health-state would change in a direction compatible with the Ratio-bias effect. To illustrate, it was hypothesized, that a higher elicited utility (because of the lower degree of risk accepted, i.e., the lower indifference point) be elicited when the probabilities in the gamble were presented out of 1000 people, than when they were presented out of 100 people. Results confirmed authors' supposition, as participants accepted a lower risk of death when probabilities of success and failure in the gamble were expressed as "Y deaths out of 1000" rather than as "X deaths out of 100."
Study 2 of the same research represented a further confirmation that the superficial way in which health risks are framed has the power to lead to inconsistent preferences. A new elicitation technique called "double-lottery method" was used, and the general population tested. The technique was a variation of the SG question in which risk was present in both parts, and participants had to answer to two questions; only one of those questions, anyway, was used to test the existence of the Ratio-bias effect, while the other served as a control question. For what concerned the main issue of interest, the scenario read,

Assume that you have been injured in a road accident. If you do not receive medical treatment you will experience situation X. There are two alternative treatments available, C and D. When treatment C is applied to 100 [1000] people, 1 [10] patient experience situation X and 99 [990] patients experience situation W. When treatment B is applied to 100 [1000] people, N patients experience situation X and (100-N) [1000- N] patients return to normal health in 3-4 days.
(Pinto-Prades et al., 2006, p.127)

Participants received either the version where ratios were framed in 100 or that where they were framed in 1000. The number N (and its complementary) was varied using a ping-pong technique as in Study 1 until indifference was reached. Authors' prediction (based on mathemathical calculations), was that participants would have accepted a lower degree of risk in the experimental conditions where the frequencies were expressed out of a 1000 denominator rather than when they were expressed out of a 100 denominator. Results supported authors' hypothesis, thus further corroborated the idea that the format in which a frequency is expressed (X out of 100 or 10X out of 1000) matters in terms of elicited health utilities.

The elicitation method was also at the basis of another effect that has been explained by means of the Ratio-bias phenomenon. Yamagishi (1994-a, 1994-b) referred about his and some colleagues unpublished study (Saito, Kawabata, and Yamagishi) where an interesting "response-range effect" was observed that could be attributed to the response ranges provided. In fact, results of the study, inspired by that conducted by Fischoff and MacGregor (reported in Fischoff, Slovic, and Litchenstein, 1982) showed, that when numerical estimates of certain social facts (e.g., the probability of being victimized by violent crimes) were required in frequency terms, radically different estimates were returned by participants who had to provide an estimation for a relatively small range (i.e., 'out of 100') and by those who instead had to provide an estimation for a relatively large range (i.e., 'out of 10,000') of population. In particular, the estimations elicited out of a population of 100 persons were generally lower than those elicited out of a population of 10,000 persons, an effect in line with the Ratio-bias. Authors' explanation for the phenomenon concerned with the idea that perception of riskiness depends highly on method of elicitation (Fischoff, Litchenstein, Slovic, Derby, and Keeney, 1981; Fischoff, Slovic, and Lichtenstein, 1982; Slovic, Kraus, and Covello, 1990). In particular, it referred to a common and underestimated bias in frequency judgments connected to the survey methodology, i.e., the fact that study participants infer normal tendencies in the reference population from the presentation of verbal response ranges (Schwarz and Hippler, 1987; Schwarz, Strack, Muller, and Chassein, 1988). Saito, Kawabata, and Yamagishi (quoted in Yamagishi, 1994-a, p. 652) argued that, "the presentation of the narrow (100) response range implicitly suggests that the least possible occurrence is one in 100 […]. Yet the use of the wide (10,000) range suggests that the least possible occurrence is one in 10,000, retaining much less lower frequency rates in concern". In Yamagishi (1994-a) the author claimed that such a response-range effect he could replicate was nothing else but a particular type of anchoring procedure. In particular, the anchoring procedure author referred to was an implicit one, as differently from the anchoring and adjustment process described by Kahneman and Tversky's (1974), no numerical expression had been provided as an anchor on the response scale. Instead, embedded clues from the range provided (1/100 or 1/10,000 in the proposed scenarios) were picked up spontaneously by individuals, who used them to ease the evaluative procedure. In Yamagishi (1994-a), participants were asked to provide an estimate of frequencies of death for 11 well-known causes in a sample of 10,000 people (i.e., "wide range" condition), in a sample of 100 people (i.e., "narrow range" condition), and to rank-order those lethal causes from most to least frequent, in three different sessions. In order to compare the reported evaluations in the two conditions differing for size range, the measure obtained in the wide condition were divided by 100. Two different analyses- aggregated level and individual subjects' level- of the same data obtained from 49 participants were performed. Significantly greater mean estimates of mortality rate were found in the narrow range condition (i.e., 29.06) than in the wide range one (i.e., 12.95), thus demonstrating a clear influence of the superficial way of expressing the response-range. Subjects' performance analysis showed a high consistence among subjective risk scales, in other words "the relationship that one cause was judged as more or less frequent than another was highly well preserved across methods of elicitation" (Yamagishi, 1994-b, p.

663). Nevertheless, these results must be taken with caution, considering that they were obtained from a within-subjects design for what concerned the ratio format presentation[35]. Comparable results, though, were obtained by Slovic, Monahan, and MacGregor (2000, Study 1 and Study 2 described here below).

Finally, elicitation of probability has been shown as influenced by Ratio-bias even in Slovic et al. studies (2000). Forensic psychologists and psychiatrists shown six case summaries of mental hospital patients were asked to evaluate the probability that the protagonist of the case would harm someone else within six months after discharge from the hospital. Also, they were required a judgment on the risk level posed by the person, an opinion on the need of monitoring the individual, and an evaluation of the likely necessity of rehospitalisation or involuntary outpatient commitment in the event of the patient's failure to comply with the prescribed medication. The variable manipulated within-subjects was the format of the response scale used to assess the probability of the question on the judgment of the risk level, namely that of interest for the present aims. Indeed, participants were randomly assigned to one of five conditions, LP (Large Probability), SP (Small Probability), LF (Large Frequency), SF (Small Frequency), or SF1000 (Small Frequency out of 1000). Results showed lower mean probability judgments for the scale with comparatively smaller than that with comparatively larger numbers, both for scales where points were labelled in percentages and in relative frequencies, showing a clear response-range effect. The same effect was replicated in Study 2 with the same scenarios, but an additional tutorial at the beginning of the session on how to make probability or relative frequency judgments. Overall responses aggregated across all six cases for each of the five experimental conditions showed that within relative frequency formats, a scale expressed in terms of 1000 rather than 100 individuals elicited larger frequency judgments than the same scale expressed in terms of 100 individuals. For instance, while 15.9% of the case summaries were considered at a 1/100 probability of harming someone else during the six months after discharge (Condition SF), more than double the number of such cases (i.e., 32.5%) were judged at a 10/1000 probability of harming someone (Condition SF1000), despite the two probabilities were in fact equivalent. The risk judgment on the probability that the patient would harm someone else followed the same trend; in other words, more people assessed the same risk as low when that had been evaluated on the SF scale (out of 100 people) than when it had been evaluated on the SF1000 scale, e.g. 81.2% vs. 76.6% for the 1% judged probability of harm, respectively.

[35] See observations in Khaneman (2003), reported at the beginning of Chapter 3.

SECTION 2.

When 1 in 10 is higher than 10 in 100: Results in favour of neglect of the numerator

The group-diffusion effect and the reference group effect

Opposite results to the Ratio-bias (or denominator neglect), thus indicating that people attend to and weigh the denominator more than the numerator, have been obtained by Yamaguchi (1998), when ratios used different superficial formats of expression. In two studies, author demonstrated that individuals feel illusory safer (more in control) in the presence of others than when they are alone, an effect that he dubbed the "*group-diffusion effect*". As the number of these "others" grew, the feeling of safety increased and people tended to assess risk at lower levels.

In fact, in Study 1 (Yamaguchi, 1998) Japanese female students evaluated risk as lower when the number of their peers exposed to the same risk increased. Participants presented with six hypothetical risk scenarios in the physical and financial domain, and asked to role-play to have part in them, were asked to estimate the amount of risk (in %) involved in each of the six situations. The independent variable manipulated between-subjects was the number of people potentially at risk, i.e., the participant alone, a small group (i.e., 10 people), or a large group (i.e., 100-1 million people). As a result, not only the group size showed a main significant effect on the perceived probability level, but also the two variables inverse relationship was obtained, since the declared level of probability decreased as the magnitude of the number of companions increased (as function of the number of risk companions). For example, one scenario ("Carcinogen 2" scenario) read:

The underground water supply in your area was recently found to be contaminated with carcinogens. It is estimated that it will take about five years before the effects of these carcinogens appear. It is also uncertain how many people will eventually get cancer from the contaminated water. There are 10 [100, 1 million] people including yourself [you are the only one] in three families who are drinking the contaminated water every day from wells in your area.

How likely do you think it is that you will develop a cancer?

(Yamaguchi, 1998, p. 128)

In this case, participants estimated, that their personal probability of getting a cancer due to water contamination through a carcinogen would be lower as more people drank the water on a daily basis. Obviously, their judgmental behaviour was irrational, as in fact the presence of other people in the same situation would have not reduced the protagonist's objective risk level.

In Study 2, results of Study 1 were further confirmed in a more realistic situation (i.e., laboratory setting), thus broadening their external validity. Female students participated individually by reading the description of an experiment involving electric shocks, in which the total number of prospective participants that would have communicated to them was varied in a way similar to that adopted in the preceding experiment. It could be one person only (i.e., the participant alone), a small group (i.e., 10 people), or a large group (i.e., 500 people). After reading the scenario, students were asked to estimate, in %, 1) the probability of suffering from the after-effects if they had taken part in the trial, 2) the severity of those effects, as well as 3) their willingness to participate into the experiment, all on 7-point Likert scales. Reasonably, and comprehensibly, eagerness to participate was constant across conditions, and generally low. Instead, the probability of suffering from after-effects decreased as the number of people involved in the experiment increased, and its trend was paralleled by that of the degree of intensity of those effects.

According to author, findings of both studies would show that individuals demonstrate an irrational tendency to feel safer, i.e., to judge a risk as lower, when in group than alone, a tendency of illusionary "safety in numbers" (Price and Mattews, 2009). Such feeling would be the result of a notion that is the product of the natural human propensity to group with other humans in order to face otherwise impossible-to-hand threats to their survival, or in order to better face those, as in case of, for instance, natural disasters. If the history of humankind is full of episodes where indeed the fact of being in a group has shown essential (or at least useful) for individual survival, the mere application of such notion to all situations appear notwithstanding irrational, as the presence of other individuals not always works as a benefit for ameliorating an individual condition. For example, in those circumstances where collaboration has no room to offer an advantage, people are not necessarily safer in a group than alone (see, for instance, the probability to get home safely on a plane in a bad storm is not function of the number of people flying on board, rather is dependent on the pilot's skills).

Such "group-diffusion effect", demonstrated on a Japanese sample, received further confirmation in Hong Kong by Ho and Leung (1998), and in the USA, albeit partially, by Chua, Yamaguchi, and Yates (unpublished, cited in Yamaguchi, Okumura, Chua, Morio, and Yates, 2008). The study in Hong-Kong aimed at testing several potential explanations for the effect, among others the "interdependence heuristic bias" and the "comfort hypothesis", both offered by Yamaguchi (1998). According to the "interdependence heuristic bias", the tendency of feeling safer in a group than alone would be the result, in collectivist cultures like the Japanese one, of people's attitude to see the group rather than the single person as the basic unit of survival. As a consequence of the effect of solidarity, and of a high concern for others in their society, collectivist people would perceive a lower degree of threat in presence of risk companions than alone. Nevertheless, the interdependence heuristic bias hypothesis has been disconfirmed by Ho and Leung (1998) who found no evidence that the effect was stronger for individuals evaluated as collectivists rather than individualists on an appropriate scale (the INDCOL, in Triandis, Bontempo, Villareal, Asai, and Lucca, 1988, translated in Chinese) intended to measure the collectivism orientation. Similarly, in the same study the "comfort hypothesis" arguing for a sense of security in presence of other people in a fearful situation did not receive support. Indeed, the same intensity for the bias was found when participants were provided with an alternative source of comfort (i.e., insurance) and when they were not.

A similarity between the group-diffusion effect and the so-called "*reference group effect explanation*" (Jenni and Loewenstein, 1997) can be noticed. Indeed, according to this latter account, in brief, individuals would feel greater distress for victims as "the reference group they are part of grows smaller" (Slovic, Fischhoff, and Lichtenstein, quoted in Small and Lowenstein, 2003, p. 6), namely they would feel more distress for a single victim than for a statistical life. "Identifiable victims represent the most highly concentrated distribution of risk (an n of n) because identifiable victims become, in effect, their own reference group. In contrast, a statistical life has a much larger denominator, because the risk is typically spread across a large population." (ibid), despite the value of life should be the same independently of the total number of individuals at risk. This "greater sensitivity to proportions than to absolute numbers of lives" (Small, Loewenstein, and Slovic, 2007, p. 144) obtained empirical confirmation in some subsequent studies (e.g., Fetherstonhaugh, Slovic, Johnson, and Friedrich, 1997; Baron, 1997).

From group diffusion to Ratio-bias

A link between Ratio-bias and group-diffusion effect has been traced by Price and Mattews (2009). In a series of four studies, authors first replicated Yamaguchi's finding (Exp. 1) and extended it to a positive outcome, thus ruling out the interdependence heuristic as explanation (Exp. 2), then made the group-diffusion effect disappear (Exp. 3), and finally argued Ratio-bias as the reverse of the group-diffusion effect.

In Study 1, a "bacteria scenario" (here following), or a "carcinogen scenario" were employed:

Imagine that you are one of N people eating at a restaurant. Afterward, you find out that you were exposed to a certain kind of bacteria in the food. Furthermore, medical experts say that people exposed to these bacteria have a P probability of becoming seriously ill as a result.

(Price and Mattews, 2009, p.439), [N (1, 10, 100, 1000); P (1%, 20%)]

Despite a variant of Yamaguchi's paradigm was used (within-subjects presentation of the scenarios, of the probability, and of the number of threat companions; non-numeric judgment on a 12-point scale rather than on a percentage scale), the typical group-diffusion effect was found. In fact, participants evaluated their chances of experiencing the negative event as lower when the number of people exposed to the risk increased.

In a further experiment presented in the study, both scenarios were slightly modified to include an explicit mention to the numerator of the ratio, that in Study 1 and 2 had instead to be deducted through a mathematical operation on the probability (%) to experience the negative outcome and the number of people exposed to the threat. It has to be noticed, that such mathematical operation was rather difficult because not everybody knows in similar cases, that a) the information (% and N) have to be integrated to return the personal probability of experiencing the negative outcome; and b) the mathematical rules necessary to integrate information. The general form of the new version of the scenario read,

Picture yourself as one of N people in a room. n of the N people in the room will be randomly selected to [win/lose] $ 50.

(Price and Mattews, 2009, p. 439), [N (10, 100, 1000), P (10%, 30%), n accordingly]

A null effect of the size of the group of probability companions was detected, namely an absence of the group-diffusion effect. Authors interpreted similar findings as due to the fact that participants could now focus on both the numerator and the denominator of the ratio they were required to judge.

In the final Experiment presented in their study, Price and Mattews demonstrated, as they had hypothesised, that a group-diffusion effect was present in the condition where only the denominator was made salient ("*denominator-only condition*"), while no effect was present when both numerator and denominator were made explicit (i.e., "*numerator-plus-denominator condition*"). A Ratio-bias effect was present instead when the numerator of the ratio was made salient (i.e., "*numerator-only condition*"). Nevertheless, it has to be noticed, for reasons which will be clarified at the beginning of next chapter), that in the latter condition the items indicated the number of people expected to be affected by the condition, but the exact number of the total people at risk was not specified, i.e., it was not clearly stated- since the scenario read: "several" people were exposed to the risk).

Other researchers apart from Price and Mattews had argued for, among different mechanisms, the role of saliency in creating the Ratio-bias (see Pinto-Prades, Martinez-Perez, and Abellán-Perpiñán, 2006), without nevertheless testing experimentally such supposition. In authors' words, "[...] the group-diffusion effect is related to the ratio-bias. Both effects occur when people make risks or likelihood judgments based on information presented as a ratio. The difference is that the group-diffusion effect occurs when the denominator of the relevant ratio is more salient than the numerator, while the ratio-bias occurs when the numerator is more salient than the denominator"(p. 436). Relative saliency of either of the two features in a ratio has been indicated as the cause generating either one or the other effect. In line with Epstein and colleagues' idea that "information can be presented so that participants attend to and weight either the denominator or the numerator more heavily in their likelihood judgments" (Price and Mattews, 2009, p. 445), Price and Mattews argued that there is not an intrinsic feature in the ratio that can be deemed responsible of one either of the two biases, rather one of the biases is generated in the conjunction of a specific ratio presentation and the problem examined.

The reason why people would focus on one feature rather than the other could be, as suggested by Pinto-Prades and colleagues (2006), the motivational concern. In their study, the numerator was the object of motivational concern as it was in that feature of the ratio that the numerical outcome of the medical treatment (success or failure) was communicated. The same idea had been claimed by Denes-Raj and Epstein (1994), to explain results obtained in the typical game of chance: individuals would have focused on the numerator since the red beans were of motivational concern as participants wanted to select them (gain outcome) or avoid them (loss outcome),

CHAPTER 3
Empirical studies

A manuscript partly based on this chapter is under review (2nd revision) in *Medical Decision Making* as: Pighin, S., Savadori, L., Barilli, E., Cremonesi, L., Ferrari, M., and Bonnefon, J.-F. "The '1 in X effect' on the subjective assessment of medical probabilities".

Rationale for research studies

The main rationale for the research was that of examining how numerical format affects perceived probability in single risk communications. The rationale stemmed from recent theories in the psychology of risk perception, that separate the role of experiential vs. analytical reasoning in the way individuals form perceptions of risk (see beginning of Chapter 2). Much evidence reviewed in Chapter 2 suggests that presentation formats have been interpreted as a way to convey an experiential vs. analytical processing of information (see, for instance, Denes-Raj, Epstein, and Cole, 1995; Slovic, Monahan, and MacGregor, 2000). However, some of the researches described in Section 1 of Chapter 2 could be criticised for ecological validity if they ought to be transferred, as they are, into the medical health domain. Most of the studies, indeed, presented people with the same probability expressed in two different formats, and asked them for a direct comparison between the two formulations. This method might compromise the ecological validity of the results, if one considers the domain of medical risk communication. Presumably, doctors rarely explain to patients that one treatment has, for instance, a 1 in 10 chance to succeed, whereas the other has a 10 in 100 chance to succeed. Using different ratio formats for two probabilities that are trivially the same, besides being in principle not advisable because the employment of different denominators in the ratios does not favour a sound comprehension and comparison of probability magnitudes (Burkell, 2004; Paling, 2003), would surely be an odd communication move. A more realistic situation, instead, is one where the practitioner chooses a specific format to express the equal chance of success of two different treatments, and coherently sticks to it for the description of the probability of both cures. Moreover, there are even occasions when patients are presented with information relating to a single treatment or with the probability of an adverse outcome, rather than being required the comparison between two or more options. In fact, frequently they are required to evaluate clinical risks having little contextual knowledge to support consistent risk perceptions. Very often, risks cannot be compared against each other, and they have to be evaluated on an absolute scale (for an exhaustive review concerning the differences between joint and separate evaluations and on how these two evaluation modes differently affect preferences, see Lichtenstein and Slovic, 2006).

In both the case of two different treatments, and in that of a single one/ of the probability of an adverse outcome, knowing whether there are any differences in the subjective probability assessment of a health-related outcome as a function of the magnitude of the numbers employed in the ratio expressing its probability can be useful for easing the choice between them.

Methodological considerations

The present research focuses on a relatively less evaluable context where only separate evaluations are possible. From the methodological point of view, two main considerations have been made on the basis of both the aim just delineated and the literature described in Chapter 2. First, in accordance with what expressed above on the ecological validity of studies in medical risk communication, it was reasoned that experiments aimed at assessing effects on perception of ratio formats should use between, rather than within-subjects, evaluations of the ratio formats. That happened in the studies on group-diffusion effect described in Section 2 of the previous chapter; nevertheless, differently from those, both numerator and denominator of a ratio should now be made explicit. Additionally, a between- rather than within-subjects design for the ratio formats under study was judged as more appropriate, in order to avoid that participants could disclose the equality of the two formulations[36]. Thus, unlike in Price and Mattews (2009, Study 1-4, still in Section 2) in our studies different participants would have assessed different ratio formats.

The research work included twelve main studies, description of which has been organized in four distinct sections.

Preliminary brief characterisation of the basic constructs investigated

Focus of investigation regards the effects of ratio format presentation (smaller/larger numbers in the ratios) on perceived probability. Nevertheless, some other related measures normally assessed in the risk communication literature will be examined as well, in particular worrisomeness for the outcome probability, the perceived severity of the outcome, behavioural intentions, and numeracy level. In order to make reading as smooth as possible, it was judged convenient to briefly delineate here, the general utility of those measures, and to specify the rationale for their use in the present research. Here following, the concepts tapped by those constructs will be briefly described, together with the modality normally adopted to elicit them in the decision making literature, so to make any change adopted in the present studies evident to the reader.

[36] As affirmed by Kahneman (2003, p. 712), "Intuitive judgments and preferences are therefore best studied in between-participants designs and in short experiments that provide little information about the experimenter's aims. Within-participant designs with multiple trials should be avoided because they encourage the participants to search for consistent strategies to deal with the task. Within-participant factorial designs are particularly undesirable because they provide an unmistakable cue that any factor that is varied systematically must be relevant to the target attribute (Kahneman & Frederick, 2002)."

1. Perceived probability of an outcome

The weight of a risk can be described as the severity of the possible harm that might occur from a given activity or event multiplied by the probability of the harm occurring (see, for instance, Savadori and Rubaltelli, 2008). This can be expressed by the following equation:

(weight of risk) = (severity of harm) x (probability of occurrence)

Thus, people's judgment on the level of perceived risk (i.e., the weight of the risk, e.g., "In your opinion, how risky is X?") will not only enclose the subjective level of perceived probability of the event occurring ("In your opinion, how likely is that X happens?"), but also the personal assessment of the gravity of that outcome ("How severe would X be?")- for a more detailed description of the severity construct, please see point 3. If the objective is to measure perceived probability in terms of risk magnitude estimation, thus, when the outcome is rather severe (e.g., Down syndrome), it is preferable to focus the question on probability rather than risk in order to properly tap one rather than the other concept (i.e., "In your opinion, the probability of X happening is___?", rather than "In your opinion, the riskiness of X happening is ___"). This expedient will be adopted in the studies of the present dissertation.

2. Perceived worry for an outcome

Studies of risk perception normally differentiate between cognitive and emotional reactions to threats. Normally, the experience of worry does not correspond to that of intellectual judgment.

"With the word 'worry' is denoted preoccupation with thoughts about uncertain and unpleasant events" (Sjoberg, 1998, p. 85). Worry is normally assessed by means of an item like, "To what extend are you concerned about X?" Worry ratings give a different picture compared with risk judgments. In line with what affirmed by Sjoberg, namely that "It can often be a good idea to use both types of dimensions [probability and worrisomeness perceptions] in a study on risk perception" (Sjoberg, 1998, p. 92), both dimensions will be assessed in the first study of this thesis.

3. Perceived severity of an outcome

The severity of a risk can be defined as the perception of the badness of the outcome in the case a hazard happens for sure. Generally speaking, higher risk activities have higher probabilities of more severe harm, while lower risk activities have lower probabilities of less severe harm. For example, Russian roulette would be considered a very high risk activity (perhaps a 1 in 5 chance of death) while reading a book would be considered a very low risk activity (perhaps a 1 in a million chance of a paper cut). Nevertheless, like benefits, harms are also largely subjective, and their severity can be assessed differently by different people. In other words, harm that one person considers extremely severe another person might not care about very much at all. In case the two persons are asked an evaluation of the same probability of that harm happening, two profoundly different judgments could be given by them. It is expected that the first person gives a higher risk judgment than the second, because the higher perceived riskiness of that harm reflects a higher severity assessment. Thus, when eliciting a probability judgement on a negative event or outcome, it is important to disentangle the person's probability evaluation (i.e., perceived *"probability of occurrence*", in the weighting function above) from her assessment of the gravity of the outcome at stake (i.e., perceived "*severity of harm*", in the weighting function above). Perceived severity of the outcome at stake will be assessed in the first study of this thesis to separately account for that variable. Nevertheless, for the specific outcome investigated in this research (i.e., Down syndrome) one more aspect related to severity has to be considered.

The extreme case of probability neglect

One of the consequence of the "risk as feelings" hypothesis (Loewenstein, Weber, Hsee, and Welch, 2001) according to which people are more insensitive to probability variations for emotional and vivid outcomes than for pallid outcomes, is that the impact of probability depends strongly on the nature of the outcome. In other words, the probability weighting function simplified above would be flatter (i.e., overweighting would occur, especially for small probabilities) for vivid outcomes evoking emotions than for pallid outcomes. Taken to the extreme, such consequence could result in extreme insensitivity to probability, "when the adverse event carries sharp and strong affective meaning, as in the case with a lottery jackpot or a cancer. In such situations, variation in probability often carries too little weight" (Slovic, 2004, pp. 13-14). Similarly, but in a positive domain, as Lowenstein et al. (2001) observed, one's images and feelings toward winning the lottery are likely to be similar whether the probability is one in 10 million or 1 in 10,000. Authors further noted that responses to uncertain situations appear to have an all or none characteristic that is sensitive to the *possibility* rather than the *probability* of strong positive or negative consequences, causing very small probabilities to carry great weight.

Support for these arguments can be found even in research of Rottenstreich and Hsee (2001) that showed that, if the potential outcome of a gamble is emotionally powerful, its attractiveness or unattractiveness is relatively insensitive to changes in probability as great as from .99 to .01. Authors found, not only that people were insensitive to probability variations, but also that such insensitivity depended on the emotional impact of the associated outcomes. That basic point has received its clearest empirical confirmation in a striking study of people's willingness to pay to avoid electric shocks. The central purpose of the study was to test the relevance of probability in "affect rich" decisions. The experiment that is relevant to the present case (i.e., Study 3) attempted to see whether varying the probability of harm would matter more, or less, in settings triggering strong emotions than in settings that seem relatively emotion-free. In the "strong emotion" setting, participants were asked to imagine that they would participate in an experiment involving some chance of a "short, painful, but not dangerous electric shock" (Rottenstreich and Hsee, 2001, p. 188). In the relatively emotion-free setting, participants were told that the experiment entailed some chance of a $20 penalty. Participants were asked to say how much they would be willing to pay to avoid participating in the relevant experiment. In a similar way, a measure of the Willingness to Pay to avoid a risk was raised to calculate the Willingness to Reduce various fatality risks (Jones-Lee, Hammerton, and Philips, 1985) or that of preventing non-fatal road injuries (Jones-Lee, Loomes, and Philips, 1995), or of illness from eating oysters, for instance (Lin and Milon, 1995). In Rottenstreich and Hsee's study, some participants were told that there was a 1% chance of receiving the bad outcome (either the $20 loss or the electric shock); others were told that the chance was 99%, and still others were told that the chance was 100%. The central result was that variations in probability affected those facing the relatively emotion-free injury, the $20 penalty, far more than they affected people facing the more emotionally evocative outcome of an electric shock. For the cash penalty, the difference between the median payment for a 1% chance and the median payment for a 99% chance was predictably large and indeed consistent with the standard model: $1 to avoid a 1% chance, and $18 to avoid a 99% chance. For the electric shock, by contrast, the difference in probability made little difference to median willingness to pay: $7 to avoid a 1% chance, and $10 to avoid a 99% chance. Thus, apparently people pay a significant amount to avoid a small probability of an affect-laden hazard, and the amount that they pay does not vary greatly with changes in probability. In a similar fashion, in the health-related field. Levy and Baron (2005) examined people's assessment of badness of 56 different medical risks varying in the probability from a 1 to 100%. Participants' judgments showed a complete independence from the amount of probability stated, while they depended entirely on the outcome.

Sunstein (2003) argued that that insensitivity, which he dubbed "probability neglect," explains overreaction to certain rare but emotionally powerful events such as terrorist acts. In practice, as a result of probability neglect people are often much more concerned about risks from terrorism than about statistically greater risks that they confront in ordinary life. That would happen because people tend to focus on the badness of the outcome, rather than on the probability that the outcome will occur, when strong emotions are involved in a judgment. In these conditions, people "[...] are not closely attuned to the probability that harm will occur."(Sunstein, 2003, p. 62). Among the studies described by Sunstein, a relevant one is that of Weinstein, Sandman, and Hallman (1998), where two conditions were compared, a high and a low outrage condition.

Results showed that a large difference in statistical seriousness had no effect in the "high outrage" condition, with people responding the same way to a risk of 1 in 100,000 as to a risk of 1 in 1,000,000. Even when the statistical risk was identical in the high outrage (nuclear waste) and low outrage (radon) cases, participants in the nuclear waste case reported a much greater perceived threat and a much higher intention to act to reduce that threat. What is more important for the present case, Sunstein (2003) argued "[…] that when emotions are intense, calculation is less likely to occur, or at least that form of calculation that involves assessment of risks in terms of not only the magnitude but also the probability of the outcome" (p. 66- 67). We will return to this point in Study 1.1.

4. Numeracy, or on "collective statistical illiteracy"

For an educated citizenship in a modern democracy, statistical thinking would be as indispensable as reading and writing (attributed to H.G. Wells, quoted in Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, and Woloshin, 2007).

How large is a 1 in 250 risk? People might view an expression like the one just presented without being able to translate it into a meaningful representation. The reasons for this phenomenon are mainly two.

First of all, individuals experience difficulties in sensing numbers, and these difficulties become even greater when people have to grasp those that are not integer. Much evidence exists in fact in the developmental literature, proving that fractions and other related ratio concepts (rational numbers, decimals, proportions, and part–whole concepts) are especially difficult to learn (e.g., Hecht, Close, and Santisi, 2003; Reyna, 1991). When these numerical expressions are used to convey the probability of an uncertain outcome occurring rather than a fix partition of a quantity, people's difficulties are even furthermore increased due to the non intuitive ideas of probability and randomness. Indeed, several findings have showed that probabilities attached to outcomes are poorly comprehended by children and adolescents (e.g., Hoemann and Ross, 1982; Piaget and Inhelder, 1975; Reyna and Brainerd, 1991, 1994; Siegler, 1981) as well as by adults (i.e., see the entire heuristic and biases approach). Structural as well as evolutional accounts have been proposed to explain laypeople's low understanding of fractions, partitions, and probabilistic concepts. In sum, individuals would not be able to translate a probability expression like the one stated above into some meaningful concept because of their poor comprehension of numbers and the concept of uncertainty. Some authors have argued, that as a consequence of such poor comprehension laypeople face difficult and stressful decisions, even in health-care (see, for instance, Estrada, Martin-Hrynlewicz, Peek, Collins, and Byrd, 2004; Rothman, Housam, Weiss, Davis, Gregory, Gebretsadik, et al., 2006).

Secondly, people’s hard times in sensing numbers can be partly ascribed to the educational system of our societies. A widespread innumeracy– “an inability to deal comfortably with the fundamental notions of numbers and chance” (Paulos, quoted in Hertwig, Zangerl, ,Biedert, and Margraf, 2008, p.3) - had been diagnosed long ago. Despite even a raise of requirements for numeracy, performance of 12th graders has not changed in decades (see NAEP, for instance, Lee, Grigg, and Dion, 2007). Indeed, quite recent data from the American National Adult Literacy Survey (NALS) showed, that about half of the American population lacks the minimal mathematical abilities necessary to use numbers inserted in printed materials or to perform a numerical task requiring at least two sequential steps (Kirsch, Jungeblut, Jenkins, and Kolstad, reported in Reyna, Nelson, Han, and Dieckmann, 2009). Data have been further confirmed in a subsequent assessment of the nation’s literacy (NAAL, 2003), estimating around 36% of the adult population (i.e., something like 93 million people) at a below-basic or basic level of performance, with quantitative items eliciting the worst results (Kutner, Greenberg, Jin, and Paulsen, also researches reported in Reyna, Nelson, Han, and Dieckmann, 2009). The picture gets even more worrisome considering that some vulnerable groups (non-native English speakers, poor people, older or less educated individuals) that have been identified as those more in need of a public health-service, have also been estimated as those in fact performing as the worst in those quantitative operations. The situation appears slightly better in Europe than in the U.S., as the gap between lower and higher educated people are smaller than in North America (Galesic and Garcia-Retamero, in press). Nevertheless, the level is still too low to ensure a sound comprehension of risk messages, as for instance a recent research on a representative German sample showed that only the 68.5% of items assessing the abilities required to understand numeric data in messages could be correctly answered (Galesic and Garcia-Retamero, 2010).

Both rationales reported here for people’s low ability in translating probability expressions into meaningful representations have evidenced the central role that individuals’ level of numeracy- namely people’s “facility with basic probability and numerical concepts” (Schwartz, Woloshin, Black, and Welch, 1997, p. 966)- has in explaining possible problems associated with risk perception. While even researchers have long recognized the importance of literacy for making informed health decisions, the same cannot be stated for numeracy (Reyna, Nelson, Han, and Dieckmann, 2009) despite its evident importance. Indeed, the acknowledgement of this construct’s relevance is only a relatively recent thing, especially in the decision making literature.

Subsumed by some authors within the broader concept of literacy, numeracy can be equated to quantitative literacy (Lipkus and Peters, 2009) as it encompasses the possession of those basic mathematical skills necessary to comprehend and use numerical information expressed in texts or displayed in charts, graphs, or tables. As such it has been defined as “the ability to locate numbers within graphs, charts, prose texts; and to perform appropriate arithmetical operations on text-based quantitative data" (Bernhardt, Brownfiled, and Parker, p. 6, quoted in Reyna et al., 2009). However, as Reyna and colleagues have highlighted, numeracy has to be distinguished from the broader concept of health numeracy (for a comprehensive and detailed review of differences between the two constructs, see that source), in that the latter "is not simply the ability to understand numbers but rather to apply numbers and quantitative reasoning skills in order to access health care, engage in medical treatment, and make informed health decisions" (Reyna et al., 2009, p. 947)[37].

In the studies reported in the present dissertation, the decision to adopt the numeracy instead than the health numeracy construct relied on the consideration, that despite given the domain of study (i.e., a health condition) health numeracy could have probably been more appropriate than numeracy, the latter had more often been object of analysis in researches in decision making than the former. As a corollary of such an extensive use, there was a relatively large number of evidence available for numeracy, while the opposite was true for what concerned health numeracy. That evidence provided, beside relevant instances of implementation of the construct assessment, a good basis for theoretical considerations. Furthermore, numeracy level had also been related with several individual and contextual features definitely relevant for the issue at study.

Indeed, it has been demonstrated that those who are highly numerate have more ability to retrieve and use appropriate numerical principles, tend to be better at reading graphical aids (e.g., Zigmund-Fisher, Smith, Ubel, and Fagerlin, 2007), tend to rely more on numeric rather than verbal risk information from physicians (e.g., Black, Nease, and Toteson, 1995; Gurmankin, Baron, and Armstrong, 2004), and what is more relevant for present aims, tend to be less influenced by presentation format effects, like framing (Reyna, Nelson, Han, and Dieckmann, 2009). By contrast, those who are less numerate would perceive higher degrees of

[37] Among other skills, health numeracy includes indeed even the ability to assess the magnitude of risks, compare likelihood values, and "understand decimals, fractions, percentages, probabilities, and frequencies, as these are the formats in which risk and benefit information is most often presented […]" (Reyna et al., 2009, p.946). As authors emphasized, though, these abilities are not possessed by everyone, but only by those who can be classified as at the highest level of numeracy, i.e. that encompassing "[…] basic logic and quantitative reasoning skills, knowing when and how to perform multistep operations, and an understanding of ratio concepts, notably fractions, proportions, percentages, and probabilities […]" (p. 945), Such level, though, is rarely achieved even by highly educated people (e.g., those with a University degree), or experts (e.g., health professionals). In a study of Estrada, Barnes, Collins, and Byrd (2004), for instance, a definitely low degree of ability with numbers was showed even in health personnel (e.g., medical students, nurses, doctors). Findings of that survey represented the practical demonstration that the education level is a misleading proxy for individuals' numeracy ability, as "educational attainment does not ensure grade-level skills, and this is particularly true for mathematical skills […]" (Reyna et al., 2009, p. 948). In a similar way, Dehaene (1997) had already indicated that education attainment is not the proper factor to define numeracy: despite the fact children spend a significant amount of time learning the methods of mathematics at school, they might nevertheless not really comprehend how to apply those rules even in adulthood. Hence, numeracy has been claimed as an individual trait, as its degree has been shown to vary substantially across individuals (e.g., Lipkus, Samsa, and Rimer, 2001; Peters, Västfjäll, Slovic, Mertz, Mazzocco, and Dickert, 2006).

risk (i.e., overestimation of a probability), and would also lack a clear affective understanding of numbers (Peters, Västfjäll, Slovic, Mertz, Mazzocco, and Dickert, 2006; but see Reyna and Brainerd, 2008). This lack of numerical skills would represent for them a crucial deficit, resulting in reduced medication compliance, difficult access to treatments, impaired risk communication, and poor medical outcomes (Reyna and Brainerd, 2007). Low numeracy ability has also been associated with greater susceptibility to extraneous factors, like effects of mood, or the way information is presented, as well as biases in judgment and decision making (Reyna et al., 2009, see also paragraph 1.1.3 in Chapter 2 for a more detailed description). Among systematic tendencies, even denominator neglect has been documented (see Peters et al., 2006), showing that numeracy affected participants' performance in a non-optimal Ratio-bias-type study involving the classic urn and balls game of chance.

For all the reasons just mentioned , it was expected that our study participants' numeracy level could influence the degree to which they would have been biased by the variation of superficial details like the magnitude of the numbers employed in ratios used to express probability.

For what concerns the choice of the inventories to assess the numeracy construct, both an objective and a subjective scale were used in the studies. Distinction between the two types of scales has been made in the literature based on the perspective of assessment for the construct. While in the case of objective scales people's level of numeracy is assessed from "the outside" by means of tests calculating their performance, in the second case (i.e., subjective inventory) self-assessment of the level of confidence in numerical ability (i.e., subjective evaluation) takes place[38].

As objective scale, the one ideated by Lipkus, Samsa, and Rimer (2001) has been chosen; seemingly, the inventory is seemingly one of the most used in decision-making research, likely because its devisers expressly linked performance on such a scale to the way people perceive risk. Lipkus et al.'s scale (2001) has been developed on the basis of the scale of Schwartz, Woloshin, Black, and Welch (1997), a three-item inventory that tested 1) the understanding of the chance concept ("Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads? X times out of 1,000"), as well as 2) people's ability to convert a percentage like 1% into a proportion, namely into 10 in 1000 ("In the BIG BUCKS LOTTERY, the chance of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1,000 people each buy a single ticket to BIG BUCKS? __ person(s) out of 1,000"), and finally, 3) the ability to convert a proportion like 1 in 1000 into a percentage, namely into 0.1% ("In ACME PUBLISHING SWEEPSAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING

[38] Yet, one more distinction in objective scales has been made depending on whether they focus on numbers rather than on a situation or a specific disease, with the first type of scale dubbed "disease-general", while the second "disease-specific" (Reyna et al., 2009).

SWEEPSAKES win a car? ___%."). Beside a minor change to Schwartz and colleagues' three items (dice instead than lottery scenarios), Lipkus and others' scale further included eight questions referring to the health sector domain (but to non specified diseases/infections). Such expanded scale measures how well individuals can perform easy mathematical operations on risk magnitudes using proportions and percentages, how well they can convert percentages to proportions and vice-versa, or convert probabilities to proportions (for the complete list of items see Study 1.1. of Section 1, in which the scale has been used).

As a subjective measure, instead, the Subjective Numeracy Scale (SNS) of Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, and Smith (2007)- validated by Zikmund-Fisher, Smith, Ubel, and Fagerlin (2007)- has been used, see Study 2.2 (Section 2). Indeed, recently, some subjective measures have been proposed as valid substitutes of objective numeracy inventories, in that they do not contain mathematical questions and have no right/wrong answers. They showed to be good proxies for the objective concept of numeracy (see introduction to Study 2.2.)

SECTION 1, or
A new bizarre systematic effect

In the present section, two studies are exposed. In Study 1 it will be showed that the probability of a clinical condition was subjectively perceived as higher and more worrisome when expressed as “1 in 200” than when expressed as “5 in 1000” by women who were pregnant or had just given birth to a child. In Study 2, the effect was generalized to both genders in an adult population.

Study 1.1

Goals

Study 1 was intended to investigate if a “1 in 200” instead of “5 in 1000” format used to convey the same .005 medical risk affected patients’ subjective assessment of that risk. According to the literature previously reviewed, two antithetic hypotheses were putted forward:

> **Hypothesis 1: According to the theories reviewed in Section 1 of Chapter 2 (suggesting that, when evaluating probability conveyed through a ratio, people show a tendency to neglect denominators and to judge on the basis of the magnitude of numerators), a higher magnitude perception should be observed for a probability expressed as “5 in 1000” than as “1 in 200” even in single presentation conditions.**
>
> **Hypothesis 2: According to the group-diffusion effect described in Section 2 of Chapter 2 (suggesting that people neglect numerators, rather than denominators, when they evaluate a probability conveyed through a ratio), a higher magnitude perception should be expected for a probability expressed as “1 in 200” than as “5 in 1000” when both terms (i.e., numerator and denominator) are explicit.**

It was hard to lean toward one of those hypotheses rather than the other, given that, as affirmed in the description of the rationale for the studies, neither experimental designs and material employed in one group of theories nor in the other’s exactly reproduced the situation at study in the present dissertation. Nevertheless, following data of the literature on numerical cognition (see studies on the difference between encoding of small and large numbers- but integer ones only, for instance that on Weber’s law (i.e., Fechner, 1860), intuitively it was judged more likely that individuals would assign higher evaluations of probability to ratios employing larger (i.e., 5 in 1000) rather than smaller (i.e., 1 in 200) numbers, even in single presentation conditions. It was reasoned, that very likely individuals would have given an assessment of the magnitude of the ratio by looking at the magnitude conveyed of both terms (i.e., the numerator and the denominator): the higher the numbers were, the higher the perceived probability for the event. Thus, Hypothesis 1 was thought of as the most likely one.

The same hypotheses (1 and 2), and the same logic adopted for perceived probability were deemed as valid for the degree of worry expressed for the outcome. Instead, it was thought, that severity should have not been affected by a format effect, since the severity of an outcome is a judgment of an event as already occurred (hence, no probability is involved, but certainty).

Finally, it was reasoned of the possible role in perceived probability of people's facility with numbers and probability concepts. In accordance with the literature demonstrating that those who are highly numerate tend to be less biased by presentation format effects (e.g., framing, see Peters and Levin, 2008), it was expected that those showing less competence in performing numerical operations and dealing with the concept of uncertainty could also be those showing a higher degree of influence of the ratio format in which the probability was expressed (smaller/larger numbers in the numerator and denominator of the ratio)- see Hypothesis 3, below. Indeed, those that would be classified as low numerates were expected not to possess the necessary abilities to draw a meaning from the probability information communicated through the ratio format, because of their difficulties with numbers, proportions, and more generally with the concept of probability. As a consequence of that, it was likely that they would have tended in a larger measure than the other participants to rely on a heuristic-type of processing of the information, one for instance where superficial details like the absolute magnitude of the numbers employed would be determinant for the probability assessment. On the contrary, thanks to their higher ability with numbers and the idea of probability, high numerates were expected to be less (or not at all) prone to the bias under study, likely because of the supposed higher capacity to draw a meaning from the probability information communicated. Thus, the hypothesis on numeracy was:

Hypothesis 3: More numerate individuals should be less susceptible to numerical format effects in perceived probability (i.e., 1 in 200 perceived as conveying a smaller/larger probability than 5 in 1000).

Method

Participants. A total of a 63 women, patients of a hospital in North Italy, took part in the study voluntarily. The large majority of them (74.6%, n = 47) were recruited at an out-patients' (gynaecologist), some of them (23.8%, n = 15) at the Sterility Centre of the hospital, while one person was recruited in a maternity ward of the hospital. In the present study and in the others described in this dissertation (where not specified differently) individuals did not receive reward for participation.

Mean age was 33.63 years (*SD* = 4.71), varying between a minimum of 22 and a maximum of 45 (one participant did not disclose her age). Most of participants had completed high school (*n* = 33, 53%) or had already a University degree (*n* = 21, 34%), and only few of them (*n* = 8%) had completed the lowest education level in Italy. Almost all women (*n* = 61, 95.2%) had an occupation at the moment of survey completion, and only one was unemployed (one person did not answer to this question).

The large majority of women (87.1%, *n* = 54) were pregnant at the moment of survey completion, 8 (12.9%) had just given birth (one person did not disclose this information). The thirty-nine percent of them (*n* = 24) had already one or more children, while the remaining 60.7% (*n* = 37) did not have children (2 people did not answer to this question).

Design and material. Participants read a scenario (for the exact wording, please see the Appendix[39]) where they were exhorted to imagine that one of their friends had just bought a trip to an exotic country. They were informed on the statistical risk of contracting malaria in that country: the probability was expressed through a numerical ratio. The single independent variable manipulated between-subjects in the study was the format in which the value of the probability could be expressed. Indeed, its numerator and denominator were numbers that could be comparatively smaller or larger. In the first case (i.e., smaller numbers), the ratio used was 1 in 200[40], while in the second case (i.e., larger numbers), 5 in 1000 was used.

Four dependent variables were assessed: the subjective magnitude of the probability of infection, the degree to which this information would be found worrisome by participants, the perceived severity of the health-related outcome at stake (i.e., malaria), and the degree of participants' numeracy.

Perceived probability was measured by asking a personal assessment of the friend's probability of contracting malaria going to Kenya on a seven-point Likert scale anchored to "negligible" and "almost certain"- all points in this and the other scales employed in the present study were verbally labelled. Similarly, worry for infection (in the hypothetical situation that the participant would go to Kenya) was measured on a 7-point Likert scale anchored to "not at all worried" and "very much worried". Finally, perceived severity of malaria was measured on a 7-point Likert scale anchored to "very severe" and "not at all severe"- this question had to be reverse scored for analysis.

[39]The experimental material of all the studies described in this dissertation (where not specified differently) is reported in the Appendix.

[40]1 in 200 is the approximated and rounded frequency of fetuses with Down syndrome to normal fetuses at 16 weeks of pregnancy for a 36-year- old woman (Hook, Cross, and Schreinemachers, 1983).

On a separate page, participants read the 11-item inventory measuring numeracy, substantially corresponding to that of Lipkus, Samsa, and Rimer (2001) apart from minor changes not altering its sense[41].

Procedure. Each participant was randomly assigned to one of the two conditions (numbers in the ratio format: smaller/larger) constructed to follow the experimental design. A paper version of the questionnaire was handed out personally to each individual by the experimenter after the participant had read and signed an informed consent form. The completed copies were collected by the experimenter who also debriefed participants on the study aims. Individuals could complete the questionnaire at their pace, without any time constraints.

Ethical approval. The study required an ethical approval because it was addressed to a special population (i.e., patients, many of them pregnant women). Under the requirement of the Clinical Ethical Committee, the questionnaire did not include any sensitive question that participants might feel unconformable answering. Indeed, the scenario employed in the questionnaire should have included the doctor-patient communication of the probability of having a child with Down syndrome, rather than the probability of contracting malaria in an exotic country. However, Down syndrome had to be substituted with an outcome not directly related to the current health condition of participants, in such a way to possibly avoid anxiety in the already sensitive individuals. For the same reason, the risk mentioned in the scenario had to attributed to a participant's friend instead than to the participant herself.

Anonymity and confidentiality in the treatment of the data gathered were ensured to participants.

Statistics. Most of the analyses of the present study and of those described in this dissertation were run by means of widely known analytical tools available in SPSS.

Results

Perceived probability, worry, and severity. Figure 1 displays the mean values of the three dependent measures as a function of the ratio format used to communicate the probability in the two experimental conditions of the present study. The visual inspection of Figure 1 immediately suggests that changing the ratio format changed patients' subjective assessments of probability, as well as the degree they found it worrisome, while not affecting the subjective severity of the negative event (i.e., malaria).

[41]A as question number 5 of the original expanded numeracy scale included two subquestions (i.e. A: how many people […] out of 100?, and B: how many people […] out of 1000?), we preferred to split it into 2 separate questions to increase its clarity.

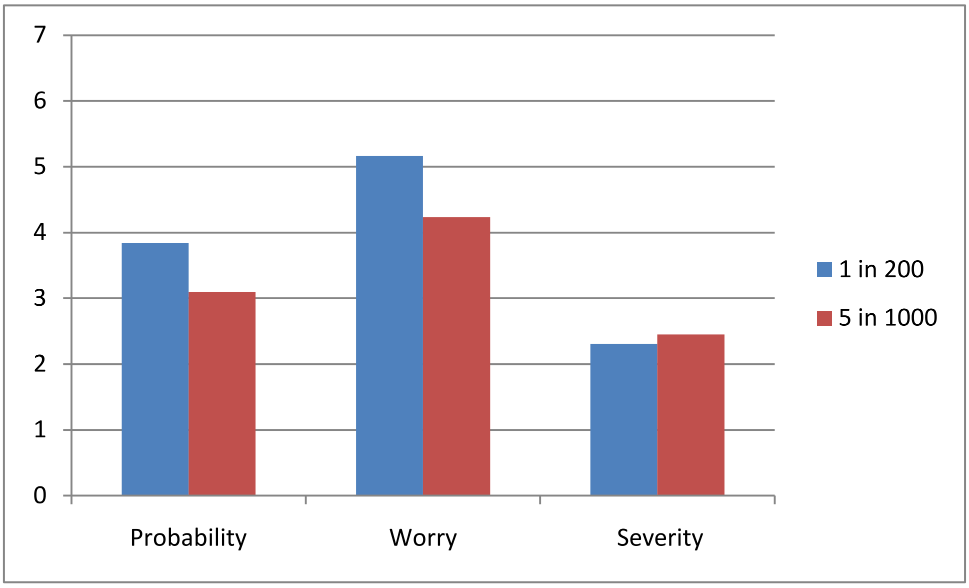


***Fig. 1.* Subjective ratings of probability, severity, and worry for the outcome, as a function of ratio format in the two experimental conditions**

This visual impression was confirmed by a multivariate analysis of variance[42]. Ratio format had a global impact on the set of dependent measures, $F(3,59) = 2.84$, $p = .045$, $\eta_p^2 = .13$. This global impact, though, was the result of a localized impact on the probability and worry measures, rather than on the severity measure. The subjective magnitude of the probability was 3.84 ($SD = 1.08$) when it was phrased as “1 in 200”, and 3.10 ($SD = 1.27$) when it was phrased as “5 in 1000”, a significant difference, $F(1,61) = 6.31$, $p = .02$, $\eta_p^2 = .09$. A higher perceived probability was detected in the format using smaller instead than larger numbers at the numerator and at the denominator, contrary to what expected. Similarly, the degree each participants would worry about the risk was 5.16 ($SD = 1.22$) when the probability was phrased as “1 in 200”, and 4.23 ($SD = 1.50$) when it was phrased as “5 in 1000”, a significant difference, $F(1,61) = 7.32$, $p = .01$, $\eta_p^2 = .11$. However, the subjective severity of malaria was the same in the two conditions ($M = 2.31$, $SD = .82$ vs. $M = 2.45$, $SD = .72$), $F(1\ 61) = 0.51$, $p = .48$, $\eta_p^2 = .01$. Thus, results of this first part of the analyses supported Hypothesis 2, but rejected Hypothesis 1, contrary to our expectations.

To further establish that the increased worry expressed by the participants in the “1 in 200”condition was mediated by an increase in subjective probability, a path analysis by means of a series of three regressions was conducted. The ratio format (dummy coded, 0 standing for “5 in 1000”) was a significant predictor of how much the risk was worrisome, as shown by the standardized regression coefficient $\beta = .33$, $t = 2.7$, $p = .01$, and a significant predictor of

[42] Except when differently stated, all tests used a critical alpha value of .05, and the p-value presented is for a two-tailed test.

subjective probability, β = .31, t = 2.5, p= .01. Subjective probability also was a significant predictor of how worrisome the risk was, β = .61, t = 6.1, p < .001. When ratio format and subjective probability were simultaneously entered as predictors of how worrisome the risk was, subjective probability remained a significant predictor, β = .57, t = 1.5, p = .15. A Sobel test (value of which was 2.3, p = .02) confirmed that the contribution of the ratio format dropped significantly when subjective probability was entered into the regression.

The moderating role of numeracy. The overall percentage of correct responses to each of the 11 items of the Numeracy scale is presented in Table 1. Performance on single items was comparable to that found for a sample of adult people in Lipkus, Samsa, and Rimer (2001, Exp. 3) and highlighted the difficulty with some simple questions testing rather basic abilities required in important decisions (see, for instance, items 2 and 4).

The mean numeracy score was 7.11 out of 11 possible (range 0-11). Because the distribution was highly skewed[43] (see Fig. 2), a median split (*Mdn* = 8) was performed that allowed to form two groups, one including those participants (n = 32, "*high numerates*" from now on) who had more facility with numbers and calculations (8 or more correct items on the scale) and the other (n = 31, "*low numerates*" from now on) those who were less able in that respect (1- 7 correct items). When the two groups were analysed separately for the main dependent variable (i.e., probability perception) according to the ratio format[44], Mann-Whitney tests[45] returned a non significant difference in the two experimental conditions for low numerates (Z= -.95, p = .34), but a significant difference for high numerates (Z= -2.54, p = .01). This finding, in contrast with Hypothesis 3, meant that, only among the high numerate participants the 1 in 200 format was perceived as indicating a significantly higher probability than the 5 in 1000 format, while instead low numerates did not perceive any difference between the two formats.

[43] The distribution of the Numeracy scale was not normal (Kolmogorov-Smirnov statistic = $.154_{(63)}$, p = .001).
[44] This procedure is similar to that adopted in Peters, Västfjäll, Slovic, Mertz, Mazzocco, and Dickert (2006).
[45] Non parametric tests were employed since the numerosities in each of the groups were too low (n < 30) to apply parametric tests.

Table 1
Overall participants' performance to each of the items composing the Numeracy inventory employed, in order of decreasing accuracy

Item	Percentage correct
8. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100? |______|	81
9. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000? |______|	79
6. If person A's risk of getting a disease is 1% in ten years, and person B's risk is double that of A's, what is B's risk? |_____|%	78
7. If person A's chance of getting a disease is 1 in 100 in ten years, and person B's risk is double of that of A's, what is B's risk? |_________| out of |________|	73
5. Which of the following numbers represents the highest probability? |__| 1%, |__| 10%, |__| 5%	73
10. If the chance of getting a disease is 20 out of 100, this would be the same as having a ___% chance of getting the disease.	71
4. Which of the following numbers represents the highest probability? |__| 1 in 100, |__| 1 in 1000, |__| 1 in 10	61
2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize is 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket to BIG BUCKS? |________|	57
1. 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 (2, 4, or 6)? |_______|	54
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? |______|%	44
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? |______|	38

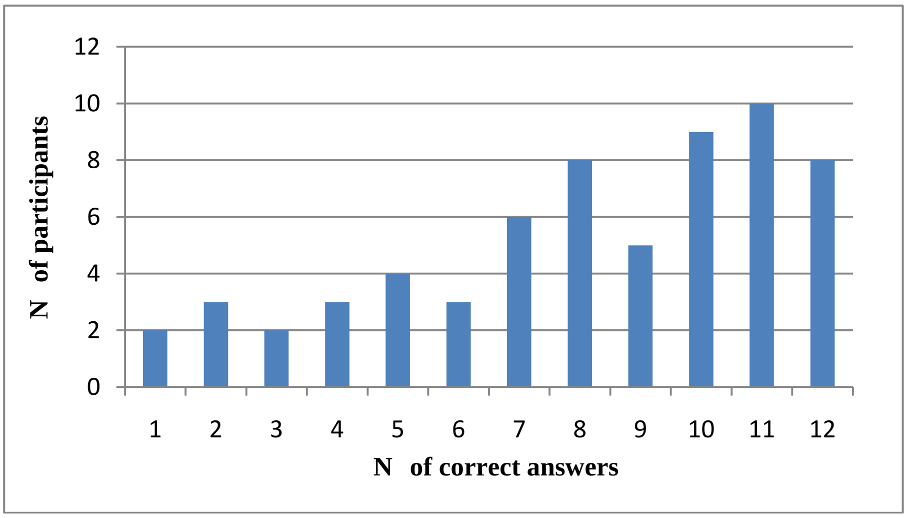


Fig. 2. **Distribution of participants' performance according to the overall score in the Numeracy scale**

Discussion

Using either 1 in 200 or 5 in 1000 when communicating a .005 health-related risk induced a bias in the degree of perceived probability. However, this distortion was not in the direction suggested by studies consistent with the idea of denominator neglect implied in explanations described in Section 1 of the theoretical chapter, which we had intuitively supported. Indeed, patients who would neglect denominators when assessing probability ratios would perceive "5 in 1000" as conveying a higher probability than "1 in 200", and not the contrary, as present results instead showed. Thus, findings were congruent with Hypothesis 2 suggesting that people neglect numerators, rather than denominators, when evaluating a probability conveyed through a ratio in single presentation conditions, and that they focus on the magnitude of the denominator. Indeed, a higher magnitude perception for probability was found when expressed as "1 in 200" than as "5 in 1000". However, applicability of the group-diffusion explanation implied in the findings to the present case was perceived as rather awkward. Could it be, that people's higher perceived probability of having a Down syndrome-affected child detected in the 1 in 200 rather than in the 5 in 1000 condition be due to a lower feeling of safety when being alone (compared to when other 4 possibilities were there, as in the 5 in 1000 case)? The question was challenging, but such explanation had, anyway, to be taken with caution, given that present experimental design and material were different from those used in group-diffusion experiments.

Undoubtedly, instead, people's evaluative behaviour proved irrational, as the two different expressions of probability correspond in fact to the same .005 value. Looking at it from the point of view of classic economic theory, it was a case of contravention of the invariance principle, as completely "rational" people should not modify their assessments on the basis of the presentation format of problems. If people could really grasp the numerical meaning of proportions and rates, and had a clear understanding of those numbers, the probability they perceive should have not varied depending on the format used for expression. However, it is known instead, that 1) the human representational system for quantities is not a perfect one, in that mental representation of numbers follows some rules that result in well known simplifying effects (see, for instance, Weber's and Fechner's law, 1860); and that 2) fractions and ratios are especially hard concepts for the human mind (see, for instance Hecht, Close, and Santisi, 2003). Thus, considering that no objectively "right" representation of numbers exists, it seemed more appropriate not to talk about "irrational" participants, but to affirm that they used some kind of heuristic based on the superficial features of the stimuli (i.e., magnitude of the numbers in the ratios) to build a personal assessment of those quantities.

Worry for the condition showed a tendency similar to that described for perceived probability, but when investigating the causal direction between these two variables, it was found that people judged the risk more worrying because it was evaluated as more probable, and not vice versa. In accordance with that, it was decided that in Study 2, and in the next experiments, focus of research would have mainly been on the subjective assessment of probability.

The third construct measured, namely perceived severity of the outcome, did not show a significant variation across the two experimental conditions, as hypothesized.

Furthermore, data highlighted a general lack of understanding of both the concept of fractions/ ratios, and of that of risk. Although the predictions were that low rather than high numerates would be those more exposed to the bias, because of their higher degree of difficulty with numbers, findings showed the opposite. Indeed, the high numerates were those perceiving a difference between the two formats in the degree of probability conveyed, with the 1 in 200 ratio expressing a higher probability than the 5 in 1000 one. It might be, that highly numerate individuals showed the bias because they were those who really processed numerical information of the ratio expressions. Low numerate individuals, instead, could have not processed numerical information, and therefore did not show the bias. Whatever the right interpretation, nevertheless, these findings called for a replica given the low number of individuals tested on whose basis the analyses were performed.

Study 1.2

Goals

In Study 1 a bizarre bias in probability evaluation was detected showing female patients' irrational tendency to judge the same health-related probability value as higher if stated through a ratio employing comparatively smaller (i.e., 1, 200) rather than larger (i.e., 5, 1000) numbers at the numerator and denominator. The result was clearly inconsistent with the idea of denominator neglect suggested by Chapter 2 Section1-like studies, and expressed in Hypothesis 1. Instead, it was coherent with Hypothesis 2 in line with the theories reviewed in Section 2 of Chapter 2, suggesting that, when evaluating probability conveyed through a ratio, people show a tendency to neglect denominators and judge on the basis of the magnitude of numerators.

Study 1.2 aimed at achieving two goals, namely

1) to replicate the main result obtained in Study 1, extending it to the Down syndrome domain, and to the general adult population;
2) to verify if the bias could be moderated by the perceived severity and the familiarity with the clinical condition.

Regarding the first goal, i.e. replica, beside the obvious consideration that a tendency found only in one study should not be taken seriously unless replicated, other reasons urged us to replicate the findings. First of all, it has to be considered that the population tested in Study 1 included a rather sensitive group of individuals (i.e., pregnant women, and new mothers), thus extending the finding to non-sensitive individuals (i.e., general population) was seen as advisable. Secondly, the effect had been tested on women only, thus both females and males were meant to be involved if the aim was that of generalizing results to all the adult population (i.e., prospective parents of both genders). Third, in Study 1 the outcome was fictitiously attributed to a friend, that is a person who is close to the protagonist but nonetheless is not the protagonist (for the differences between self and other perspective, see e.g., Kirkpatrick and Epstein, 1992). Beside the fact that self- rather than others-perspective constituted our real object of interest, a replica of the effect in self-attribution conditions seemed also more ecological, as in real life many more situations see people faced with choices related to their own health rather than to one friend's or relative's.

The second goal, considered the fact that perceived severity and familiarity with the disease could have been two potential moderators of the effect found in perceived probability, because some studies have argued for an influence of the availability of examples of cases involving the occurrence of the health outcome on its perceived probability: The greater the availability, the greater the probability of the outcome would seem. For instance, Gates (2004) mentioned the moderating role of already having, or knowing someone who has, a child with a particular birth defect. According to the author, for such a person it would be easier that examples of disabled children be brought to mind by a communication of risk related to birth defects than for another who does not have similar familiarity with the issue. As a consequence, such person would perceive the same probability of the risky outcome as higher than a person who does not have a similar degree of familiarity with the issue. In a analogous direction can be classified Wertz, Sorenson, and Heeren (1986)'s claim that the risk of having a disabled child tends to be interpreted as higher if a woman has personal experience with an affected child than if she does not. Scholars, indeed, found that having a disabled child at home was associated with a more pessimistic interpretation of risk than that of people without an affected child at home, an interpretation which in turn tended to produce a higher risk perception.

Thus, summing up these considerations with the original need of broadening Study 1 findings to the general population, three were the specific hypotheses for the present study, namely:

> **Hypothesis 1: Probability will be judged as higher when expressed as "1 in 200" rather than as "5 in 1000".**
>
> **Hypothesis 2: Those individuals who perceive greater severity of the disease will show the bias (i.e., 1 in 200 > 5 in 1000) to a greater extent.**
>
> **Hypothesis 3: Those individuals who are more familiar with the health-related outcome (i.e., that already have, or know someone who has, a Ds affected child) will show the bias (i.e., 1 in 200 > 5 in 1000) to a greater extent.**

Method

Participants. One hundred and thirty-four individuals (90 women, 43 men, and 1 individual not reporting the gender) from the general population took part in the study voluntarily. They were approached at several public locations (e.g., libraries, IT rooms, common areas), mainly in a University in North Italy. The experimenter asked them individually to participate in the research. Mean age of participants was 24.90 years (*SD* = 8.14, ranging from 18 to 66). Many of them (47%) were University students. Accordingly, education level was fairly high, as 80.5% of the sample had completed college (*n* = 95), 16.1% (*n* = 19) had a University degree, and 1 person had a higher level education, while 3 participants (2.5%) had the lowest level of education in Italy (6 participants did not answer to this question). The large majority of the sample- 92% (*n* = 102)- had no children (13 individuals did not answer to this question). The majority of participants (*n* = 107) did not report whether they had been or not communicated the probability of having a child with Down syndrome by a health professional; only 2 of those who answered (i.e., 11.8%) did it in an affirmative way.

Design and material. Participants read a fictitious scenario describing a couple of parents-to-be being communicated by the gynaecologist the maternal-age probability that their child would be affected by Down syndrome. Such probability, expressed through a numerical ratio, could be expressed in two variants- the single factor manipulated between-subjects in the study: the magnitude of numerator and denominator was modified, and could be either comparatively smaller (i.e., "1 in 200"), or comparatively larger (i.e., "5 in 1000").
After having read the scenario (whose instructions asked to role-play one of the protagonists, namely, Anna, the mother-to-be, if females, or Luca, the father-to-be if males), participants were asked to rate within this context their subjective perception of the magnitude of the probability of having a Down syndrome affected child on a 7-point Likert-scale anchored to "extremely low" and "extremely high" with all points verbally labelled. Next, two further questions followed. The first question assessed the perceived severity of the disease on a 5-point Likert scale from "not at all severe" to "extremely severe"[46], while the second one assessed the degree of knowledge of the disease in terms of personal experiences with people affected by the condition (i.e., "familiarity")- for the experimental material employed, please see the Appendix[47].

Procedure. The study was described as part of a broader study on risk communication. A paper version of the questionnaire was handed out personally to each individual by the experimenter, who later collected the completed copies and debriefed participants on the study aims; individuals could complete the questionnaire at their pace, without any time constraints[48]. Participants were randomly assigned to one of the two conditions constructed to follow the experimental design. A minimum of 30 participants per condition was necessary to make statistical comparison valid. Individuals were allowed an unrestricted amount of time to complete the questionnaire.

---

[46] This time, severity did not measure the probability of the outcome, but the outcome alone- i.e., "According to you, how serious is Down syndrome?" instead of "According to you, how serious would be the X probability of having a Down syndrome affected child?"
[47] All the questions in the questionnaire appeared on a separate page so to maximize participants' attention to the information presented in that only page- the question on perceived probability appeared in the same page as the related scenario.
[48] The same procedure was employed for all the other studies reported in the present dissertation. When not differently stated, thus, it will not be repeated in the next studies.

Results

Data of ten participants had to be removed from the sample because they declared not knowing how to rate the severity of the Down syndrome disease. Likely, their probability assessment of the ratio was not the result of a sound evaluation. Thus, the final number of participants in the sample was 124.

The effect due to the magnitude of the numbers employed in the ratio. Ratio format had an impact on the main dependent measure, namely perceived probability, as showed by the significant result of a t-test comparison of the means in the two experimental conditions, $t(122)= 2.60$, $p = .01$. Indeed, the subjective magnitude of the probability was 3.75 ($SD = 1.43$) when that was phrased as "1 in 200", and 3.12 ($SD = 1.27$) when it was phrased as "5 in 1000". Thus, Hypothesis 1 was supported by data.

The influence of perceived severity of the disease and familiarity with the disease. Aim of these analyses was that of understanding whether the format bias (1 in 200 > 5 in 1000) was greater/lower/the same for a) those individuals that perceived the possible outcome as highly severe and for those who perceived it as lowly severe (i.e., severity variable); b) those individuals who were familiar to a lower vs. greater extent with the disease under study (i.e., familiarity variable).

a) Neither the median split (*Mdn* = 3), nor that on the basis of the 33rd and 66th percentiles (i.e., 3 and 4) on the severity measure returned groups that could be compared in terms of numerosity- see Table 1. Indeed, the majority of the sample was distributed between the two answers, namely that affirming that Down syndrome is a "quite severe" disease (46% of the sample, $n = 57$), and that affirming it is a "very severe" disease (29%, $n = 36$). Hence, it was not possible to test Hypothesis 2.

Table 1
Distribution of participants according to the perceived severity of the disease (Down syndrome)

Answer	Frequency	%	Valid %
Not all severe	3	2.4	2.4
Slightly severe	19	15.3	15.3
Quite severe	57	46.0	46.0
Very severe	36	29.0	29.0
Extremely severe	9	7.3	7.3
Total	124	100.0	100.0

b) In a similar way, unfortunately participants' distribution on the degree of familiarity with the disease was highly skewed, because the majority of answers were either on the 2nd ("I have seen Down syndrome affected-people but never interacted with them") or 3rd answer ("I have been interacting with them, but sporadically") - see Table 2. Thus, it was not possible to split participants on the basis of the mean familiarity measure, nor on that of the median familiarity measure as correct analyses would require; both procedures resulted in largely unbalanced groups.

It was decided, nevertheless, to try some explorative analyses, and hence such a different dichotomisation of participants was created. When those participants occupying the most extreme positions on the variable were excluded from analysis (i.e., those who had never heard about the disease, n = 12; and those who had been interacting with, or normally used to interact with, Down syndrome-affected people, n = 17), the remaining participants were split in two groups, namely those "lowly familiar" [with the disease], n = 44 (i.e., those who declared to have heard about Down syndrome disease, but never seen a person affected by the disease, and those who stated they have seen Down syndrome-affected people, but had never been interacting with one of them), and those "highly familiar" [with the disease], n =59 (i.e., those who have interacted, even if sporadically, with people having Down syndrome). While there was no effect due to the format of the ratios in the former type of participants (Z = - .68, p = . 50), the latter instead showed the bias in perceived probability (Z = - 2.00, p = .046). Thus, despite these results would call for a replica, it can be affirmed that Hypothesis 3 was corroborated by data.

Table 2
Distribution of participants according to the declared degree of familiarity with the disease (Down syndrome)

Answer	Frequency	%	Valid %
Never heard about it	0	0	0
Heard about but never seen	12	9.9	9.8
Seen but never interacted with	32	26.4	28.5
Interacted with, but sporadically	59	48.8	48.0
Frequently interacted with	17	14.0	13.8
Total	123	99.2	100.0
Missing answers	1	0.8	
Total	124	100.0	

Discussion

Study 1.2 generalized the effect described for female patients in Study 1 to the general adult population and for the risk that was the actual object of study of the present work (i.e., that of having a Down syndrome- affected child). In particular, it was found that when evaluating a probability, laypeople's perception was influenced by the magnitude of the numbers employed in the ratios: Using "1 in 200" resulted in a larger probability perception than using "5 in 1000", despite corresponding to the same .005 probability. Such bizarre bias was found not only in the case the outcome at study was described as potentially affecting a person which was close to the participant (i.e., a friend, see Study 1), but also when the participant her/himself was the person potentially affected by the outcome (i.e., in self-perspective, see Study 2).

Unfortunately, it was not possible to test whether the perceived severity of the disease had influenced the bias under examination, due to the fact that an exact median split was not feasible (see the highly skewed distribution of participants along this variable). The same has to be affirmed for what concerned the level of familiarity with the disease, but when those individuals at the extremes of the scale (i.e. those that can be considered in the dark about the medical condition under study, and those, that on the contrary, are completely familiar with it) were excluded from analysis, familiarity level showed an interesting mediating effect on the appearance of the bias under study. Indeed, only participants that had high familiarity with the disease showed a higher probability perception for the 1 in 200 rather than the 5 in 1000 format.

SECTION 2, or
Boundaries of the bias

In Section 1, a new bizarre bias has been established: When the same probability of a medical outcome was expressed through a ratio format, its perception was systematically higher when the numbers in the ratio were smaller (i.e., 1 in 200) rather than larger (i.e., 5 in 1000). In the present section, six experiments are presented.

In Study 2.1, the bias in probability perception found in Section 1 will not only receive additional corroboration, but further extension to those less frequent but still possible cases where a ratio employs a larger denominator (i.e., 10,000). Indeed, it will be showed, that in terms of probability perceptions 1 in 200 > 5 in 1000, and also that 1 in 200 > 50 in 10,000. Exploratory analyses conducted on individual's decision style (despite numerous problems blurring indexes on which basis the categorisation will be made) will reveal an apparent association of the bias with individuals' experientiality score. Nevertheless, the expected higher appearance of the bias in those individuals high in experientiality compared with those low in experientiality will not manifest; on the contrary, low experientials will be those showing a stronger bias, probably because they elaborated numbers at least to a certain degree. Finally, in this study previous results showing that a high level of numeracy is associated with the bias (see Study 1.1) will be confirmed with a subjective measure of the construct.

In the next study (i.e., 2.2), it will be showed that the bias extends well beyond the specific numerical ratios considered till then, as it will prove valid for comparisons between proportions (i.e., ratios in a "1 in X" format) and rates with 1000 as denominator. The systematic tendency to perceive the same probability as higher when expressed in a "1 in X" than in a generic "N in NX" format will be dubbed the "*1 in X effect"* as effect of the generalisation of results to ratios with that superficial structure.

Study 2.3 will confirm the validity of the bias when rates using 100 (120, to be precise) as denominator will be employed; furthermore, exploratory analyses on the degree of understanding of the two types of format, will confirm results obtained in Study 2.2 on the same variable but through a different inventory. Indeed, in a similar fashion, no difference will be found between "1 in X" and "Y in 100" (or "Y in 1000") formats in the difficulty they pose for comprehension- an unexpected result if one considers claims of a higher difficulty of proportions rather than of rates stemming from the literature.

Till then, results will seem to converge on some specific features of the "1 in X" format. A further study, i.e., 2.4, will be conducted to exclude, that the higher probability perception for "1 in X" than for "N in NX" formats found reflects in fact a general focus on the denominator of the ratio, rather than a specific effect of the "1 in X" format. It will be showed that such possibility is very unlikely, as the probability perceptions of two equivalent ratios, none appearing in formats featuring directly a “1 in X”, will not differ significantly. However, a direct comparison between pairs of equivalent “1 in X” and “N in NX” formats, and pairs of equivalent N in NX formats will be deemed as necessary to demonstrate, that only when the “1 in X” format is present in the comparison, the effect in perceived probability manifest itself. Such demonstration will be showed in Study 2.6, where the idiosyncrasy of the “1 in X” format found for abstract ratio values (namely, not referred to an outcome, see Study 2.5) will receive further confirmation in a contextualized situation. Indeed, while people will generally tend to transform ratios expressed in “N in NX” formats, and they will not do that with a ratio in “1 in X” format.

Study 2.1

A special thanks to Sandra Eccel for collection of data presented in this study

Goals

Studies 1 and 2 of the previous section assessed a bias in laypeople's probability perception due to two specific ratio formats (i.e., 1 in 200 and 5 in 1000). In particular, the bias consisted in an irrational tendency to judge the same health-related probability value as higher if stated through a ratio employing comparatively smaller (i.e., 1, 200) rather than larger (i.e., 5, 1000) numbers at the numerator and denominator.

The first issue addressed in the present study was the following: What about this tendency when numbers even higher than those already analyzed are used in the ratios? Despite probably less frequent in face-to-face communication, the employment of large numbers in a ratio expressing a probability takes place even for medical risks. It is especially the case of mass-media-driven communications on base-rates of events happening in a population. But it can also be the case of a physician-patient communication, when for instance the doctor is suggesting vaccination against a disease to her/his patient about to go on vacation in an exotic country. Or, that of different treatments for a disease in terms of survival rates (i.e., number of people that get completely cured, for instance, per 100,000 people). Garcia-Retamero and Galesic (in press-b, p. 1), for example, referred to the recent case of communications on swine flu pandemic deaths ("about 5700 deaths out of 440,000 laboratory-confirmed cases in the world"), and that of information on lung and stomach cancer in the United Kingdom ("40.1 and 5.7 deaths per 100,000 people per year") as example of the use of ratios employing larger denominators.

If in similar cases, the use of large denominators can be necessary, nevertheless, grasping the meaning of such numbers may be particularly difficult both for laypeople and specialists, for mainly two reasons.

First of all, because of the fuzzy representation of such quantities in the mind: some authors have claimed that numbers follow the same "psychophysical function" described by work of Weber and Fechner and known as the "Weber's law" (Weber, 1834; Fechner, 1860). Weber's law is believed to characterize "our diminished sensitivity to a wide range of perceptual and cognitive entities—brightness, loudness, heaviness, and money—as their underlying magnitudes increase" (Slovic, 2007, p. 6). According to some scholars (see Slovic, Finucane, Peters, and MacGregor, 2004), encoding of quantities in the experiential system (besides in the analytic one), would follow those rules. In other words, some scholars think, that the affective system is designed to make individuals sensitive to small changes in the environment (e.g., the difference between 0 and 1 deaths) with the drawback of rendering them insensitive to larger changes further away from zero (e.g., the difference between 800 and 900 deaths), a tendency that has been dubbed "psychophysical numbing" (Fetherstonhaugh, Slovic, Johnson, and Friedrich, 1997). If that principle is valid even in the field of medical diseases, and for the probability of having a Down syndrome-affected child, then some specific predictions can be made for the outcome analysed in the present work.

Also, to many proportions with large denominators are confusing, as it has been stressed by Grimes and Snively (1999). That is exemplified in the words reported by authors (p. 910) and attribute to Walker, namely "To many, 1/400 sounds higher than 1/200 because the denominator is bigger". Even literature in decision making has disclosed laypeople general incapacity to appreciate differences among variations in low probabilities expressed by means of ratios employing extremely large numbers at the denominators (e.g., 1000, 10,000, 100,000, a million), especially when an assessment of these values in isolation is required. For instance, in one study, Kunreuther, Novemsky, and Kahneman (2001) showed that risks of 1 in 100,000, 1 in 1 million, and 1 in 10 million elicited the same subjective perception in participants. In a similar fashion, they also found little difference in probability perceptions for probabilities ranging from 1 in 650, to 1 in 6300, to 1 in 68,000. Thus, when assessing ratios using denominators of such orders of magnitude, people might well show a relative "numbness for numbers", namely a difficulty in grasping the meaning of statistical information.

Secondly, the difficulty with large numbers might take place "because direct experience with large groups of people is relatively rare both in our evolutionary history and in daily life", as it has been argued by Garcia-Retamero and Galesic (in press-b, p.1). Indeed, as authors claimed referring to the work of Dunbar (1993) in line with the "frequency hypothesis" (e.g., Gigerenzer, 1994; Gigerenzer and Hoffrage, 1995), "the typical size of human groups in a wide range of ancient and modern human societies is 100–200 people" (Dunbar, 1993, p. 1). As a consequence of that, authors believe, individuals would find it easier to imagine, understand, and recall "smaller, evolutionarily plausible groups of people" (Dunbar, 1993, p. 1), than larger ones. That is exactly what authors have demonstrated in their study (Garcia-Retamero and Galesic, in press-b)[49].

[49] Analyzing the literature on the specific issue addressed in the present thesis (i.e., Chapter 2), only one study can be mentioned that dealt with probabilities expressed through ratios employing "large" numbers. It is the case of Denes-Raj, Epstein, and Cole (1995)' study, where the ratios tested by means of the usual urn-and-ball or lottery scenario were "1 in 1000" and "10 in 10,000" (Study 1 and 2), in a within-subjects design. In those experiments, people generally expressed a judgment of higher probability both in self and others-perspective for the option whose likelihood was stated in (comparatively) larger (i.e., 10 in 10,000) numbers. Nevertheless, in the third study, i.e. when a health outcome was at stake (HIV contamination), no effect was found in self-perspective with the same ratio values of the previous two studies. On the other hand, if it is true that in studies on the group-diffusion effect, implicit rates using numbers in these

Given results obtained in the previous section, intuitively one would have expected that the same tendency (i.e., a larger perceived assessment for the ratio expressed by means of comparatively larger numbers at the numerator and denominator) be found even when the ratio employed a denominator in the 10-thousand-order of magnitude; more precisely: a negatively linear relationship between the magnitude of the numbers used at the numerator and denominator of the ratios and the level of perceived probability for the ratio. That translated in the following first hypothesis for the present study, namely:

Hypothesis 1: A higher magnitude perception for a .005 probability will be found when expressed as "1 in 200" than as "5 in 1000", and when expressed as "5 in 1000" than as "50 in 10,000".

Also, another issue was judged relevant for the topic, namely discovering whether the effect found in probability perception was moderated by the prevalence of one of the two decision styles over the other (i.e., rational vs. experiential- for a characterization, see the section on Slovic' s theory on the two ways of risk comprehension, in Chapter 2, and that on CEST, ibid.). In line with the experiential/ analytical model of risk perception (see Finucane, Alhakami, Slovic, and Johnson, 2000; Slovic, Finucane, Peters, and MacGregor, 2004; Slovic and Peters, 2006) and the System1/ System2 model of decision making (for a characterisation, see Kahneman, 2003), it was expected that:

Hypothesis 2: The bias should have been more frequent in those persons who show a highly experiential decision style (mainly in the highly-experiential-lowly rational individuals) rather than in those who have a highly-analytical decision style.

In order to test this hypothesis, it was decided to use the REI (Rational-Experiential Inventory) scale. REI is an inventory designed to assess individuals' preferences for information-processing styles, available in various versions (the original one: Epstein, Pacini, Denes-Raj, and Heier, 1995/6; Norris, Pacini, and Epstein, 1998; Pacini and Epstein, 1999-b, with 40-items). Theoretically motivated by the Cognitive-Experiential Self-Theory, all the versions cited above distinguish between the two cognitive styles, namely a rational one, measured by an adapted Need for Cognition (NFC) scale (see Cacioppo, Petty, and Kao, 1984), and emphasizing a conscious, analytical, intentional, approach, and an experiential one, measured by

---

order of magnitude have been used in between-subject designs (e.g., a large group ranging from 100 to 1 million people), it is also true that those rates were only implicit.

the Faith in Intuition (FI) scale (Epstein, Pacini, Denes-Raj, and Heier, 1996) emphasizing a pre-conscious, affective, automatic, holistic approach.

In this study, the REI version adopted was the 40-item one (Pacini and Epstein, 1999-b), as recommended in both “dmidi.net”[50] and by Pacini and colleague, that used it to corroborate their hypothesis on a prevalence of the experiential system in individuals showing the Ratio-bias (Pacini and Epstein, 1999-b). Authors judged this version of the inventory as superior to the preceding ones they had created, in particular that of ‘96, because: a) in the preceding version, contents of the two sub-scales (NFC and FI) did not parallel each other, while in the new version they did, as each scale had comparable subscales, namely engagement and ability; b) in the preceding version, the NFC scale was more reliable than the FI scale, while in the new version they both were; c) in the previous version, the two scales were unbalanced in terms of item valence, while now negatively and positively worded items were equally numerous (Pacini and Epstein, 1999-b). The inventory includes twenty items measuring rationality, and twenty measuring experentiality; each of the scales is constituted by two subscales comprising ten items assessing the ability to use each of the systems and ten assessing the engagement in each of the systems (i.e., reliance on use and enjoyment in using)- see the section of the Appendix corresponding to the present study for full lists of items.

A third hypothesis concerned the replica of results found in Study 1.1 for what regarded ability with the concept of ratio and numbers, i.e., numeracy. Those findings seemed to indicate, that the new bias was more likely for highly-numerate individuals. It has to be reminded, that such a result would not be, in principle, in contradiction with theories on error-leading heuristics. Indeed, highly numerate individuals are not excepted from biased perceptions, despite being true that in many tasks they do seem to suffer in a lower extent from well-known systematic tendencies (e.g., Reyna, Nelson, Han, and Dieckmann, 2009; Garcia-Retamero and Galesic, 2009; Garcia-Retamero, Galesic, and Gigerenzer, in press; Reyna and Brainerd, 2007; Reyna and Brainerd, 2008)- see also dedicated paragraph at the beginning of the present chapter.

A different inventory was used to assess numeracy in this study, more precisely an inventory measuring the construct in a *subjective* rather than *objective* way. To that end, the Subjective Numeracy Scale (SNS) developed by Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, and Smith (2007) was chosen instead than the objective numeracy inventory (Lipkus, Samsa, and Rimer, 2001) used in Study 1.1. The reasons for this substitution are described here following.

First of all, as the present study included already a long inventory (i.e., the 40-items REI scale), the use of this shorter version seemed more appropriate. When they developed the SNS,

[50] The DMIDI (Decision Making Individual Differences Inventory) is a catalogue comprising the main individual difference measures commonly used in judgment and decision-making research (over 170, in fact), and available at the URL: http://www.dmidi.net/. DMIDI was developed by Kirstin Appelt with Kerry Milch, Michel Handgraaf, and Elke Weber.

Fagerlin and colleagues' aim was to create an easier, shorter, more enjoyable way to assess facility with various numerical tasks involving percentages, ratios, and transformations on such formats. Authors were successful, as the SNS showed to be a good proxy for the original numeracy scale, with the following interesting plus (in comparison): the scale was a) quicker (completing time: about 2.5 min less than those who had to complete the objective numeracy test); b) less annoying, less stressing, less frustrating; c) it prompted a higher will to repeat a similar survey; c) it caused less missing data (Fagerlin et al., 2007).

Moreover, Keller, Siegrist, and Visschers (2009) also advocated in favour of such scale as it measures perceived self-efficacy (i.e., it is a self-report measure of the perceived ability to perform some mathematical operations) rather than ability with those operations. The first construct seems to be a better predictor than the second of the extent to which people actually engage in a particular task. "Based on the self-efficacy literature, it seems plausible that subjective numeracy skills are more important than objective numeracy skills for the successful interpretation of risk information", authors conclude (Keller, et al., 2009, p. 1256). Finally, another positive feature of the SNS consists in the possibility to measure, beside cognitive abilities, people's preferences for the display of numerical information, an issue not tapped by the objective numeracy scale.

Following findings of Study 1.1, the hypothesis for the present study was:

Hypothesis 3: High-numerate individuals are more likely to show the bias in probability perception than low-numerate individuals.

Finally, the fourth hypothesis of this study regarded the possible moderating effect of the degree of familiarity with the outcome of the scenario (see Study 1.2). Indeed, as explained in the previous study attempting to assess this variable, some literature pointed to a higher probability perception for those people who show a certain degree of familiarity with the outcome at study (in particular, the greater the availability of examples in the person' mind, the greater the probability of the outcome in question would seem). The assumption was that, those who had a higher familiarity with Down syndrome would have been more likely to process the numbers and hence show the bias, than those who had a lower familiarity with the outcome.

Hypothesis 4: Those individuals who are more familiar with the health-related outcome depicted in the scenario (i.e., that already have, or know someone who has, a disabled child) show the bias to a larger extent.

Method

Participants. Two hundred and forty-eight individuals (138 males and 110 females), from the general population took part in the experiment voluntarily, all recruited face-to-face in several public locations (e.g., University cafes, library).

Mean age was 31.91 (*SD* = 10.97), ranging from a minimum of 18 to a maximum of 60 (5 participants did not disclose this information). Of the sample, 66.1% (*n* = 162) had completed the high school, 14.7% (*n* = 36) had a University degree (3/5 years), 3,7% had a higher title like Master, PhD, or Specialization, and the 15.5% had a low level of education ("licenza media" or "licenza elementare"). A large part of the sample (*n* = 113), namely almost half of it (45.7%), had one or more children, while 134 people stated they did not have any (1 person did not answer to this question). Some of the participants (*n* = 44; 26%) affirmed having being communicated by a medical specialist the probability that the future child could have Down syndrome (79 people did not answer to this question).

Design and material. Participants read instructions asking them to role-play and imagine themselves as prospective parents (i.e., Anna, the mother-to-be, if females, or Luca, the father-to-be if males) at risk for a child affected with Down syndrome. In the scenario, parents-to-be received information from the gynaecologist of their maternal age-related probability of having an affected child. Such probability, expressed through a numerical ratio, could assume three variants- the single factor manipulated between-subjects in the study- where the magnitude of numerator and denominator was modified, thus resulting in 1 in 200, 5 in 1000 (as in previous studies), and 50 in 10,000.

Within that context, participants were asked to rate their subjective perception of the magnitude of the probability ratio on a 7-point Likert-scale from "extremely low" to "extremely high". On a separate page, then, participants were required to answer the 40-item REI scale of Pacini and Epstein (1999-b), translated into Italian- five lists were created where the order of the questions was varied to control for possible order effects. Alike in the original version, in the present study a 5-point Likert scale from "definitely not true of myself" to "definitely true of myself" was available to respondents for each item.

On a separate page, participants read the 8-item inventory assessing numeracy developed by Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, and Smith (2007), that is the SNS (Subjective Numeracy Scale), translated into Italian.

Procedure. Participants were randomly assigned to one of the fifteen conditions constructed to follow the experimental design. A minimum of 30 participants per ratio format condition was necessary to make statistical comparison valid on this variable, but that number was doubled in order to be able to dichotomize each group on the basis of the low/high rationality/

experentiality score on the REI scale, in such a way to examine possible effects of these thinking-style features on the main dependent variable. Individuals were allowed an unrestricted amount of time to complete the questionnaire by themselves.

Results

Reliability and validity of the REI scales. The reliability and validity of the REI scales received little support in the present sample, as well as their orthogonal two-factor structure, both necessary prerequisites to their employment for analysis. Nevertheless, some exploratory analyses on the basis of the original complete scales were made.

Table 1 summarises the rather ambiguous results obtained when considering all items of the REI. Reliability of both main scales was quite low (Rationality scale, α = .61 almost acceptable; Experentiality scale, α = .41, liable to problems)[51]. Indeed, while each main scale was strongly positively correlated with both its Ability and Engagement subscales (in accordance with the literature), and the latter were moderately related within each mode, other findings were either diverging with those expected theoretically, or unexplainable. For example, the correlation between the main scales was significant, even if low, $r(246) = .25$, $p < .001$, a data that has never been reported in the literature as, in line with the CEST assumption of the existence of two parallel but autonomous information-processing modes, the two scales are supposed to be independent. The average inter-item correlation was extremely low in both scales (.071 in the Rationality scale, and 0.39 in the Experentiality scale), a rather problematic data as it suggested that the constructs were not being measured reliably-that is, there were sources of unexplained error in the measurement- or, that the instruments were not measuring the intended elements. In addition, two rather suspect correlations were found, that were that between Rationality and Experiential Engagement, $r(246) = .31$, $p < .001$, and that between Rational Ability and Experiential Engagement, $r(246) = .31$, $p = .001$, which could not be explained- see again Table 1.

[51] In the original Pacini and Epstein (1999-b) article, both scales reliability were rather high (i.e., α = .90 for the Rationality scale, and α = .87 for the Experentiality scale), and the same can be affirmed for subsequent studies employing the same cognitive style scales, either in the English version (e.g., Marks, Hine, Blore, and Phillips, 2008), or translated for use in non-English speaking Countries (e.g., Bjorklund and Backstrom, 2008; Shiloh, Salton, and Sharabi, 2002; Witteman, van den Bercken, Claes, and Godoy, 2009).

Table 1
Intercorrelations and Reliabilities of Rational-Experiential Inventory (REI) scales

REI scale	1.	2.	3.	4.	5.	6.
1. Rationality	(.61)	.85°°°	.88°°°	.25°°°	.11	.30°°°
2. Rational Ability		(.41)	.50°°°	.23°°°	.08	.31°°°
3. Rational Engagement			(.33)	.20	.12	. 22
4. Experientiality				(.48)	.83°°°	.81°°°
5. Experiential Ability					(.29)	.35°°°
6. Experiential Engagem.						(.05)

Note. N = 248. Reliabilities appear in diagonal in parentheses.
°°p = .01 °°° p = .001

Thinking-styles typology. Despite the problems highlighted in the preceding section, participants were dichotomized on the basis of the median scores on Rationality and Experentiality[52], with the final aim of performing an exploratory analysis of variance on the perceived probability level. Thus, based on the median splits, a substantial equal numerosity was obtained in each thinking-style group, with the 27.8% of the sample (n=69) classified as low on both scales, the 25.4% of the sample (n = 63) as high on both scales, while the 24.2% (n = 60) as high in Rationality but low in Experentiality, and the 22.6% (n = 56) as high in Experentiality and low in Rationality.

The Subjective Numeracy Scale (SNS). The scale showed satisfactory internal consistency-Cronbach α was .88, a very good value. In order to calculate each participant SNS score, the same methodology used in Fagerlin et al. (2007) was employed: the ratings across all SNS items answered were averaged for all participants who had missing data yet completed more than half of the SNS questionnaire.[53]

Participants' scores on the SNS ranged from 1 to 5[54]. Table 2 reports the mean score in the entire sample for each item composing the scale, while Fig. 1 reports the distribution of each answer for the 8 questions.

52 Despite being associated with problems such as loss of power, the lack of symmetry in the scales (Kolmogorov-Smirnov test: p values < .05), justified the median split approach used to divide the sample-$Mdn_{Rationality}$ = 2.85, $Mdn_{Experentiality}$ = 3.25.

53 The same results were obtained even when excluding item number 3 from the calculations alike in Witteman, van den Bercken, Claes, and Godoy (2009), considering that due to cultural reasons the present sample could have not had familiarity with the task figured in the question.

54 All items in the inventory had six-point Likert scales, but we preferred to reduce them to 5 due to a better adaptation with the range of available terms for labels in Italian. While questions n° 1, 2, 3, 4, and 8 ranged from "not at all" to "extremely good", questions n° 6 and 7 ranged from "always prefer words" to "always prefer numbers".

Table 2
Mean score for each of the eight five-point items composing the SNS inventory (SD in parentheses)

Questions	Mean score
1. How good are you at working with fractions?	2.98 (1.01)
2. How good are you at working with percentages?	3.00 (.93)
3. How good are you at calculating a 15% tip?	3.10 (.96)
4. How good are you at figuring out how much a shirt will cost if it is 25% off?	3.21 (.97)
5. When reading the newspaper, how helpful do you find tables and graphs that are parts of a story?	3.02 (.92)
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’’)?	3.13 (1.05)
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”?	2.78(1.09)
8. How often do you find numerical information to be useful?	3.28 (.86)

As it is visible from Figure 2, the trend was very similar for all items of the scale apart from n° 8 and 7, since answers were distributed normally along the 5-points scale, with the majority of people choosing the central answer (= 3, i.e., “slightly”). Indeed, the mean score on the scale was 3.06 (SD = .72).

A median split (Mdn = 3.00), was adopted on the SNS score to divide the sample into low (n = 128) and high (n = 120) numerates. An analysis of variance (ANOVA) on perceived probability was separately conducted in each numeracy group with ratio format (1 in 200, 5 in 1000, 50 in 10,000) as between-subjects variable. A significant effect of ratio format was found in high numerates, $F(2, 120) = 6.71$, $p = .002$, $\eta_p^2 = .10$, while it was not found in low numerates, $F(2, 128) = 1.40$, $p = .25$, $\eta_p^2 = .02$, indicating that only for the former participants the three ratio formats evoked different levels of magnitude on the assessment scale. In high numerates, post hoc comparisons using Fisher’s LSD tests on perceived probability for the three ratio formats showed that 1 in 200 was perceived as significantly different from 5 in 1000 ($p = .04$) and from 50 in 10,000 ($p < .001$), but 5 in 1000 and 50 in 10,000 were not perceived as significantly different ($p = .10$). Mean measures of perceived probability in each ratio condition are reported in Table 3, and visually represented in Fig. 2. Hence, for what regarded subjective numeracy, Hypothesis 3 was confirmed by data (obviously, only for the 1 in 200 and 5 in 1000 comparison).

Table 3
Mean perceived probability for high and low numerates in the three experimental conditions (SD in parentheses)

	Numeracy		
Ratio Format	High (n = 120)	Low (n = 128)	Total
1 in 200	4.23 (1.66)	3.68 (1.49)	3.96
5 in 1000	3.55 (1.37)	3.27 (1.21)	3.41
50 in 10,000	2.97 (1.52)	3.24 (1.37)	3.11
Total	3.62 (1.59)	3.39 (1.37)	3.51

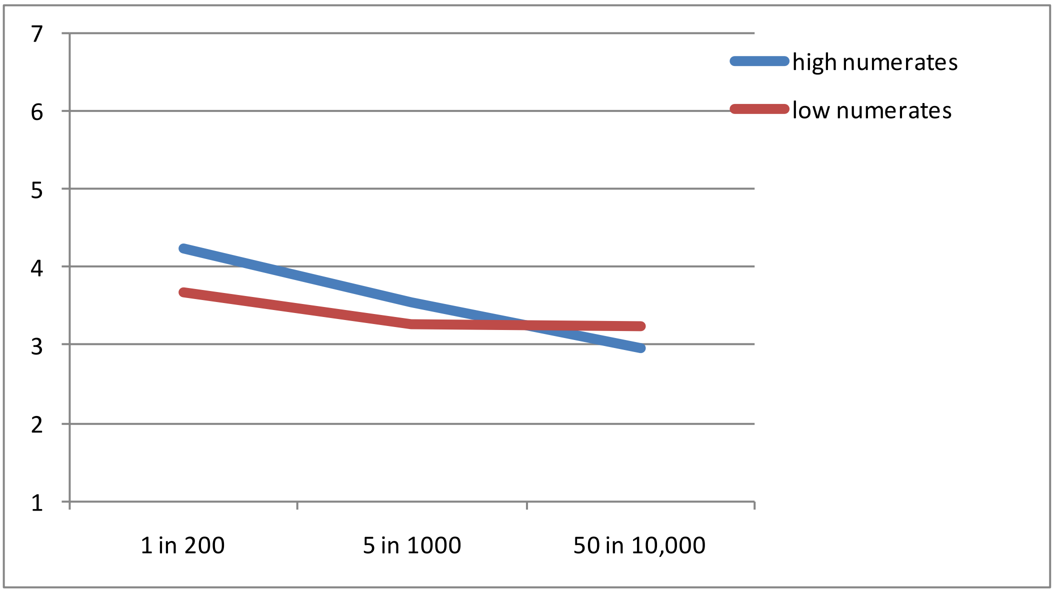
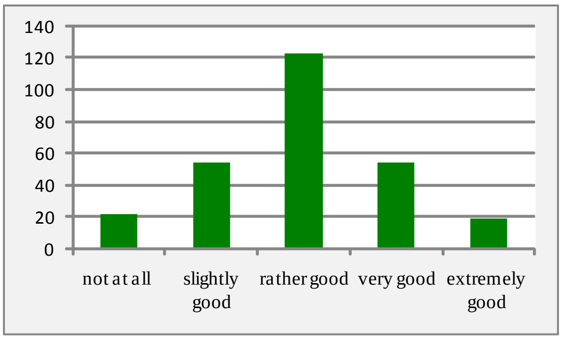


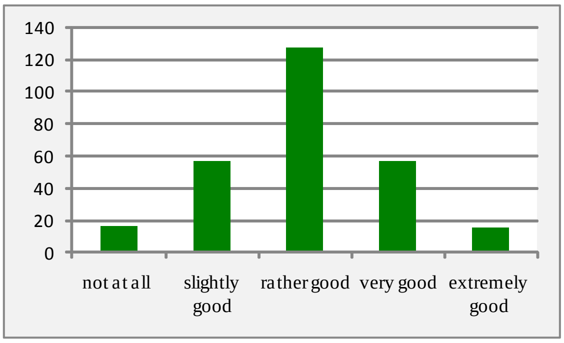
Fig. 2. **Perceived probability in the three experimental conditions as a function of participants' (low/high) numeracy level**

Fig. 1. (following page) **Distributions of answers in each of the 8 items composing the SNS scale**

1. How good are you at working with fractions?



2. How good are you at working with % ?

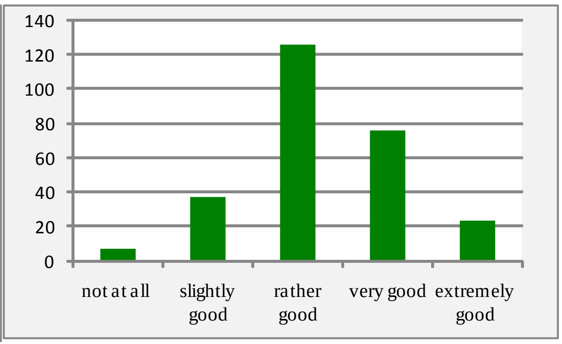


3. How good are you at calculating a 15% tip?

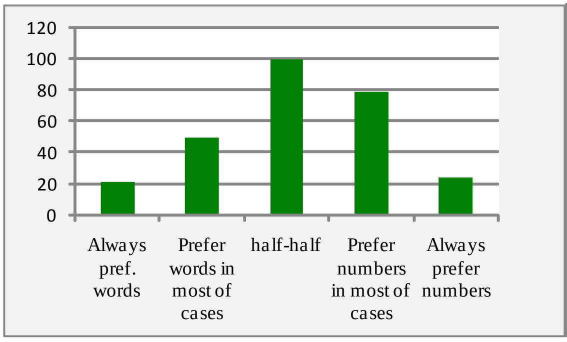
4. How good are you at calculating a 25% off price?

5. When reading the newspaper, how helpful do you find tables and graphs?

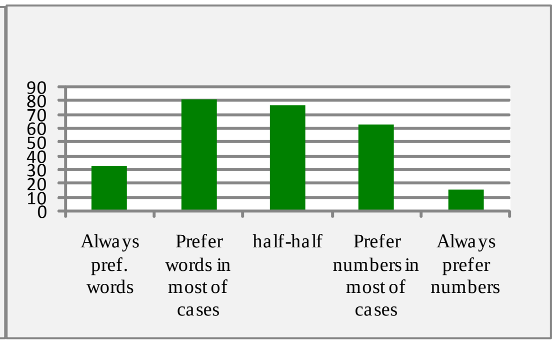
8. How often do yo find numerical information useful?



6. Do you prefer words or numbers for an outcome chance?



7. Do you prefer % or words only in the weather forecast?



The effects on perceived probability. Analyses with the entire REI scale. An analysis of variance (ANOVA) was conducted on perceived probability, with ratio format (1 in 200, 5 in 1000, 50 in 10,000), rationality score (low/high), experentiality score (low/high), and numeracy score (low/high) as between-subject variables. A significant main effect was found for the ratio format, $F(2, 224) = 5.98$, $p = .01$, $\eta_p^2 = .05$, indicating that different ratio formats evoked different levels of probability perception ($M_{1\ in\ 200}$: 3.94, $SD = 1.59$; $M_{5\ in\ 1000}$: 3.41, $SD = 1.29$; $M_{50\ in\ 10,000}$: 3.12, $SD = 1.44$). Post hoc comparisons using Fisher's LSD tests revealed that 1 in 200 was perceived as significantly different from 5 in 1000 ($p = .001$) and from 50 in 10,000 ($p < .001$), but 5 in 1000 and 50 in 10,000 were not perceived as significantly different ($p = .20$)- see Fig. 3. Taken together, these results suggested, that the bias found in preceding studies was supported, and they also indicated a formal perceived equivalence among 5 in 1000 and 50 in 10,000, despite the average perception of 50 in 10,000 was positioned in between 5 in 1000 and 1 in 200. This result will be commented in a subsequent section.

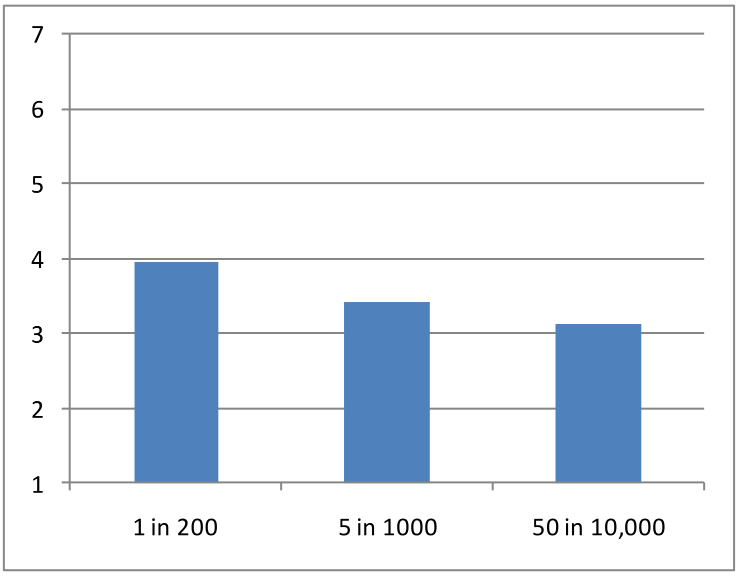


***Fig. 3*. Mean perceived probability in the three experimental conditions**

Nevertheless, no main effects of the high/low score levels of Rationality and Experentiality scales, were detected on probability perception, $F(1, 224) = 1.67$, $p = .20$, $\eta_p^2 = .01$, and $F(1, 224) = .08$, $p = .78$, $\eta_p^2 = .00$, respectively, neither of the mean SNS score (low/high), $F(1, 224) = 1.30$, $p = .26$, $\eta_p^2 = .01$. Instead, a significant interaction among ratio format and Experentiality (low/high) was found, $F(1,224) = 5.39$, $p = .01$, $\eta_p^2 = .05$, an interesting result that will be commented in a further section, as it will be replicated with the new (refined) Experentiality index.

The refinement of the scales. As both main scales of the REI were unrealiable, and items not internally consistent, the item-to-scale (i.e., item-to-total) correlation was used for scale refinement (Nunnally, 1978; Saraph, Benson, and Schroeder, 1989). In other words, those items from each scale that did not statistically agree with the other items of the scale were eliminated. For a list of the items originating the new Rationality and Experentiality scales, i.e. Rationality2 and Experentiality2, see Table 4.

With the scale refinement, reliability of the Rationality2 scale raised to .78. Despite that level could be considered satisfactory, as was in fact the average inter-item correlation (.31, equivalent to that found in Pacini and Epstein's original study), it should not be forgiven that 12 items had to be erased from the original scale, thus chasing doubts on whether remaining items could adequately tap the desired concept. The same considerations are valid for the Experentiality2 scale (α = .83, average inter-item correlation = .28) calculated on the basis of 12 items only.

Table 4
List of the items retained in the two scales composing the REI after the scale refinement

Rationality scale	Experentiality scale
1. I try to avoid situations that require thinking in depth about something.	21. I like to rely on my intuitive impressions.
2. I' m not that good at figuring out complicated problems.	23. Using my gut feelings usually work well for me in figuring out problems in my life.
4. I am not that good at solving problems that require careful logical analysis.	24. I believe in trusting my hunches.
5. I don't like to have to do a lot of thinking.	25. Intuition can be a very good way to solve problems.
7. Thinking is not my idea o fan enjoyable activity.	26. I often go by my instincts when deciding on a course of action.
8. I am not a very analytical thinker.	27. I trust my initial feelings about people.
9. Reasoning things out carefully is not one of my strong points.	28. When it comes to trusting people, I can usually rely on my gut feelings.
11. Thinking hard and for long about something gives me little satisfaction.	31. I think there are times when one should rely on one's intuitions.
	35. I hardly ever go wrong when I listen to my deepest gut feeling to find an answer.
	37. My snap judgments are probably not as good as most people' s.
	38. I tend to use my heart as a guide for my actions.
	39. I can usually feel when a person is right or wrong, even if I can't explain why I know.

Rationality (M = 2.30, SD = .68) and Experentiality scores (M = 3.70, SD = .59) were now reliable (Rationality2, α = .78; Experentiality2, α = .83), and independent, $r(248) = .12$, p = .05. The lack of a correlation between the REI scores for the two scales confirmed the orthogonal structure of the two factors. The low numbers of items composing the new revised scales impeded any further analyses (e.g., on subscales).

Thinking-styles types. Similarly to the proceeding adopted for unrefined REI scales, participants were dichotomized on the basis of the median scores on the new indexes (i.e., Rationality2 and Experentiality2[55]). Groups of unequal numerosities were obtained, namely the 22.2% of the sample (n = 55) was classified as low on both scales, the 16.9% of the sample (n = 42) as high on both scales, while the 30.2% (n = 75) as high in Rationality but low in Experentiality, and the 30.6% (n = 76) as high in Experentiality and low in Rationality. The fact, that now the most "extreme" thinking-style groups (i.e., those with a high score on one thinking-style and a low one on the other) were more numerous than the others could be due to the scale refinement process.

The effects on perceived probability. Analyses with the refined REI scales. An analysis of variance (ANOVA) was conducted on perceived probability, with ratio format (1 in 200, 5 in 1000, 50 in 10,000), Rationality2 (low/high), Experentiality2 (low/high)- i.e. dichotomisation of participants on the basis of the score obtained on each of the refined scales- and numeracy (low/high) as between-subject variables. As before, a significant main effect was found for the ratio format , $F(2, 224) = 6.93$, $p = .001$, $\eta_p^2 = .06$, indicating that different ratio formats evoked different levels of probability perception ($M_{1\ in\ 200}$ = 3.96, SD = 1.59; $M_{5\ in\ 1000}$= 3.41, SD = 1.29; $M_{50\ in\ 10,000}$ = 3.12, SD = 1.44)– see Fig. 4 which corresponds to Fig. 3 (i.e., when the entire REI scale had been considered). Post hoc comparisons using the Fisher's LSD test revealed that 1 in 200 was perceived as significantly different from 5 in 1000 (p = .01) and from 50 in 10,000 (p < .001), but 5 in 1000 and 50 in 10,000 were not perceived as significantly different (p = .18). Thus, even when using refined REI scales it appeared that Hypothesis1 was only partially supported by data, in that the effect found in the previous section of this chapter got confirmed (i.e., 1 in 200 was irrationally perceived as larger than 5 in 1000) and extended (i.e., 1 in 200 irrationally perceived as larger than 50 in 10,000), but not for the pair 5 in 1000 and 50 in 10,000, as indeed it had been expected. In sum, there was not a completely (and perfectly) scalar decrease in perceived likelihood as the magnitude of the numbers at the numerator and denominator of the ratios increased.

[55] Despite being associated with problems such as loss of power, the lack of symmetry in the scales (Kolmogorov-Smirnov test: p values < .05), justified the median split approach used to divide the sample- $Mdn_{\text{Rationality}}$ = 2.29, $Mdn_{\text{Experientiality}}$ = 3.75.

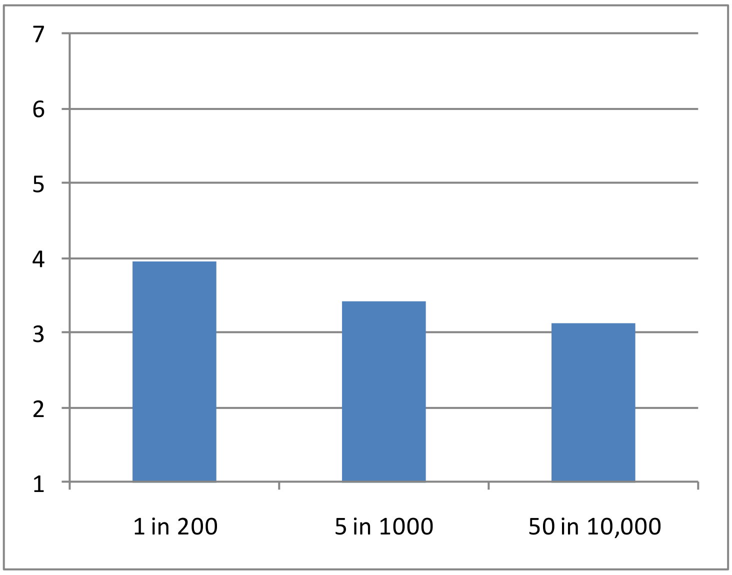


***Fig. 4*. Mean perceived probability in the three ratio format experimental conditions when the REI scale had been refined**

No main effect of Rationality2 on the basis of the revised (i.e., α- improved-) REI scale was detected on probability perception, $F(1, 224) = 2.44$, $p = .12$, $\eta_p^2 = .01$, but a main effect of Experentiality2, $F(1, 224) = 6.43$, $p = .01$, $\eta_p^2 = .03$. A higher probability perception was found in those having a high Experentiality2 score ($M = 3.65$, $SD = 1.52$) than in those having a low Experentiality2 score ($M = 3.37$, $SD = 1.44$., that will be commented in a later paragraph.

Remarkably, there was a significant interaction between the format employed in the ratio and the Experentiality score on the basis of the revised (α-improved-) REI scale, $F(2, 224) = 6.83$, $p = .001$, $\eta_p^2 = .06$- see Fig. 5.

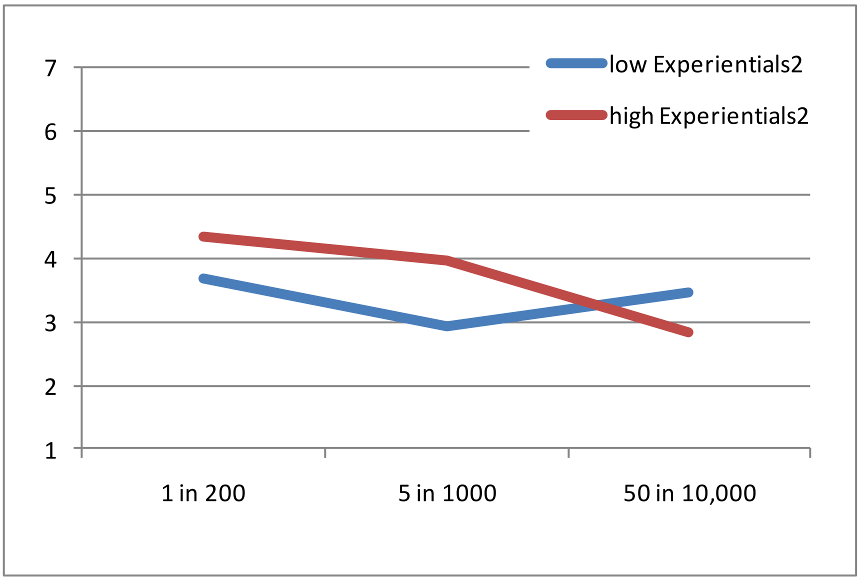


***Fig. 5.* Mean Perceived probability in the three experimental conditions as a function of participants' (low/high) Experentiality2 level**

The two-way interaction between Experentiality2 and ratio format was analyzed using simple main effect analysis. The ratio format influenced the perceived probability in both Experentiality2 levels, namely low, $F(2, 242) = 3.61$, $p = .03$, and high, $F(2, 242) = 12.79$, $p < .001$.

The significant simple main effects of ratio format were further analyzed by pair wise comparison using the Sidak adjustment for multiple comparisons. For those participants classified as having a low experientiality processing style, the significant effect on perceived probability was due only to the higher perceived probability in the 1 in 200 condition (*M*= 3.69, *SD* = 1.62) than in the 5 in 1000 one (*M* = 2.93, *SD* =.1.16, *p*= .03). Instead, perceived probability in the 50 in 10,000 condition (*M*= 3.47, *SD*= 1.44) fell between the other two conditions, but was not significantly different from either of them.

For those participants classified as having a high experientiality processing style, the significant effect on perceived probability was due to the 50 in 10,000 ratio format being perceived as lower (*M* = 2.84, *SD* = 1.41) than both the 1 in 200 (*M* = 4.34, *SD* = 1.49, $p < .001$) and the 5 in 1000 (*M* = 3.97, *SD* = 1.22, $p = .001$) ratio format conditions. Perceived probability in the 1 in 200 and 5 in 1000 conditions, though, did not appear as significantly different ($p = .60$).

Since the visual representation of the magnitude perceptions given by each group seemed anyway to suggest, in high experientials, the presence of a tendency to perceive 1 in 200 as conveying a higher probability than 5 in 1000 (i.e., in the same direction of the effect found for low experientials), further analyses were conducted. To this aim, indexes of Rationality and Experientiality were re-calculated, still on the basis of the revised scales, but only for the 1 in 200 and 5 in 1000 conditions (i.e., $n = 167$). New dichotomisations- i.e., Rationality3 (low/high), and Experentiality3 (low/high)- were computed on the basis of those indexes.

When an analysis of variance (ANOVA) was conducted for these individuals on perceived probability, with ratio format (1 in 200, 5 in 1000), Rationality3 (low/high), and Experentiality3 (low/high) as between-subject variables, a significant main effect for the ratio format was found as before, $F(1, 167) = 7.37$, $p = .007$, $\eta_p^2 = .04$, no main effect of Rationality3, $F(1, 167) = .004$, $p = .95$, $\eta_p^2 = .01$, but a main effect of Experentiality3, $F(1, 167) = 15.42$, $p < .001$, $\eta_p^2 = .09$. Remarkably, now no significant interaction between the format employed in the ratio and the Experentiality score was detected, $F(2, 167) = 2.90$, $p = .09$, $\eta_p^2 = .02$- see Fig. 6- although the p value was not that distant from the significance level.

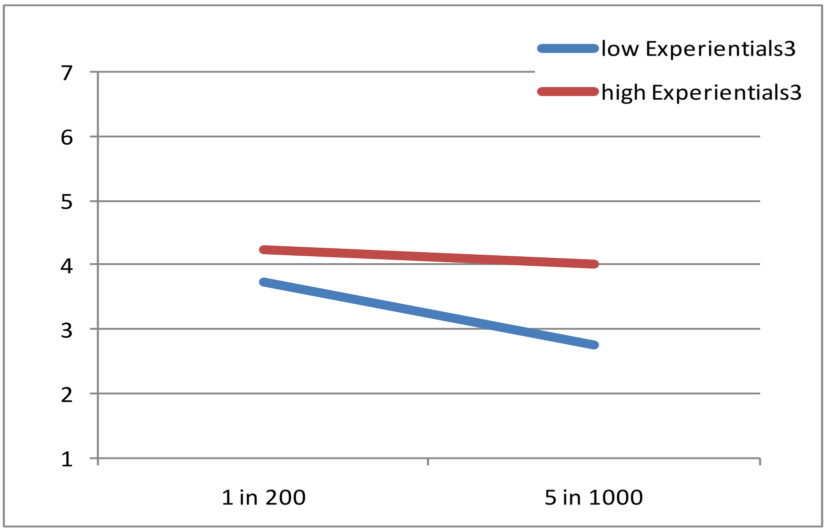


Fig. 6. **Mean perceived probability in the 1 in 200 and 5 in 1000 experimental conditions (*n* = 167) as a function of participants' (low/high) Experentiality3 level**

To confirm visual impressions generated by the figure, separate t-tests were calculated on low and high experientials which showed, that only low experientials evaluated 1 in 200 (M = 3.69, SD = 1.62) as higher than 5 in 1000 (M = 2.93, SD = 1.16), $t(92) = 2.60$, $p = .01$. Instead, the two values did not differ significantly in high experientials, $M_{1\ in\ 200}$ = 4.34 (SD = 1.49), and $M_{5\ in\ 1000}$ = 3.97 (SD = 1.22), $t(71) = 1.16$, $p = .25$.

Thus,

1) The fact the interaction among Experientiality3 and ratio format was not found when only the 1 in 200 and 5 in 1000 conditions were analysed seemed to indicate, that the significant effect obtained in the interaction in the previous analyses of variance including also the 50 in 10,000 ratio was due exclusively to the intersection of graphical lines AFTER the 5 in 1000 value. Thus, it could be affirmed, that both high and low experientials manifested the systematic tendency (despite it being non-significant among high experientials) to evaluate probability as higher when expressed through a ratio in a 1 in 200 rather than in a 5 in 1000 format.

2) For what regarded Hypothesis 1, it could be précised that the scalar decrease in probability perception expected as long as the numbers at the numerator and denominator of the ratios increased (for 5 in 1000 to 50 in 10,000) hold, but only in high experiential individuals. Furthermore, these individuals showed a tendency, albeit non significant, in line with the bias observed (i.e., 1 in 200 > 5 in 1000) in low experientials.

The influence of familiarity with the outcome at study. Differently from Study 1.2, the degree of familiarity was not assessed by means of an explicit question in this study, but a new variable was created *ex-post* that took into account individual differences on factors significant for the construct. Such variable, re-named "*familiarity*", intended to return a measure of participants' relative novelty to the situation described in the scenario. People were classified as "familiar" if at least one of the following two conditions was satisfied: having children, and having been communicated the probability of having a Down syndrome child. They were classified as "not familiar" if conditions were not both satisfied at the same time, or answers were both missing[56].

For those (n = 116, 46.8%) that could be classified as familiar with the problem a significant effect of ratio format was found on perceived probability, $F(2, 113) = 8.87$, $p < .001$, $\eta_p^2 = .14$; post hoc comparisons using Fisher's LSD test revealed that when described with 1 in 200 format, the probability was perceived as significantly higher (M = 4.42, SD = 1.73) than both when described with the 5 in 1000 (M = 3.23, SD = 1.25, $p < .001$) and with the 50 in 10,000 (M = 3.07, SD = 1.49, $p < .001$) formats. The latter two formats, though, did not raise significantly different probability perceptions. Instead, for those (n = 131, 52.8%) that could be classified as unfamiliar, no significant effect of the ratio format was detected on perceived probability, $F(2, 131) = 1.18$, $p = .31$, $\eta_p^2 = .02$. Thus, high familiarity with the outcome under study revealed a mediator of the systematic effect found in probability perception, in accordance with Hypothesis 2.

Discussion

In this study, a .005 probability expressed as "1 in 200" was subjectively evaluated as larger than when expressed as "5 in 1000", as well as larger than when expressed as "50 in 10,000". Thus, results not only corroborated the effect found in Study 1.1 and 1.2, but extended it to a denominator of an even larger order of magnitude. Nevertheless, the perception of "5 in 1000" and "50 in 10,000" did not differ. Thus, it could not be affirmed, that the relationship between the magnitude of the numbers used at the numerator and denominator of the ratios and the level of perceived probability for the ratio was negatively linear. All in all, findings seemed to imply:

1) that boundaries were present for the existence of the bias at study, thus not just whatever smaller-numbered ratio would have elicited a larger probability perception

[56] It has to be recognized that for a mistake in the construction of the questionnaire, the question on having or not been communicated by a specialist the probability of having a child with Down syndrome did not allow those who might have received such communication but did not ended up having a child due to other reasons (e.g., abortion or miscarriage, sudden death of child) a possibility to declare that without somehow appearing irrational (i.e., they appeared to have ignored the logical condition contained in the premise to that question, i.e., "If you answered yes to previous question [..]"). Only two people were in such a condition, therefore the mistake did not pose problems to the analyses

than larger-numbered ratios. If that would have been the case, then a difference in perceived probabilities conveyed by 5 in 1000 and 50 in 10,000 should have been found as well (with the first higher than the second one), but that did not happen, and

2) that since probability conveyed by 1 in 200 was judged at the same time higher than that conveyed by 5 in 1000, and than that conveyed by 50 in 10,000, thus the presence of 1 in 200 was clearly indispensable for the appearance of the bias defined so far.

Taken together, these considerations suggested that there was something peculiar in the 1 in 200 format, and that some of its features could be deemed responsible of the bias found in perceived probability. These will have been addressed in following studies.

Several problems occurred in the assessment of the REI scale, and results were difficult to interpret without a scale refinement which, anyway, dramatically reduced the number of valid items. Explorative analyses on those data highlighted, as expected, a significant association of the bias with individuals’ experientiality score, while the rational thinking-style did not show any effect at all on the bias. Nevertheless, contrary to expected findings, those high in experientiality did not show a larger bias than those low in experientiality; on the contrary, the bias appeared more established among low experientials. Result, awaiting further confirmation, that highly experientials perceived a difference between 5 in 1000 and 50 in 10,000 judging the first expression as conveying a larger probability than the second, could be explained taking into account the possibility, that those high in experientiality might have not actually elaborated the ratio expressions. Instead, they could have intuitively assessed the magnitude of the ratio expressions taking the size of the denominator as a hint.

Findings on the degree of individuals’ numeracy as assessed by the SNS supported previous results (i.e., Study 1.1) obtained by means of an objective measure of the construct, namely that only highly experientials showed the bias.

Finally, for what concerned familiarity with the outcome at study, an important moderation of the effect was shown, as only individuals so-defined “familiar” with the scenario showed the bias in perceived probability.

Study 2.2

Goals

Findings of the previous study indicated that a .005 probability expressed as "1 in 200" was at the same time subjectively evaluated as larger than when expressed as "5 in 1000", as well as larger than when expressed as "50 in 10,000".

Despite the rather interesting and surprising nature of the phenomenon described, it had to be recognized that results were contingent to specific ratio values. It could well be, that ratios whose numbers at the numerator and denominator were, for some reasons, peculiar, had been chosen by chance, and that when employing different numbers the same effect would not be found. In the conclusions of the past study, it has been pointed at possible specific features peculiar of the 1 in 200 ratio that could be deemed responsible of the bias found in perceived probability. One of such features could have been the type of format in which the probability had been stated: indeed, while "1 in 200" was in a "1 in X" format, the other two ratios examined had an "N in NX" format. Could it be possible that the "1 out of some number" format might have explained the increased level of probability found for the "1 in 200" ratio when compared to the other ratios expressed in different formats? In other words, was a ratio in a "1 in X" format generally perceived as higher than a ratio in an "N in NX" format? In order to test that, and at the same time to reject the chance, that the effect in probability perception would not be found when ratios different from 1 in 200 and 5 in 1000 (or 50 in 10,000) were employed, it was deemed necessary to attempt a generalization of results to other conditions. In other words, it was necessary to compare the probability perception elicited by "1 in X" formats (also different from the one tested until that moment) with that elicited by corresponding equivalent "N in NX" formats (also more variants than the two tested until then).

In this experiment, ratios with denominator 1000 were chosen as "N in NX" formats, for mainly two reasons. First of all, because the "Y in 1000" format represented the structure of the ratios where the effect had been found until then. Secondly, ratios with 1000 at the denominator are frequently used mainly in institutional communications (e.g., sanitary campaigns; scientific reports, but also in doctor-patient interactions, e.g., to describe the incidence of a disease in the population). Thus, NX was equated to 1000 in the present study. Only with the aim of simplifying the description, in this study it will be referred to "Y in 1000" to mean those proportions having 1000 as denominator, even if Y would not be an adequate term[57].

Drawing from results obtained in previous studies, it was predicted that a probability would be judged as higher when the ratio expressing it be in a "1 in X" format rather than in a "Y in

[57] In fact, as each pair wise comparison of proportions vs. rates corresponds to the same likelihood value, and the rate is always a not simplified version of the proportion, the two expressions should always share a common variable term (i.e., X, in the cases illustrated above).

1000” format (i.e., a rate with 1000 as denominator). Therefore, the first specific hypothesis for this study was:

Hypothesis 1: A probability expressed in a “1 in X” format is perceived as greater than when expressed in a “Y in 1000” format.

Moreover, in accord with Study 1.1, the same trend was expected for worry. Thus, a further hypothesis to be tested in the Study was:

Hypothesis 2: A probability expressed in a “1 in X” format induces greater worry than when expressed in a “Y in 1000” format.

Furthermore, an additional hypothesis was formulated regarding the degree of comprehension of the two formats. Following CEST’s second principle (i.e., the small-numbers effect- see for instance, Pacini and Epstein, 1999-b), prediction should have been, that ratios employing smaller numbers would have been more easily understood than ratios employing larger numbers, exactly because of their general higher clarity. Such a prediction, nevertheless, was unbearable for mainly two reasons. First of all, as it had not yet been demonstrated (and indeed, the opposite seems true), that people reason on ratios the same way they do on integer numbers. Transferring a rule accounting for how people evaluate integers to the case of fractions was in principle not seen as correct. Indeed, numerous findings stemming from the literature on mathematical cognition point to the fact, that a) people process numerical information expressed through a ratio in a different way to information expressed in integer numbers (see, for instance, Bonato, Fabbri, Umiltà, and Zorzi, 2007; Reyna and Brainerd, 2007), and b) that fractions pose more difficulty than integer numbers (same sources, but see also the previous study, i.e., 2.1).

In second instance, the prediction was unbearable because supporting the hypothesis of a greater clarity of ratios employing smaller rather than larger numbers seemed hard in the specific case at study: results had showed an opposite tendency to denominator neglect for ratios when in single presentation conditions in all experiments described up to the present one, hence apparently disconfirming the applicability of the explanation in terms of Ratio-bias to the case under study. Moreover, some clear evidence in the literature are present arguing for a lower comprehension of proportions rather than of rates (see results described in the first paragraph of Chapter 2, e.g., Grimes and Snively, 1999; van Vliet, Grimes, Popkina, and Smith, 2001).

All these considerations leaded to a forecast opposite to the one stemming from CEST, namely that lower comprehension had to be expected for smaller-numbered ratios than for larger-numbered ones. Hence, the hypothesis predicted, that ratios expressed in (comparatively) smaller numbers (schematised as in “1 in X” format) should have been less understood than ratios using (comparatively) larger numbers (i.e., in “Y in 1000” format). Such hypothesis read:

> **Hypothesis 3: Ratios in “1 in X” format are less understood than ratios in “Y in 1000” format.**

Method

Participants. A total of 254 pregnant women attending the maternity wards of San Raffaele Hospital in Milan (Italy) were contacted by the experimenter during their first visit at the gynaecologist. Of these, 54 (21.2 %) refused to take part in the study, while 200 accepted. The reasons for refusals were generally not specified or brought back to time constraints and to the desire to avoid mathematical tasks. The sample therefore included 200 pregnant women but 1 was excluded for providing incomplete data, thus additionally reducing the sample to 199 participants.

Mean age of participants was 32.77 years, ranging between 18 and 47 years. Most of the women had completed high school (44.8%) or had already a University degree (43.3%), while only few of them (11.9%) had completed the lowest education level in Italy. The 59.6% of them were at their first pregnancy, while the remaining 40.4% of participants had already one or more children (one participant did not answer to this question).

Design and material. In the questionnaire, a table was presented summarizing the risk of having a child affected by Down syndrome according to the maternal age (from 18 to > 51 years), where the maternal age had been classified into twelve 3 year- span groups (e.g., 18-20, 21-23)[58].

The risk was expressed by means of a verbalized ratio, whose superficial format was the single factor manipulated in the study. Indeed, the ratio appeared in the entire table either as “1 in every X births” or as “Y in every 1000 births”- see the Appendix for the two tables employed in the material. Therefore two versions of the same questionnaire were elaborated, one where numerical risks were expressed in “1 in X” format and the other where they were expressed in “Y in 1000” formats. The table served as a reference source for the woman to identify her age-related risk; after having identified it, she was asked to circle it using a pen.

Three dependent variables were assessed: woman’s perceived risk of having a child affected by Down syndrome given her age, the degree to which she would find this information worrisome, and a construct that measured comprehension of the numerical format.

[58] Similar intervals have been used, for instance, even in a landmark study (Hook,1981) to indicate the rates of all clinically significant cytogenetic abnormalities in live births by 1-year- maternal age intervals ranging from ages 15 to 49.

The degree of risk had to be judged on a graphical 7-point scale with the two extremes labelled as “not at all high” and “extremely high”, respectively. The scale was not the typical Likert scale, but a visual scale created on purpose for the study, with adjacent steps whose dimension was getting steadily bigger and the colour darker as moving towards the right (to symbolize an increasing quantity of the dimension assessed)- see the Appendix.

Perceived worry was measured asking participants to answer the question, “How worried are you concerning the risk you have just circled?”. Alike the previous question, answers had to be provided on a 7-point graphic scale, ranging from “not worried at all” to “extremely worried”.

Finally, participants were asked to answer to a set of 6 questions created on purpose for the study, and intended to assess comprehension of the format stating the numerical probability value circled in the table -see the results section for the list of questions along with the answers available according to the experimental condition.

Procedure. Participants were presented with a paper version of the questionnaire- they were randomly assigned to one of the two conditions stating the risk as a “1 in every X” or as a “Y in 1000 births”. They completed the questionnaire individually and at their pace, namely without any time constraints.

Ethical approval. The study received the approval of the Clinical Ethical Committee of San Raffaele Hospital. All participants answering the questionnaire had preliminarily signed an informed consent form describing both the purpose and the method of the study, and guaranteeing them anonymity and confidentiality in the treatment of data.

Results

As it is visible from Table 1, three risk classes had no participants (i.e., the 18-20, and the 48-50, and over 51)[59]. The mode maternal age class was 30-32 years, while the median was still in the same class (*Mdn* = 32.50). Distribution of participants in the 9 risk classes according to the experimental condition is illustrated in Fig. 1[60]. The number of participants per risk class was equally divided among experimental condition, $\chi^2(8, N = 199) = 7.21$, $p = .51$.

The “1 in X effect”. A multivariate analysis on perceived probability and perceived worry with ratio format as between-subjects variable showed a global effect of ratio format, $F(2,195) = 3.92$, $p = .02$, $\eta_p^2 = .04$. Nevertheless, while the impact of ratio format on the probability variable was significant, $F(1,196) = 7.61$, $p = .01$, $\eta_p^2 = .04$, the same could not be affirmed for the worry measure, $F(1,196) = 1.93$, $p = .17$, $\eta_p^2 = .01$.

[59] These data can be considered coherent with the actual trends in human reproductive age.
[60] The “1 in every X” ratio format condition was presented to slightly more than half sample (*n* = 101 women), while the “Y in every 1000” ratio format was presented to 98 women.

The mean subjective magnitude of the probability was 3.57 (*SD* = 1.45) when expressed as "1 in X", and 3.03 (*SD*=1.37) when expressed as "Y in 1000". Hence, overall it could be affirmed that Hypothesis 1 was supported by data. The systematic tendency to evaluate a probability expressed in a "1 in X" format as greater than the same probability expressed in a "Y in 1000" format was dubbed the "*1 in X effect*".

When looking at the data divided by mother's age, as in Table 2, it can be noticed, that all the means, except two, went in the predicted direction, even if, due to reduction in sample sizes, only two reached significance (1 in 475 vs. 2.1 in 1000, and 1 in 795 vs. 1.3 in 1000).

Table 1
Distribution of participants in the 12 risk classes

Maternal Age	No. of participants
18-20	0
21-23	4
24-26	14
27-29	25
30-32	57
33-35	43
36-38	37
39-41	10
42-44	7
45-47	2
48-50	0
Over 51	0

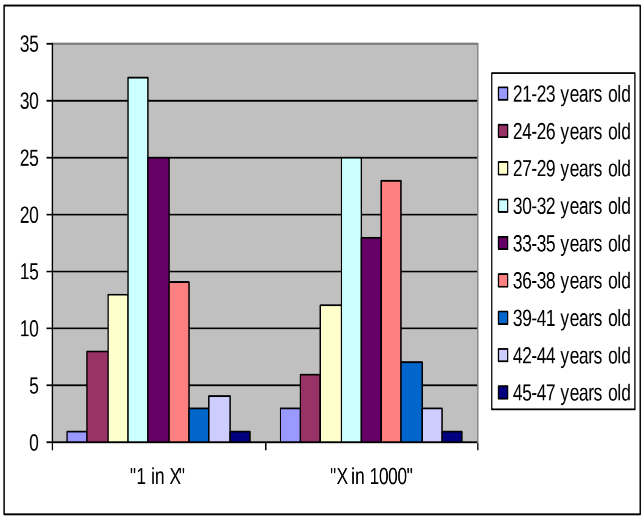


Fig. 1. **Participants' age class distribution between the two experimental conditions**

Table 2
Results of the paired Mann-Whitney tests on equality of the means between the two experimental conditions for each valid risk class

Mother's Age	1 in X format	n	Y in 1000 format	n	Mean Perceived risk 1 in x	Mean Perceived risk Y in 1000	Mann-Whitney p value (2-tail)
18-20	1 in 1540	0	0.6 in 1000	0			
21-23	1 in 1480	1	0.7 in 1000	3	**2.00**	1.67	1.00
24-26	1 in 1350	8	0.8 in 1000	6	**2.25**	2.17	.852
27-29	1 in 1120	13	0.9 in 1000	12	**2.92**	2.08	.406
30-32	1 in 795	32	1.3 in 1000	25	**3.38**	2.80	.033
33-35	1 in 475	25	2.1 in 1000	18	**3.88**	2.78	.001
36-38	1 in 240	14	4.2 in 1000	23	**4.14**	3.65	.077
39-41	1 in 110	3	9.1 in 1000	7	4.00	4.14	.833
42-44	1 in 49	4	20.4 in 1000	3	**6.00**	5.67	.629
45-47	1 in 21	1	47.6 in 1000	1	4.00	4.00	1.00
48-50	1 in 8	0	125 in 1000	0			
> 51	> 1 in 6	0	> 166.7 in 1000	0			
		101		98	3.57	3.03	.007

For what regarded perceived worry, overall a mean assessment of 3.95 (SD = 1.69) was found for "1 in X" formats, and a mean of 3.60 (SD = 1.87) for "Y in 1000" formats. Perceived worry in the two conditions showed the same tendency of perceived probability, despite not resulting in a significant difference. However, the correlation between the two variables was significant (r_{199}= .65, p < .001), thus, it could be affirmed that data indicated even a corroboration of Hypothesis 2.

Comprehension of the format. Participants' overall performance on the comprehension inventory is presented in Table 3 according to each of the six questions. The general difficulty demonstrated in answering apparently simple questions (see the large number of participants- 50% or more- who found answering to questions number 1, 4, 5, and 6 hard) further corroborated the idea that laypeople (i.e., patients) experience difficulties in understanding and grasping the meaning of numbers, even more when these are not integer. Two of those questions, in particular, (i.e., number 1 and 5) involved calculations which are normally required for making sound health decisions (see, for instance the case where patients are communicated the rate of success of two different treatments, and have to opt for one of them, or that where they are in need to compare rates of success of two alternative treatments). In both cases, a sound answer would require that one correctly handles the rules of comparisons between ratios; question number 5, moreover, implies exactly the mastery of that skill which could prevent people from falling prey to the "1 in X" effect. The fact, that only the 37% of the sample could solve the problem posed by that question correctly, is undoubtedly of a certain relevance.

Table 3
Overall participants' performance in each of the six items of the comprehension scale

Items	% correct
1. The risk you circled above is: higher/lower a. higher than 1 in 250 b. lower than 1 in 250	40
2. Compared to the risk of a 23 years old woman, the risk you circled above is: a. higher b. lower c. equal	88
3. The risk you circled above means that the probability of having a child affected by Down syndrome is: a. over 50% b. below 50%	86
4. Imagine that the risk of another woman is double than the one you circled above. What would it be? ___________ in ____________	49
5. The risk you circled above is equal to the risk of _______ in 1000 (for condition "1 in X") 1 in ________ (for the condition "Y in 1000")	37
6. What is the probability that your child will NOT be affected by Down syndrome given your age?	35

Overall, mean comprehension score was 3.36 (*SD* = 1.38), while the median was 3 out of 6 possible correct answers.

The distribution of participants according to the number of correct answers they could give in the six-item inventory measuring comprehension is represented in Fig. 2, separately showing participants who were in a "1 in X" condition from those who were in a "Y in 1000" condition. However, it was thought, that a median split on the comprehension index was not appropriate in the present case as participants addressed non comparable inventories, and the fact participants have answered to the same inventory is a necessary prerequisite for such type of analysis. Some exploratory analyses were made, though.

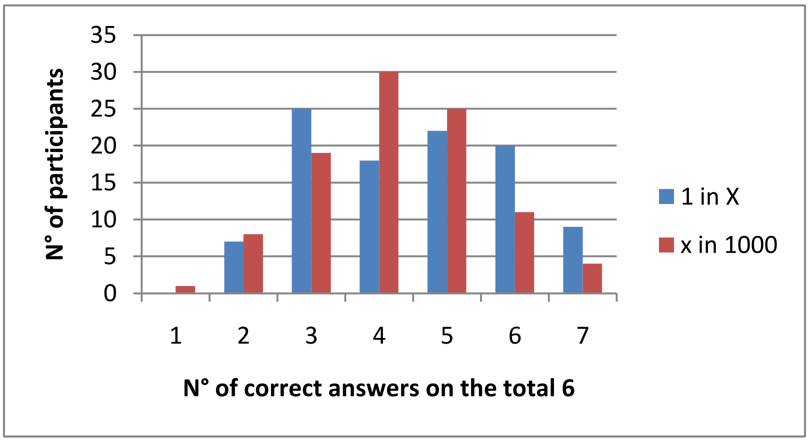


Fig. 2. **Distribution of participants according to the performance in the six-item comprehension inventory in the two experimental conditions**

Overall, a test on the equality of the means between the two groups returned a non-significant difference between the mean number of correct answers given, $t(197) = 1.44$, $p = .15$. Thus, from these data it seemed that Hypothesis 3 was not supported. However, it was also true, that the null effect did not allow to state the opposite trend to the one tested and hypothetically stemming from CEST's second facet for Ratio-bias explanation, namely that "1 in X" formats would be more comprehended than "Y in 1000" formats,

In Table 5, percentages of correct answers by question are reported according to experimental conditions. Further explorative analysis segmented by question showed that the two conditions did not differ significantly in the number of participants answering correctly for any of the six questions, respectively [$\chi^2(1, N = 199) = .965$, $p = .33$] for question number 1, [$\chi^2(1, N= 199) = 1.41$, $p = .24$] for question number 2, [$\chi^2(1, N = 199) = 1.72$, $p = .19$] for question number 3, [$\chi^2(1, N = 199) = 1.12$, $p = .29$] for question number 4, [$\chi^2(1, N =199) = 2.55$, $p = .11$] for question number 5, and [$\chi^2(1, N = 199) = .79$, $p = .38$] for question number 6.

Table 5

Percentages of participants' correct answers to each of the six questions in the two experimental conditions

Items	% correct	
	"1 in X"*	"Y in 1000"°
1. The risk you circled above is: c. higher than 1 in 250; d. lower than 1 in 250.	43.5	36.7
2. Compared to the risk of a 23 years old woman, the risk you circled above is: a. higher; b. lower; c. equal.	91.1	85.7
3. The risk you circled above means that the probability of having a child affected by Down syndrome is: a. over 50% b. below 50%	89.1	82.7
4. Imagine that the risk of another woman is double than the one you circled above. What would it be? ___________ in ____________	45.5	53.1
5. The risk you circled above is equal to the risk of _______ in 1000 (for condition "1 in x") 1 in ________ (for the condition "Y in 1000")	42.6	31.6
6. What is the probability that your child will NOT be affected by Down syndrome given your age?	37.6	31.6

* = originally out of 101 participants; ° = originally out of 98 participants

Discussion

Results obtained in the present study demonstrated that the bias in probability perception extended well beyond the specific numerical values considered in Section 1 of the present chapter (i.e., 1 in 200 and 5 in 1000) and in Study 2.1 (i.e., 50 in 10,000). Individuals showed a tendency to perceive a probability as higher when expressed through a proportion (i.e., in a "1 in X" format) than through a rate whose denominator was 1000. The tendency was further confirmed by a similar, albeit not statistically significant, trend in the degree of perceived worry, as showed by a positive correlation with the perceived probability measure.

Also, overall participants revealed a low degree of comprehension of the meaning of ratios, as a high percentage of them made mistakes on relatively simple mathematical questions. The hypothesis made on comprehension, namely that ratios in a "1 in X" format employing (comparatively) smaller numbers are less clear than ratios in a "Y in 1000" format employing (comparatively) larger numbers, did not receive support in the present study.

Study 2.3

Goals

In the previous study, a "1 in X effect" had been established, as probability statements in a "1 in X" format increased the probability perception compared with probability statements in "N in NX" formats, when NX had been fixed to 1000. What about the effect when the ratio with an "N in NX" format had a denominator equal to 100? Indeed, a generalization of results of Study 2.2 to ratios whose denominator be in the teens-order of magnitude or in the hundreds-order of magnitude would have not been, in principle, legitimate as those values had not been directly compared. However, knowing whether people would still be prone to the bias when rates in those formats were used was judged interesting for at least three reasons.

1) There are several cases, where a doctor might employ a smaller denominator (e.g., 100) than those that had been analyzed until then in communicating a probability. An expression using such numerical information frame corresponds to "the typical size of human groups in a wide range of ancient and modern human societies" (Dunbar, cited in Garcia-Retamero and Galesic, in press-b, p. 1), and as such it can be thought as more likely perceived as plausible. Plausibility is a determinant feature to generate reliance on the data. A hundred, 1000, and 10,000 are probably the most frequently used denominators in communication of probability by means of ratios. By investigating the order of magnitude of the first one the three formats, all of them would have been covered as previous studies already investigated the second and third format.
2) By validating applicability of findings in the missing (i.e., 100) condition, boundaries of the bias would have been further specified.
3) Ratio-bias, that is mainly a within-subject phenomenon (Garcia and Tor, 2009), has mostly been found with two ratio values, namely 1 in 10 and 10 in 100 (that is, when participants compared a rate out of a 10-denominator with the equivalent proportion out of a 100-denominator). Testing similar ratio values in a within-subjects design would have further examined whether the type of presentation (apart from the type of task) might be responsible for the opposite results found in the present dissertation.

Therefore, the first specific hypothesis for this study was:

Hypothesis 1: A probability expressed in a "1 in X" format is perceived as higher than when expressed in a "Y in 100" format or similar magnitude.

Furthermore, object of interest was constituted by the possible influence of ratio format on intentions to perform amniocentesis. Behavioural intentions have been indicated by some scholars (see, for instance, the Health Belief Model, e.g., Rosenstock, 1966) as a better indicator of people's attitudes. The assumption behind the second research question was, that behavioural purposes depend on the degree of probability perceived for the condition to be tested for. Thus, the second hypothesis for the study was:

> **Hypothesis 2: A probability expressed in a "1 in X" format induces more participants to hypothetically choose to perform amniocentesis than when expressed in a "Y in 100" or similar magnitude format.**

Furthermore, the explorative analysis on the degree of comprehensibility of the two formats continued also in the present study. The same hypothesis as previous study was stated, only it was modified to apply to the case of a "1 in X" versus "Y in 100" or similar format. It read,

> **Hypothesis 3: Ratios using (comparatively) smaller numbers in a "1 in X" format, are less comprehended than ratios using (comparatively) larger ones in a "Y in 100" or similar magnitude format.**

Method

Participants. A total of a hundred individuals (58 women and 42 men) from the general population took part in the experiment voluntarily. Mean age was 38.2 yrs. (SD= 12.4), varying between a minimum of 19 to a maximum of 69.

Design and material. Participants were asked to read a very similar scenario to that of all preceding studies (apart from Study 1.1). The scenario reported the case of Anna, a 48 years-old mother-to-be who, after having been waiting for long, became pregnant; at the gynaecologist, she got informed of her the age-related probability of having a Down syndrome-affected child. Object of experimental manipulation between-subjects was, as in previous studies, the format of the ratio expressing such probability. A proportion format whose denominator was in the teen-order of magnitude (this time: 1 in 12) was compared with a rate format whose denominator was in the hundred-order of magnitude (this time: 10 in 120). The round ratios (1 in 10 and 10 in 100) were judged too easy and were avoided in this study to minimize the likelihood that participants could transform the ratio into another one, which would lead to the impossibility of controlling for format effects. Two different questionnaire versions were created to follow the experimental design.

After having read the scenario, participants were asked to role-play and imagine themselves as Anna, the prospective mother described in the scenario. Within this context, they answered two questions. The first question tested the subjective evaluation (when role-playing the protagonist) of the magnitude of the stated probability of having a child with Down syndrome. Participants had to provide an estimation on an 11-point Likert scale whose extremes were labelled as "extremely low" and "extremely high", while the half-way point was labelled as "moderate".
The second question assessed participants' intentions to perform a diagnostic test (i.e., amniocentesis, able to tell with certainty whether the baby had Down syndrome). That invasive exam was described in both experimental conditions as carrying a fix additional risk of abortion (i.e., the same 1% risk in both conditions). Answer was categorical (i.e., yes/no).
Finally, to assess comprehension for each format, a different inventory than the one employed in the previous study was used in the present one. A set of 5 questions was created on purpose for the study. Among other abilities, skills assessed by questions were that of being able to put the ratio value in correct comparison with other probability values, and that of performing some mathematical operations that are commonly required in everyday life (e.g., redoubling a risk value, transforming it into a different mathematical format).

Procedure. Each participant was randomly assigned to one of the two conditions constructed to follow the experimental design. Individuals could complete the questionnaire at their pace, without any time constraints.

Results

For what concerned perceived probability, mean ratings were greater for the 1 in 12 (M = 7.48, SD = 2.80) than for 10 in 120 format (M = 6.18, SD = 2.69), $t(98) = 2.37$, p = .02- see Fig. 1. Results corroborated the first hypothesis as when evaluating a probability expressed through a numerical ratio people perceived a proportion in a "1 in X" format with a 10 denominator as conveying a larger probability than a rate having 100 as denominator.

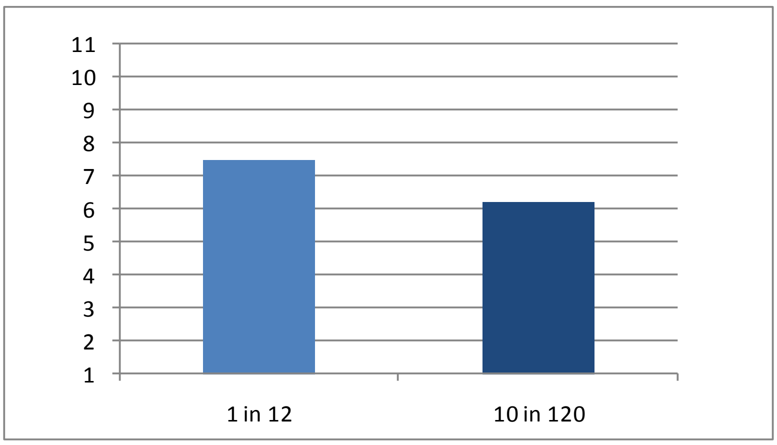


Fig. 1. **Mean perceived probability in the two experimental conditions**

For what concerned the second dependent variable, instead, no effect of the format of the ratio (smaller/bigger numbers) was detected on behavioural intentions. Participants' will of performing the diagnostic test was always high and did not change when the probability of having a child with Down syndrome was stated as 1 in 12 or 10 in 120, $\chi^2(1, N = 100) = .33$, $p = .56$, likely because the probability of .083 was perceived high (even in the "Y in 100" format) enough to consider doing an amniocentesis as a sort of binding option- see the distribution of answers in the two conditions in Table 1.

Table 1
Distribution of choices for the intention to perform a risky diagnostic test (i.e., amniocentesis) in the two experimental conditions

	Perform test?		
Condition	yes	no	Total
1 in 12	42	8	50
10 in 120	44	6	50
Total	86	14	100

For what concerned the third variable (i.e., comprehension), exploratory analyses were made that did not indicate differences across experimental conditions. Overall, the mean and median numbers of correct answers in the total four answers analysed[61] were 3.36 (*SD* = .88) and 4, respectively, thus very high. Table 2 displays the distribution of the number of correct answers according to experimental condition.

Table 2
Number of correct answers per experimental condition to each item of the numerical comprehension inventory

Item n°	1 in 12 (*n* = 50)	10 in 120 (*n* = 50)
1	41	45
2	46	50
4	41	38
5	33	42
Total	161	175

[61] Answers to four items only were analyzed. Unfortunately, in question number 3 of the 10 in 120 condition, the small number ratio format was erroneously employed instead than the large number ratio format. Because of that mistake, answers on that item in both experimental conditions were excluded from analysis.

Visual inspection of the table immediately suggests that the manipulation of the ratio format did not influence the number of correct answers participants could give on the inventory. The impression was confirmed by an overall test on the equality of the means between the two groups, that returned a non-significant difference, $t(98) = -1.60$, $p = .11$. Once again, data did not seem to indicate differences in the degree of comprehensibility of the two formats. Thus, Hypothesis 3 was again not confirmed by data.

Discussion

Results of the present study extended the bias in perceived probability to ratios with a denominator in the teen and hundred order of magnitude. That allowed to generalize previous affirmation in the following way: a ratio in a proportion format (i.e., “1 in X”) is normally perceived as conveying a higher probability than the equivalent ratio in a rate format (i.e., “N in NX”).

The use of ratios of different superficial appearance did not seem to influence people’s intentions for choice, as the number of people declaring the will of performing amniocentesis was always high, and did not change according to experimental condition. Also, the hypothesis on a supposed higher difficulty of the “1 in X” format in respect to “N in NX” ones did not receive support in the data, as participants showed the same degree of comprehension for both formats. The absence of a difference in comprehensibility of the two formats confirmed results obtained in the previous study with a different inventory; findings did not bear support to results in the literature (see, e.g., Grimes and Snively, 1999) suggesting an advantage of rates over proportions in terms of numerical comprehension. However, the absence of an effect could have been the result of the questions chosen to measure comprehension. These questions should be reviewed and ameliorated.

Study 2.4

Goals

Until now, results seemed to converge on a specific effect of the “1 in X” format. One last possibility, though had to be ruled out. In the experiments, frequently (and trivially so) the “1 in X” format featured a smaller denominator than any other format (e.g., “N in NX”).[62] The fact that participants perceived “1 in X” as larger than “N in NX” might thus have reflected a general focus on the denominator of the ratio[63], rather than a specific effect of the “1 in X” format, and people’s tendency to provide lower probability judgments as the number of people exposed to a threat increases (Yamaguchi, 1998), that is when the denominator increases. However, if the effect in probability assessment was due to a phenomenon like numerator neglect (i.e., to the magnitude of the denominator and not to the “1 in X” format), the same effect should be found even when comparing ratios not featuring 1 at the numerator. In particular, a higher probability perception should be found for the ratio having the smaller numbered-denominator than for the ratio having the larger numbered-denominator. However, it was instead believed, that the effect in perceived probability was not due to participants focusing on the magnitude of the denominator, but to the specific format employed (i.e., the “1 in X” format). In line with this prediction, the hypothesis (to be disconfirmed) for the present study was:

> **Hypothesis: In line with the numerator neglect hypothesis, people give a higher judgment of the same probability when expressed in a “N in N*X” format than in a “N_1 in N_1*X” format, if $N*X < N_1*X$ and both $N \neq 1$ and $N_1 \neq 1$.**

Method

Participants. A total of a sixty-six individuals[64] from the general population took part in the experiment voluntarily. Mean age was approximately 32 yrs. (SD = 12), varying between a minimum of 18 to a maximum of 68. Of the sample, 46.2% (n = 30) had a high school diploma, 44.6% (n = 29) a University degree/ a higher title, while the others had a low education levels (“licenza media” and “elementare”)- one person did not report this information. The majority of participants (72.7%, n = 48) declared not having children, while the 27.3% of them (n = 18), had one or more children.

[62] The only exceptions were some of the ratios employed in Study 2.1, precisely those defining the mother’s age-classes from 27-29 downwards (e.g., 1 in 1480 vs. 0.7 in 1000).
[63] As people might have realized that the division of a number for a sensitively “large one” returns a “small number”.
[64] For a mistake, the questionnaires did not contain a question assessing the gender of respondents, therefore such variable could not be controlled for.

Design and material. A scenario similar to that employed in all preceding studies (apart from Study 1.1) was used, describing a couple being informed by the gynaecologist on the maternal age-related risk of having a child with Down syndrome. The probability, expressed through a numerical ratio, could use both comparatively smaller or comparatively larger numbers at the numerator and denominator, caring that in none of the formats the number 1 would be employed at the numerator. Thus, in one version, the probability was expressed as 3 in 48 (comparatively smaller numbers), while in the other as 10 in 160 (comparatively larger numbers).
After having read the scenario, participants read instructions asking them to role-play and imagine themselves as one of the two prospective parents described in the scenario. Within that context, they rated the magnitude of the numerical probability of having a child with Down syndrome. They had to provide estimation on a graphical 7-point scale with the two extremes and the central point labelled as "low", "high", and "moderate", respectively. The scale was similar to that employed in Study 2.4- see Appendix.

Procedure. Each participant was randomly assigned to one of the two conditions constructed to follow the experimental design. Individuals could complete the questionnaire at their pace, without any time constraints. Anonymity and confidentiality were ensured to them.

Results

Mean ratings of subjective probability did not significantly differ when the probability was expressed as 3 in 48 (M = 3.88, SD = 1.87) or as 10 in 160 (M = 3.67, SD = 1.74), $t(64)$= .48, p = .64. Thus, Hypothesis 1 was not supported by data.

Discussion

Results showed, that the probability perceptions of two equivalent ratios, none appearing in a "1 in X" format, but using denominators of different magnitudes (i.e., 48 < 160) did not differ significantly. Hence, the tested hypothesis was not supported by data; however, the null result (given the rules of inferential statistics) did not allow excluding, that numerator neglect was anyway present. However, the fact, that the hypothesis in line with numerator neglect was not bear by findings, in any case indicated that results obtained in all experiments presented so far could not in principle be explained by an overlook of the numerator. Anyway, these results did not allow drawing ultimate conclusions on the fundamental function of the number 1 at the numerator of the ratio to obtain the effect of higher probability perception described. With that aim, a further experiment had to be carried out (see Study 2.6) in order to perform a direct test of the necessity of the "1 in X" format to generate the effect at study.

Study 2.5

A special thanks to Alessio Sperlinga and Mariarita Barisione for having advertised our study on their web-sites[65]

Goals

Is the "1 in X" format really peculiar? What is/are the specific feature/s of the "1 in X" format that can be deemed responsible for the higher perceived probability found compared to ratios in "N in NX" formats?

Aim of this short study was that of investigating whether a ratio value prompted a different degree of mathematical operations when expressed in a "1 in X" format rather than in its equivalent "N in NX" format. In particular, the mental operation analyzed was transformation[66], and, differently from previous studies, ratio values were not contextualized but they were pure mathematical quantities. Potential goal of this investigation was that of isolating some peculiarities of the two types of ratio formats. Indeed, intention was that of employing those expressions in a further study where they would have conveyed a probability and they would have been linked to the usual outcome (i.e., having a Down syndrome affected child)- see next study.

It was predicted that when dealing with the "1 in X" format a large number of people would have not operated a transformation, due to its particular character of expressing an irreducible fraction[67]. Instead, the opposite was expected for the "N in NX" formats, because of expressing reducible fractions[68]. The hypothesis of the study was,

Hypothesis: People do not tend to operate a transformation on a ratio expressed in a "1 in X" format, but they do tend to transform ratios expressed in "N in NX" formats.

Method

[65] The websites were http://www.bambini.it/ and http://www.ilmondodeigemelli.org/, respectively.

[66] In Mathematics, a transformation is an equivalent change in an expression or equation resulting from the substitution of one set of variables by another.

[67] An irreducible fraction (or fraction in lowest terms or reduced form) is a vulgar fraction in which the numerator and denominator are smaller than those in any other equivalent vulgar fraction. It can be shown that a fraction a/b is irreducible if and only if a and b are coprime, that is, if a and b have a greatest common divisor of 1 (Dodero, Baroncini, and Manfredi, 2001, p. 55).

[68] As the structure itself shows, i.e. "N in NX" = N(1 in X), "N in NX" ratios are reducible to ratios in a "1 in X" format by dividing both numerator and denominator by N.

Participants. Fifty-eight individuals[69] from the general population took part in the study voluntarily. The data of nine individuals had to be eliminated due to some problems[70], thus the sample employed for analyses comprised 49 people (29 females and 19 males- one person did not report the gender).

Design and material. The study manipulated a single factor, namely the format of the numerical ratio expressing the same probability (i.e., .04), in a within-subjects design. The ratio was stated in four variants, where one had a "1 in X" format (i.e., 1 in 25), while three had an "N in NX" format (i.e., 0.4 in 10, 2 in 50, and 10 in 250). The order in which the four variants appeared in the questionnaire was fully randomized thus to control for order effects.
For each format of the ratio, people were asked to report whether they were used to transform it (yes/ no answer) - a short example of what was meant with "transformation" was given at the beginning of the page.

Procedure. An electronic version of the questionnaire was accessed by participants who were randomly assigned by the program to one of the 24 conditions constructed to fully control order effects- a minimum of 2 participants per condition was considered necessary in each order condition. Individuals were allowed an unrestricted amount of time to complete the questionnaire by themselves.

Results and discussion

Contingencies on whether participants to the study were used to transform the ratios proposed, or instead they normally used to consider them the ways they were presented, are shown in Table 1.

[69] Mean age was 31.4 years (*SD* = 5.9) ranging from 22 to 46. Of the sample, 41 individuals (85.4%) had a high education level, with 24 (50%) having a (3/5 Years) University degree, and 17 (35.4%), a higher title like Master, PhD, or Specialization school diploma, while the 14.6% of them (*n* = 7) had a high school diploma- one person did not disclose information on this question.
[70] Some participants accessed more than one survey, thus data of the second or third one (individuated through data and time of access automatically registered by the program) were erased and only data of the first one completed were retained.

Table 1
Frequencies (% in parentheses) of participants declaring to be used to transform/not transform each of the four proposed formats, along with the binomial test results of equivalence of distribution

	Do you transform?			
Ratio Format	yes	no	Total	Binomial Test
0.4 in 10	32(65.3)	17(34.7)	49	$p = .04$
2 in 50	33(67.3)	16(32.7)	49	$p = .02$
10 in 250	39(79.6)	10(20.4)	49	$p < .001$
1 in 25	6(12.2)	43(87.8)	49	$p < .001$
Total	110(56)	86(44)	196(100)	

Overall people showed a tendency to operate transformations on the ratios more frequently (i.e., 56%) than not operating them, even though the difference was only almost significant, χ^2 (1, N = 196) = 2.94, p = .08. The number of participants who declared to transform a ratio format and that of those who declared not to transform it differed significant for every ratio format examined, as showed by results of the binomial tests reported in Table 1.

Results of the six pair wise comparison using Bonferroni correct p[.05/6=] .0083 were performed whose results are reported in Table 2.

Table 2
Results of the Mc-Nemar tests performed for each pair wise comparison in the four experimental conditions for those participants who did not transform the format (in grey the significant comparisons)

	Ratio Format			
Ratio Format	1 2 in 50	2 0.4 in 10	3 1 in 25	4 10 in 250
1 2 in 50		χ^2 = 29.26, p = 1.00	χ^2 =23.31, p < .001	χ^2 = 29.26, p = .07
2 0.4 in 10			χ^2 = 20.83, p < .001	χ^2 =29.26, p = .12
3 1 in 25				χ^2(49)=29.26, p < .001

Participants' behaviour when dealing with the 1 in 25 format differed from that showed when dealing with any of the other "N in NX" formats. Indeed, while the majority of them declared not to be used to transform the 1 in 25 ratio format, the majority declared the opposite when dealing with 2 in 50, 0.4 in 10, and 10 in 250 ratio formats. Furthermore, the comparison between behaviour towards each of the three ratios in an "N in NX" format and any of the remaining two did not give significant results, thus proving that participants' behaviour didn't differ across them.

Hence, the hypothesis of the study was fully confirmed, and it could be affirmed that while the majority of people tended to leave the "1 in X" format untouched, the majority of them also tended to transform "N in NX" ratio formats.

Study 2.6

A special thanks to Prof. Vittorio Girotto and Stefania Pighin for hints on the implementation of research goals of this study

Goals

Study 2.4 showed, that the probability perceptions of two equivalent ratios, none appearing in formats featuring directly a "1 in X" and using denominators of different magnitudes (i.e., 48 < 160), did not differ significantly. Anyway, these findings did not allow to draw ultimate conclusions on the fundamental function of the number 1 at the numerator of a ratio for the appearance of the bias at study, namely on the necessity to have number 1 at the numerator for the higher probability perception to be elicited. Hence, the present experiment was carried out. The specific first hypothesis of the study was the following:

Hypothesis 1: People give higher judgments of the same probability when expressed in a "1 in X" format than in a "N in NX" format, and the same probability expressed as "N in N*X" or "N_1 in N_1*X" does not evoke different judgments.

Also, in the present study, the investigation on the tendency to transform the ratios that had been started in the previous study was continued here. Differently from Study 2.5, however, where the investigation occurred for ratio formats manipulated within-subjects, in the present study analyses regarded ratio values manipulated between-subjects; not only, as here the probabilities referred to a specific outcome, i.e., probability of having a Down syndrome-affected child. It could have been the case, that people's tendencies were different when the expressions were not abstracted from an outcome. Hence, in line with findings of Study 2.5, the following prediction was made,

Hypothesis 2: When it refers to a specific outcome (i.e., having a Down syndrome-affected child), people do not tend to operate a transformation on a ratio if it has a "1 in X" format, but they do transform it if it has an "N in NX" format.

If verified, Hypothesis 2 would have further confirmed the "1 in X" format idiosyncrasy.

Method

Participants. One-hundred sixty-nine individuals (97 females and 62 males) from the general population took part in the study voluntarily, all recruited through online[71] or word of mouth advertisement. The data of 10 surveys had to be eliminated due to sampling problems[72], thus the final sample was constituted by 159 individuals.

Mean age was 31.56 years (*SD* = 6.38), ranging from 21 to 60. Of the sample, the large majority (*n* = 123, 77.4%) had a (3/5 Years) University degree (*n* = 82, 51.6%) or a higher title like Master, PhD, or Specialization school diploma (*n* = 41, 25.8%), while a 22.7% had either a high school diploma (*n* = 33) or a lower title ("*licenza media*"). Many (*n* = 127, 80.4%) did not have children - one person did not answer to this question; only ten people (the 6.5% of the sample) declared to have been reported by a health specialist the information on their probability of having a child affected by Down syndrome.

Design and material. Participants read the same scenario to that employed in Study 2.2 but on a web-page, role-playing and imagining themselves as prospective parents. The single factor manipulated between-subjects was the numerical ratio expressing the age-related probability of having an affected child: the ratio could be expressed in four variants. Only one of the ratios had the "1 in X" format (i.e., "target"), while the others had a "N in NX" format. Also, in the present study, the comparison of a new "1 in X" ratio format was extended to other not previously tested "N in NX" formats (one even had a numerator smaller than 1). Thus, the values employed were 1 in 25 as "1 in X" format, while 0.4 in 10, 2 in 50, and 10 in 250 as "N in NX" formats.

Within the described context, participants were asked a subjective magnitude evaluation of the probability on a 7-point Likert-scale from "extremely low" to "extremely high". Then, on a separate web-page, they were asked to report whether in order to perform the magnitude evaluation task required, they had or not transformed[73] the ratio format. To conclude, the next web page addressed the socio-demographic variables.

Procedure. An electronic version of the questionnaire was accessed by participants who were randomly assigned by the program to one of the four conditions constructed to follow the 4 (ratio format) between-subjects design. A minimum of 30 participants per condition was necessary to make statistical comparison valid. Individuals were allowed an unrestricted amount of time to complete the questionnaire by themselves.

[71] i.e. newsletter or link to potential participants from the University of Trento main internet page.

[72] The computer server did not allow an individual to submit a completed questionnaire more than once per browser session. Still, individuals could submit multiple questionnaires if they closed their browser, reopened it, and returned to the Web site. That had been judged unlikely, especially considering that participants were informed about the absence of payments or rewards. Nevertheless, this phenomenon was showed by participants in 5 cases, thus forcing us to eliminate 10 questionnaires from data analysis.

[73] A short example of what was meant by "transformation" was given at the beginning of the web-page.

Results

Means for perceived probability in each ratio format condition are given in Table 1 and visually represented for an even more immediate comparison in Fig. 1.

Table 1
Mean perceived probability in each of the four experimental conditions (SD in parentheses)

Ratio Format	*Mean*	*SD*
0.4 in 10	3.63	1.61
1 in 25	4.42	1.71
2 in 50	4.17	1.51
10 in 250	3.72	1.19

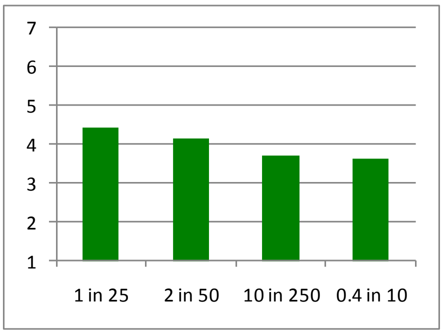


***Fig. 1.* Mean perceived probability in the four ratio format experimental conditions**

The visual inspection of Fig. 1 immediately suggests that a different subjective assessment of probability corresponded to the different ratio formats, even though not for all of them. Indeed, pair wise comparison confirmed this visual impression only in two cases, i.e., in the comparison of mean perceived probability for the 1 in 25 and the 0.4 in 10 formats, $t(79) = -2.15$, $p = .04$, and in that of the 1 in 25 and the 10 in 250 formats, $t(74) = -2.06$, $p = .04$ (see Table 2). These results showed that the "1 in X" format was perceived as significantly higher than both that of 0,4 in 10 and of 10 in 250, but not different from the 2 in 50 format. All the other pair wise comparison among formats other than "1 in X" did not return significant differences. Thus, it could be affirmed that the first hypothesis of this study was generally supported by data (if we exclude the 2 in 50 case mentioned above).

Table 2
Results of pair wise t-tests on perceived probability for each combination of the four experimental conditions (in grey the significant comparisons)

	Ratio Format			
Ratio Format	1. 2 in 50 (*n* = 42)	2. 0.4 in 10 (*n* = 41)	3. 1 in 25 (*n* = 40)	4. 10 in 250 (*n* = 36)
1. 2 in 50		$t(81) = 1.55$, $p = .12$	$t(80) = -.73$, $p = .47$	$t(76) = 1.43$, $p = .16$
2. 0.4 in 10			$t(79) = -2.15$, $p = .04$	$t(75) = -.27$, $p = .79$
3. 1 in 25				$t(74) = -2.06$, $p = .04$

For what concerns the second variable, that was, whether people had transformed the ratio format before evaluation, or instead considered it in the way it was expressed, contingencies are shown in Table 3.

Overall, across formats, the majority of participants (i.e., 57.9%) declared having not transformed the format they read in the scenario before evaluating its magnitude, a significant difference, $\chi^2(1, N = 159) = 3.93, p = .047$. Nevertheless, looking at the distributions, in Table 3, it appeared evident that this significant result was due to the significantly higher majority of people who declared not to have transformed rather than transformed the 1 in 25 ratio format (indeed, clearly the other comparisons had not generated any significance, as the distribution of answers in the two options was almost the same)- this was confirmed by a z-test for proportions between those participants who were given the "1 in X" format condition and those who were given one "N in NX" format condition, z = 3.09, 99.9% confidence interval (1-tailed).

Table 3
Frequencies of participants (% in parentheses) who transformed and did not transform the proposed format in the four experimental conditions

	Did you transform?		
Ratio Format	yes	no	*n*
1 in 25	8 (20.0)	32 (80.0)	40
2 in 50	20 (47.6)	22 (52.4)	42
10 in 250	20 (55.6)	16 (44.4)	36
0.4 in 10	19 (46.3)	22 (53.7)	41
Total	67 (42.1)	92 (57.9)	159 (100)

When separately comparing the "1 in X" format (i.e., 1 in 25) with any of the other three "N in NX" formats, all three z-tests for proportions returned significant values, namely z = 2.28, 98.9% confidence interval (with 0.4 in 10 as "N in NX format"); z = 2.40, 99.2% confidence interval (with 2 in 50 as "N in NX format"), z = 2.97, 99.9% confidence interval, (with 10 in 250 as "N in NX format"), all 1-tailed. Thus, while for the ratio expressed in the "1 in X" format, the majority of people did not tend to operate a transformation, for those ratios expressed in an "N in NX" format the majority of them did. Hence, data confirmed the second hypothesis of the study even when the ratios where referred to a specific situation.

Further analyses. Here following further analyses made on the basis of a distinction between those participants who did and those who did not transform the format received are reported. These analyses have only an exploratory value, as indeed the calculations had almost always been made on groups of evidently different numerosities, some of which were definitely too low to return reliable measures[74].

Table 4 reports the mean perceived probability in each sub-group of the four experimental conditions obtained taking into account whether participants declared having transformed or not the format received. The same descriptive means reported in Table 4 are also visually represented in Fig. 2.

Table 4
Mean perceived probability in the four experimental conditions separately assessed for those who declared to have transformed the format and for those who declared they did not (SD in parentheses)

	Ratio Format			
Transf.	1 in 25	2 in 50	10 in 250	0,4 in 10
Yes	3.62 (1.61)	4.20 (1.44)	3.55 (1.05)	4.16 (1.43)
	n =8	*n =20*	*n =20*	*n =19*
No	4.62(1.79)	4.14 (1.61)	3.94 (1.34)	3.18 (1.65)
	n =32	*n =22*	*n =16*	*n =22*

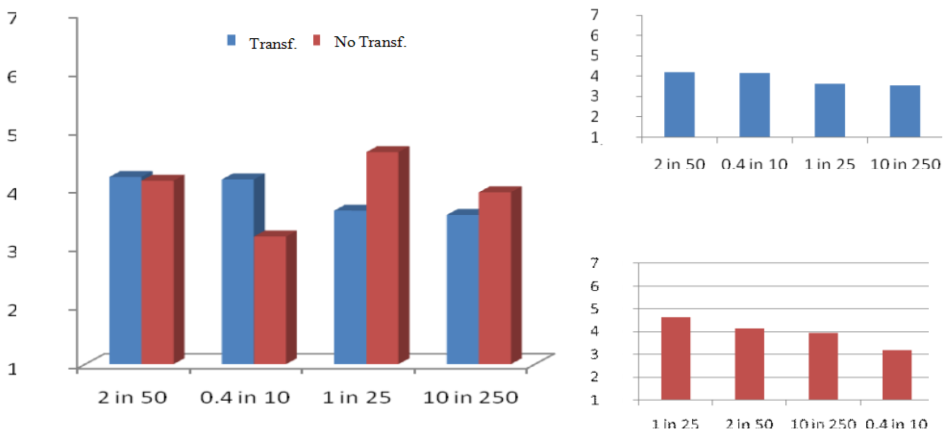


Fig. 2. **Mean perceived probability for each of the four experimental conditions distinguishing those who did/ did not transform the ratio format (whose assessments are separately represented in the two smaller figures)**

[74] Because of the low numerosity (i.e. $n < 30$) in each experimental condition of the two groups, non-parametric tests were used.

For participants who declared having not transformed the ratio format received, six pair wise independent Mann-Whitney tests were performed whose results are reported in Table 5. Only one significant difference was found.

Table 5
Results of the pairwise Mann- Whitney tests performed on perceived probability in the four experimental conditions for those who did not transform the format- in grey the significant comparison

	Ratio Format			
Ratio Format	1 2 in 50 (*n* = 22)	2 0.4 in 10 (*n* = 22)	3 1 in 25 (*n* = 32)	4 10 in 250 (*n* = 16)
1 2 in 50		Z = - 1.95, *p* = .051	Z = - 1.08, *p* = .28	Z = - .42, *p* = .693
2 0.4 in 10			Z= - 2,84, *p* = .01	Z =- 1.63, *p* = .10
3 1 in 25				Z = - 1,44, *p* = .15

For participants in the second group (i.e., those who declared having not transformed the ratio format received), six pair wise independent Mann-Whitney tests were performed whose results are reported in Table 6. No significant differences were found.

The effect of each of two of the socio-demographic variable (i.e., having children/not, and having been/not communicated by a doctor the probability of the future child be affected by Down syndrome) taken separately was not analyzable, because of the large differences in numerosity between the two groups (yes/no) to be compared. The attempt to create a new variable (i.e., "familiarity") as in Study 1.2 did not show success.

Table 6
Results of pairwise Mann- Whitney tests on perceived probability in the four experimental conditions for those who did transform the format

	Ratio Format			
Ratio Format	1 2 in 50 (*n* =20)	2 0.4 in 10 (*n* = 19)	3 1 in 25 (*n* = 8)	4 10 in 250 (*n* = 20)
1 2 in 50		Z= -.16, *p* = .87	Z= -.73, *p* = .46	Z= -1.42, *p* = .16
2 0.4 in 10			Z=-.60, *p* = .55	Z=- -1.24, *p* = .21
3 1 in 25				Z= -.29, *p* = .77

Discussion

The present study further corroborated the peculiarity of the “1 in X” format in probability perception thanks to the simultaneous demonstration of two phenomena. Indeed, results showed that the probability perceptions of two equivalent ratios none of which featuring directly the “1 in X” format did not differ significantly. At the same time, results proved that people’s probability judgments were systematically higher for the “1 in X” format in two out of the three pair wise comparison among ratios where one appeared in the “1 in X” format. Thus, it was demonstrated, that the presence versus absence of the “1 in X” format in one term of the pair wise comparison could determine the “1 in X effect”.

Findings on the variable transformation further confirmed the idiosyncrasy of the 1 in X format found in Study 2.5, as that was the only format tested on which the slight majority of participants declared not to have applied the mathematical operation.

In sum, through this study, the “1 in X effect” received additional corroboration. Its strength was additionally raised by findings indicating that the higher probability judgments could be obtained despite the 55.6% (in the case of 10 in 250) and the 46.3% (in the case of the 0.4 in 10) of the sub-samples declared to have transformed the format received.

Further investigation of the peculiarities of the 1 in X format could help to discover the mechanism at the basis of the “1 in X” effect. Future studies could, for example, identify the ratio formats in which the proposed ratios are normally transformed, in order to detect 1) which feature, if any, is responsible, in those ratios, for the higher/lower probability judgment; 2) which are the clues that could help clarify the mental processes taking place in people’s mind during the probability evaluations process.

The effect of individual differences on the bias under study

A growing body of research has documented a variety of individual differences that influence decision making accuracy. Even if the specific analyses are not reported, in each of the studies described until here both in Section 1 and Section 2, demographical and individual differences among participants had been checked for in the extent in which they were affected by the "1 in X" bias. Overall the list below summarized results:

- **Gender** (all studies): no effect
- **Age** (all studies): no effect
- **Education level** (all studies): no effect
- **Familiarity with the scenario** (Study 1.2., our question; Study 2.1, *ex-post* construct): the "1 in X effect" was present only on those participants highly "familiar" with the outcome at study (i.e., Down syndrome)- in Study 1.2, results obtained only through exploratory analyses, as the necessary conditions to test the hypothesis were not present
- **Information-processing style** (Study 2.1**,** REI scale, i.e., Pacini and Epstein, 1999-b): "1 in X effect" only on low experientials
- **Numeracy:**
    - **objective measure,** that was obtained by means of Lipkus, Samsa, and Rimer (2001)' s inventory; (Study 1.1): "1 in X effect" only on high-numerates
    - **subjective measure**, that was obtained by means of Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, and Smith (2007)'s inventory; (Study 2.1): "1 in X effect" only on high-numerates
- **Having children or not** (Study 2.1): having already one or more children was a necessary condition for the ratio format effect to appear[75]

[75] Two One-Way ANOVAs on perceived probability with ratio format as between-subjects variable were separately conducted on the those who had declared having children and on those who had instead declared they did not have any. Results showed that having already one or more children appeared to be necessary condition for the format effect to appear, $F(2, 113) = 8.41$, $p = .000$, $\eta^2 = .133$ for those who had children, while $F(2, 131) = .975$, $p = .380$ (*n.s.*) for those who did not have children.
In those participants who had children, Post hoc comparisons using Fisher's LSD tests revealed. that the 1 in 200 format was perceived as significantly higher ($M = 4.34$, $SD = 1.70$) than both the 5 in 1000 format ($M = 3.26$, $SD = 1.25$, $p = .002$) and the 50 in 10,000 format ($M = 3.00$, $SD = 1.43$, $p = .000$), but the latter two formats did not differ significantly in terms of perceived probability judgment evoked.

SECTION 3, or
An affective explanation?

The heart has its reasons, of which reason knows nothing.
Blaise Pascal

An account that has not been examined until now concerns the different degree of affective reaction to the numerical expressions of probability compared each time. The relevance of such emotional response to stimuli for probability assessment, and in turn to the bias at study can be summarized in the following two considerations.

First of all, Peters, Lipkus, and Diefenbach (2006) had showed in an experiment that the superficial features of stimuli are able to shape, among other, people' s emotional reactions, and that the latter in turn have the power of influencing people's assessments and preferences by means of the so-called "affect heuristic"- see Chapter 2 (e.g., Finucane, Alhakami, Slovic, and Johnson, 2000; Slovic, Finucane, Peters, and MacGregor, 2004; Slovic and Peters, 2006).

Secondly, also two of the theories (i.e., CEST theory, and the risk as analysis and risk as feelings approach) that have specifically addressed the phenomenon under study, namely people's evaluative behaviour when ratios of different superficial appearance are employed, have explained findings in terms of a prevalence of the experiential system over the rational one. Furthermore, some findings (awaiting further confirmation) obtained in the previous section (i.e., Study 2.1), namely that rationality was not related with the bias under study, while individuals showing a low degree of experientiality were those more prone to the bias, seemed to indicate a relevant role of the experiential system in the manifestation of the tendency.

In light of both theoretical considerations made above, and of the preliminary findings just reviewed, it was reasoned, that even in the present case the emotional responses to the superficial differences of the ratios might have explained (partly or completely) the differences found in participants' perceived probability. Hence, aim of studies described in the present section was, to inspect if the affective information conveyed by equivalent ratio formats with a different superficial structure (namely, "1 in X" and "N in NX" formats) was in fact different. It has to be reminded from the characterisation made in Chapter 2, that the experiential system is more affective in nature than the analytic one, and it is influenced by emotions, images, and intuitions. Possibly, these factors could have explained the different degree of perceived probability assessed for the two formats.

Since both direct and indirect means of measuring affect exist, both were tested. A direct mean will be employed in Study 3.1, while an indirect one in Study 3.2.

Study 3.1

Goals

In the field of risk research, affect is normally measured via self-report, that is a direct method of assessment consisting in indicating on a scale whether the hazard is judged as good or bad, positive or negative (see, for example, Alakhami and Slovic, 1994). In a similar way, in the present study the evaluation of the affective answer to stimuli was measured through two direct questions, namely by asking participants to separately estimate the degree of positivity and that of negativity of their feeling for the uncertain outcome. The specific hypothesis of the experiment was,

> **Hypothesis: When expressed in a "1 in X" format, a probability conveys a higher degree of negative feelings [and a lower degree of positive feelings] than when communicated in one of the equivalent "N in NX" formats.**

Method

Participants. A total of a hundred and twenty-one individuals (73 women and 48 men) from the general population took part in the experiment voluntarily. Mean age was approximately 28.3 yrs. (*SD* = 7.2), varying between a minimum of 18 to a maximum of 47- one person did not disclose this information.

Design and material. Participants read the usual scenario of a couple of prospective parents who received by the gynaecologist the communication of their probability of having a child affected by Down syndrome. They were exhorted to role-play one of the prospective parents described in the scenario (according to their gender).

The study manipulated a single factor between–subjects, namely the ratio format employed to communicate the probability: the numbers at the numerator and at the denominator of such a ratio could be either both comparatively "smaller" or both comparatively "larger" than in the other condition, with one ratio in a "1 in X" format and the other in a "N in NX" format. 1 in 12 and 10 in 120 were the ratio chosen for this study, alike in Study 2.3.

After having read the scenario, participants answered to two questions, whose order was controlled for[76], separately assessing[77] the degree of positivity and that of negativity of their

[76] To easily identify the conditions according to the order in which the two questions appeared in the questionnaire, the case where the negative question appeared before the positive one has been redubbed as "order 1" for the description of the study, while the other as "order 2".

[77] One question measured the degree of negativity of the feeling, while the other the degree of positivity of the feeling.

feeling for the probability of “1 in X” [“N in NX”] of having a Down syndrome- affected child. The estimation had to be provided on an 11-point Likert scale anchored to “not at all negative” and “extremely negative” in the case of the negative assessment, while to “not at all positive” and “extremely positive”, in that of the positive assessment, with no verbally labelled points in between in both cases.

Procedure. Each participant was randomly assigned to one of the four experimental conditions created to correspond to the experimental design. Individuals could complete the questionnaire at their pace, without any time constraints.

Results

One participant was excluded from the analyses because of not providing an answer to one of the two scales. Therefore, analyses were performed on a total of 120 participants.

No effects were found that could be attributed to the order of the two questions, t_{neg} (118) = - .38, p = .71 and , t_{pos} (118) = - .38, p = .62 (see Table 1).

Table 1
Mean declared degree of negative and positive affect in each order condition

	Order	*n*	Mean	*SD*
Neg	1	62	4.26	3.10
	2	58	4.47	2.95
Pos	1	62	6.90	3.29
	2	58	6.60	3.21

A one-way within-subjects (or repeated measures) ANOVA on the effect of ratio format on the two measures of affect (negative and positive) returned no significant main effect for the ratio format, *Wilks' λ* = .99, $F(2, 115) = .12$, $p = .69$, $\eta_p^2 = .01$. No significantly different negative affective evaluations were provided in the two ratio format conditions, $F(1, 116) = .08$, $p = .78$, $\eta_p^2 = .00$; the same had to be affirmed for the positive affective evaluations, $F(1, 116) = .63$, $p = .43$, $\eta_p^2 = .01$- for exact means see the following Table (i.e., Table 2), while for a graphical representation, see Fig. 1.

Table 2
Mean perceived negative and positive affect for the probability in each experimental condition

	Ratio format		
	1 in 12	10 in 120	Total
Affect	(n = 61)	(n = 59)	(n = 120)
Negative	4.26 (2.87)	4.46 (3.18)	4.36 (3.02)
Positive	7.00 (3.05)	6.51 (3.44)	6.76 (3.24)

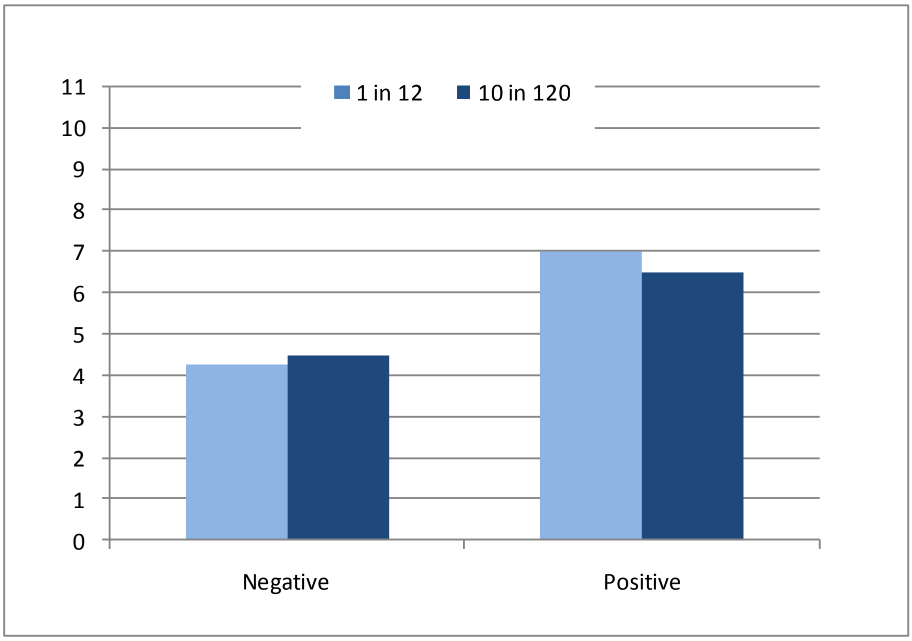


Fig. 1. **Mean negative and positive affect in the two experimental conditions**

Discussion

No difference was found in terms of positive and negative affect elicited by the ratio between the two formats ("1 in X" and "N in NX") expressing the same probability of having a Down syndrome-affected child. The two null results obtained through a direct testing methodology leaded to think that the affective explanation was not the adequate one for the systematic effect at study[78].

[78] The same result was obtained even when using the type of measure normally used in decision making studies investigating the affect variable, i.e. a single bipolar scale anchored to "extremely negative" and "extremely positive", with the half-way point labeled "neutral". Indeed, no significantly different ratings on the scale were obtained in the two ratio conditions for the 96 individuals of the general population who participated in the study, $M_{1\ in\ 10}$= 3.42 (SD = 1.90) and $M_{10\ in\ 100}$= 3.96 (SD = 1.70), $t(94) = -1.47$, $p = .14$.

Study 3.2

Goals

A direct methodology for assessing affect can sometimes not be the most adequate for some decisional situations. It could have been the case of Study 3.1, as social desirability concerns might have played a role (Down syndrome can be a displeasing or embarrassing domain for some people), making thus participants alter their responses. Or, it might have been the case, that in such study individuals had found it hard to recognize their own mental processes since, as affirmed by Dohle, Keller, and Siegrist (2010, p. 1117), "these can lie beyond one's introspective capability".

In similar cases, implicit measures of attitudes (i.e., an indirect method) represent a valid alternative methodology for assessing affective associations. In a study on the potential adverse economic impact of a proposed nuclear waste repository site, Slovic, Layman, Kraus, Flynn, Chalmers, and Gesell (1991), for instance, used an indirect strategy instead than direct questions to assess people's perceived impact of nuclear facilities on future behaviour. The aim was that of avoiding introspective judgments that are often criticized as not trustworthy because of people's difficulty in foreseeing long distance consequences of events. The study employed a version of the "method of continued associations" (Szalay and Deese, 1978) which authors described as having been very extensively used by Galton, Wundt, and Freud. Word-association techniques are easy and efficient ways of determining the contents and representational systems of human minds without requiring those contents to be expressed in the full discursive structure of human language. As a consequence, such techniques are thought to be better ways of revealing ideas or associations otherwise probably difficult to spell out. In the version modified for the experiment, participants of Slovic and others' study had to report the first 4-6 thoughts or images (normally a word or a very brief phrase) they associated with the target stimulus. After having stated their mental associations, participants were requested an evaluation of each image in terms of affective meaning on a short (4/5 point) Likert scale, ranging from "very negative"/ "very bad" to "very positive"/ "very good".[79]

[79] In order to obtain an overall imagery index, "the summation model" had been proposed by the authors, namely a scoring method consisting in summing and averaging ratings for all the images a respondent had produced. Results using this methodology are somewhat mixed. For instance, on the one side, Slovic et al. (1991)'s results showed that image ratings and consequent preference order were predictive of people's behavior (i.e. ultimate vacation choices, preferences for cities, and states in which to take a new job, retire, or site a business). Moreover, when a further survey using exactly the same methodology was performed 16-18 months later on smaller samples of the same participants, data confirmed that the affective qualities of a person's word-associations were clearly related to the probability that the person would subsequently undertake a given behavior. Similar results using the same methodology were obtained in a study on adolescents by Benthin, Slovic, Moran, Severson, Mertz, and Gerrard (1995): the affective quality of imagery acted as a strong predictor for participation in some health-related behavior.

In the present experiment, it was decided to use the same version of the "method of continued associations" used by Slovic and colleagues. In accordance with the affect heuristic it was expected, that the two superficial ways of expressing the same probability, namely a "1 in X" or in a "N in NX" format for the numerical ratios, might have elicited mental images whose degree of affective tag could vary, thus indirectly influencing the (different) degree of perceived probability in the two cases. In particular, the specific hypothesis for the study was that,

Hypothesis: The negativity of the affect conveyed by an image evoked by a given probability is higher when such probability is expressed by a ratio in a "1 in X" format than in an equivalent "N in NX" format.

Method

Participants. Sixty-seven individuals from the general population (39 females and 27 males) took part in the study voluntarily, all recruited face-to-face in public locations (e.g., libraries, town parks). The data of one person had to be eliminated due to some problems in comprehension of the Italian language. Thus, the final sample comprised 66 individuals.

Mean age was 30.25 years (*SD* = 11.3), ranging from 18 to 65- one person did not report this information. Self-reported education level was generally quite high, as the 50% of the sample (*n* = 33) had a (3/5 Years) University degree, 3 people (4.5%) a higher title like Master, PhD, or Specialization school diploma, and the 40.9 % of the sample (*n* = 27) had a high school diploma, while only 3 people (4.5%) possessed a lower title ("*licenza media*").

Only the 24.2% of the sample (*n* = 16) had one or more children. Among these, only 3 people declared to have been reported by a health specialist the information on their probability of having a child affected by Down syndrome.

Design and material. The scenario employed was similar to that used in previous studies. The maternal age-related probability that the future child would be affected by Down syndrome was expressed through a numerical ratio that could be expressed in two variants, namely one using a "1 in X" format (i.e., 1 in 200) and one using a "N in NX" format (i.e., 5 in a 1000). As in previous studies, the single factor manipulated between-subjects was the relative magnitude of both numerator and denominator in each ratio. Nevertheless, this time instead than a subjective assessment of the probability, participants were asked to report the first image coming to their

---

On the other hand, results of a study on financial judgments (MacGregor, Slovic, Dreman, and Berry, 2000) showed that despite the fact imagery and affective evaluations were highly correlated with one another, and with the likelihood of investing, the latter allowed for low degree of predicted actual market performance. This could well be a product of the difficulty in forecasting future behavior, which is particularly strong for the financial context in which conditions on whose basis decisions have to be taken are continuously changing. Anyway, as difficulty in forecasting future behavior is common to other fields, it cannot be explained why the variable would not play a role in those other studies.

mind when thinking about the specified numerical probability of having a child with Down syndrome[80]. Following the elicitation of the image, respondents were asked to rate it on a scale ranging from very negative (- 2), somewhat negative (- 1), neutral (0), somewhat positive (+ 1), or very positive (+ 2).

Procedure. Each participant received a paper copy of one out of the two versions of the survey constructed to follow the experimental design. A minimum of 30 participants per condition was necessary to make statistical comparison valid. Individuals were allowed an unrestricted amount of time to complete the questionnaire by themselves. Anonymity and confidentiality were ensured, as well as a short debrief given to them at the end of data collection.

Results

A total of 66 mental images was generated (1 each participant), whose complete list is available on request from the experimenter. Table 1 shows the distribution of such images in content categories created on the basis of a classification scheme developed ad hoc for the study. As the nature of the analysis on images was only explorative, and our interest was mainly on the affect conveyed by the images, the standard procedure for analyzing qualitative data (i.e., multiple coders) was not adopted.

The most frequently evoked idea in the sample (n = 15, 22.7 %) was a material image of the child him/herself or of a specific physical detail. Nurturance, and need of special care was the next thought in terms of frequency (n = 6, 9.1 %), and the idea of a person affected by the disease and known by the participant, together with the arousal of difficulties/problems, or the association with a specific colour were the images immediately following (n = 5 each; 7.6 %). Similarly to Finucane, Slovic, and Mertz (2000), the general affect rating across the images, as well as their mean affective valence, was calculated for all the responders. The rating was slightly negative (M = -.14, SD = 1.29). The majority of images was negative or very negative (n = 33, 50.0%), but a substantial proportion of them was positive or very positive (n = 20, 33.4 %) or neutral (n = 11, 16.7 %), a result compatible with the positive attitude shown by some people. No difference between the two experimental conditions was detected in the number of mental images of either valence generated by participants- see Table 2. Figure 1, instead, shows the distribution of answers for each affective valence available in the 5-point Likert scale in each experimental condition.

[80] An informal definition of what was meant by "mental image" was provided in a footnote, together with an example.

Table 1
Images associated with the 1 in 200 [5 in 1000] probability of having a Down syndrome-affected child

Category	**Frequency**	**%**
Visage, body, physical detail, photo	15	22.7
Nurturance, care	6	9.1
Known person	5	7.6
Difficulties, problems	5	7.6
Colour	5	7.6
Positive feeling or attitude	4	6.1
Hospital, choice, doubt	4	6.1
Anxiety, worry	3	4.5
Emargination	3	4.5
Handicap	3	4.5
Sadness	2	3.0
Probable, possible, and contraries	2	3.0
Total	57	86.4
Ambiguous category/ not classified	9	13.6
Total	66	100.0

Table 2
Number of images for each valence in the two experimental conditions

	Ratio Format		
Valence	1 in 200	5 in 1000	Total
Negative	18	15	33
Positive	11	11	22
Neutrum	4	7	11
Total	33	33	66

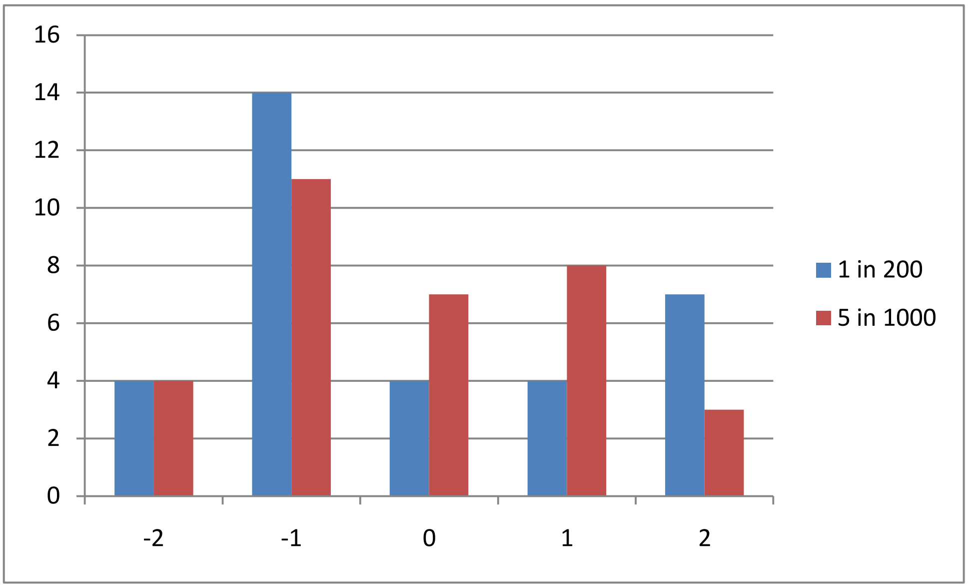


Fig. 1. **Distribution of the affective valence attached to the 66 images generated as function of the experimental condition**

A visual inspection of the figure suggests a right shift of the distribution of answers for the 5 in 1000 condition (more positive average valence) compared with the distribution of answers for the 1 in 200 condition (more negative average valence). In order to check this impression, a test on the equality of the means between the two groups was performed[81]. The test returned an equivalent affective reaction toward the 1 in 200 and the 5 in 1000 conditions, $t(64) = .10$, $p = .93$, $M_{1\ in\ 200} = .12$ ($SD = 1.39$), $M_{5\ in\ 1000} = .15$ ($SD = 1.20$), therefore the visual impression generated by observing the figure was not confirmed. Thus, the main hypothesis was not supported by data. Though, the median answer was lower in the 1 in 200 condition (i.e., - 1) than in the 5 in 1000 condition (i.e., 0)- see Fig. 2. This tendency was in the expected direction, namely a more negative affective valence for the smaller-numbered ratio condition than for the larger-numbered one.

The content of the mental associations was further examined by investigators for the attributed valence (positive/negative/neutral). These associations were distinct from the affect ratings assigned by respondents to their image. The proportional majority of the images was judged by the experimenters as negative ($n = 27$, 40.9%), with only 13 (19.7%) judged as positive; those images which could not be classified, i.e., 6 (9.1%), expressed an ambivalent thought, while 20 images (30.3% of the sample) were judged neutral, with no clear connotation. No main significant differences in the two experimental conditions in the number of images of either valence were detected.

[81] See Rubaltelli, Pasini, Rumiati, Olsen, and Slovic (2010). Obviously, the classic summation model (Slovic, Layman, Kraus, Flynn, Chalmers, and Gesell, 1991) generally used in similar researches was not appropriate here, as more than one image per person would have been necessary to apply it.

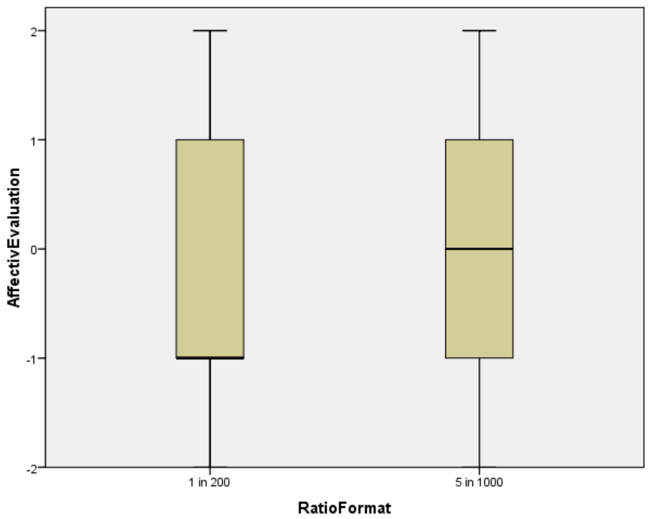


Fig. 2. **Box plots of the affective evaluation in each experimental condition**

Discussion

The present study investigated the hypothesis of a different affective response to the two ratio formats under examination (i.e., “1 in X” and “N in NX”) by means of an implicit measure. Such diverse emotional reaction was supposed to originate in the mental images evoked by either format, and if found, could be attributed the ability to mediate probability perception. However, unfortunately results showed that the degree of emotional valence attached to the mental images elicited by either format did not differ. A possible problem with the methodology might have been the fact, that differently from the original procedure (e.g., Slovic, Kraus, Flynn, Chalmers, and Gesell, 1991) only one mental image, instead than 3 or more, had been elicited. Another problem could reside in the way the procedure was translated. Participants were not asked to read the ratio (“1 in X” vs. “N in NX”) alone, but the scenario was always given (i.e., probability of having a child affected by Down syndrome). In previous studies on images, the stimulus was often just a word (e.g., nuclear power) and not a scenario or a phrase.

SECTION 4, or
Test of two practical interventions

Studies in Section 2 replicated and generalized the effect observed in Section 1: When the probability of a medical risk was expressed as “1 in X”, that probability loomed bigger (and the risk more alarming) than when it was expressed by the equivalent ratio in “N in NX” format. As a consequence of such findings, doctors and health personnel must thus be warned that their choice of a ratio format in communicating the probability of a health-related outcome (here, the risk of having a Down syndrome-affected child) does not come without consequences on patients’ assessments. Indeed, opting for a “1 in X” format might influence laypeople subjective impressions by increasing their probability perceptions in respect to other formats. Nevertheless, a question at that point was, Are professionals aware of the bias identified in laypeople? To answer such query, in Study 4.1, a group of medical doctors will be studied who will show not to be prone to the bias, probably thanks to their higher expertise with the concept of risk and probabilities. That will call for an examination of possible communicative interventions able to attenuate or even eliminate the “1 in X” effect.

In order to avoid as much as possible sub-optimal decision that might originate, for instance, in inflated assessments two such interventions will be tested. It seemed natural to turn to classic interventions aimed at coping with the poor understanding people have of probabilities and numbers (Garcia-Retamero and Galesic, 2009; Gigerenzer, 2002; Gigerenzer and Edwards, 2003; Lipkus, Samsa, and Rimer, 2001; Peters, Lipkus and Diefenback, 2006; Peters, Västfjäll, Slovic, Mertz, Mazzocco, and Dickert, 2006). As in the domain of risk communication, it is common practice to use verbal analogies (see Barilli, Savadori, Pighin, Bonalumi, Ferrari, Ferrari, et al., in press), or visual aids (i.e., graphical representations), they will both be tested. Study 1 and 2 will accordingly assess the resistance of the “1 in X effect” to these two interventions**.**

Study 4.1

A special thanks to Natasha Cont for collection of data presented in this study

Goals

Results of Sections 1 and 2 defined a new bias in the probability perception of a health-related outcome (i.e., the risk of having a Down syndrome-affected child) when the probability is expressed by means of a numerical ratio. When the same probability was expressed in a “1 in X” format, it loomed bigger (and the risk more alarming) than when it was expressed in a “N in NX” format. Thus, health-care professionals who routinely communicate probabilistic information on health-related outcomes must be warned that a probability phrased as “1 in 200”, for instance sounds bigger to patients than the same probability expressed as “5 in 1000” or “50 in 10,000”. However, it was reasoned that, it could have been not exceptional to discover that practitioners were not aware of the fact patients might interpret differently the two probability expressions, because either they were as well influenced by the same tendency, or because they were themselves de-sensitized to the bias as effect of their expertise (i.e., for them the two ratios represent the same probability, thus they do not imagine other people could not think the same). In other words, there were reasons leading to think, that practitioners’ judgment could have differed from that of their patients.

Indeed, literature has in some situations showed a systematic departure of expert judgment from that observed in laypeople (i.e., novices to the problem). Reasons for it have been recognized in the ability to organize “[…] cues into larger ‘chunks’”, and in that of recognizing “[…] patterns more easily, more accurately and more quickly” (Koehler, Brenner, and Griffin, 2002, p. 692) provided they are in the domain of expertise. Some authors (e.g., Adams and Smith, 2003, p. 78) have been referring even to two different “tribes”, namely experts (“those people who have knowledge of, and usually some responsibility for dealing with, risks in a particular area. They may not be responsible for creating the risks, but are charged both with representing those risks to others and with influencing their nature and severity”), and “consumers”, (i.e., those who lack a scientific or professional knowledge in that field; often, though not always, lay people, as “[…] it is quite possible that experts in one type of risk will also be consumers of many others”).

Following Bonini, Del Missier, and Rumiati (2008), one can define experts those people who can make an evaluation or perform a choice in a quick way that most of times represents the result of an intuition, of a rapid process based on tacit and implicit knowledge rather than on a standardized conscious procedure. This definition allows a more exhaustive recognition of experts than each of the separate criteria identified instead by Shanteau, Weiss, Thomas, and

Pounds (2003), namely 1) having a relevant number of professional tenure years; or, 2) having got formal certification qualifying them as such; or, 3) having been co-opted by people who are recognized as experts; or, 4) formulating accurate and stable definitions in their field of competence. Bonini and colleagues claimed that experts' judgment should not only be constantly (stably) accurate, but also valid, namely it should not change as a function of (really so or perceived) irrelevant information, like (to quote the case of the present studies) the magnitude of the numbers in a ratio conveying a probability.

In most of cases, the decision of so-called experts are taken following intuitions (see, for instance, the finding that expert clinicians can formulate a diagnosis in less than 15 seconds-Hamm, 1988) rather than examining situations analytically in depth. Use of heuristics by experts would correspond to the functionality of those shortcuts in complex, and uncertain situations requiring quick decisions that cannot follow from systematic and complete rational analysis (see, for instance, nurses' need of speedily identifying a heart attack). This view is summarized in Gigerenzer and colleagues' idea of "fast and frugal heuristics" (Gigerenzer, 2008; Gigerenzer, Todd, and the ABC Group, 1999), according to which cognitive shortcuts would have an ecological value as in most of the situations they lead to the best solution. In the genetic field, for example, Dewhurst, McCarthy Veach, Lampman, Petraitis, Kao, and LeRoy (2007), showed in two studies on the methods used to solve (four different) genetic problems, that while even the large majority of genetic undergraduate students made use of heuristics, the use of mental shortcuts by genetic counsellors resulted in a greater percentage of correct answers than in less-experts (see similar results in Smith and Good, 1984, reporting videotape analysis of novices and experts solving analogous problems). One of the studies argued, that explanations resided in experts' better understanding of the rules of probability, perhaps due to their clinical experience involving computation and interpretation of risk, beside the statistical education received (even though the number of years of experience was not a predictor of accuracy in that case).

Obviously, it is also possible that such cognitive shortcuts, based on intuition, lead to biased answers, namely errors. A large body of studies has showed that even experts are prone to some of the systematic and predictable cognitive biases that have been detected in laypeople, for instance availability and representativeness heuristics, anchoring and adjustment, omission fallacy, framing effect (for a review of "errors" in medical experts, see for instance Motterlini and Crupi, 2006). Similarly to patients, experts find numbers and probabilities difficult to understand (e.g., Gigerenzer, 2002), as demonstrated on tests in basic numeracy (Estrada, Martin-Hryniewicz, Peek, Collins, and Byrd, 2004; Schwartz, Woloshin, and Welch, 1999).

An analysis of experts (e.g., medical doctors)' perceptions of the same probability conveyed in ratios using numbers of different magnitudes not only would have allowed to gather further important information on this fundamental subject of the communicative pair, but also to obtain hints of undoubted useful value for the aims of the present research. Hence, the specific hypothesis for this study was the following,

> **Hypothesis: Health care professionals, differently from laypeople, are not susceptible to the "1 in X" bias in perceived probability.**

Method

Participants. Fifty-eight medical doctors recruited in hospital wards of a clinic in North Italy accepted to take part in the study for free. Seventy-three percent of them (n = 41) were males while 27% (n = 15) females- two participants did not disclose their gender. Mean age was 45.9 years (SD = 8.2) varying from a minimum of 31 to a maximum of 60 (two participants did not disclose their age). Practitioners had different specialties (see Table 1), and had a mean of 18.36 years of tenure (SD = 8.35) ranging from a minimum of 2 to a maximum of 35 years.

Design and material. The study manipulated a single factor in a between-subjects design, namely the format of the ratio employed to describe the probability of the outcome happening (i.e., in smaller numbers vs. larger numbers at the numerator and denominator). Participants read a scenario very similar to that used in Study 1: they were asked to role-play, that when considering the possibility of buying a trip to Kenya they had received information on the statistic of a severe disease. In particular, information regarded the statistics on the possibility of contracting a contagious disease in the country[82]. The probability information was expressed by means of a numerical ratio in two variants between-subjects, namely using comparatively smaller numbers in a "1 in X" format (i.e., 1 in 200), or an "N in NX" format using larger numbers at the numerator and denominator (i.e., 5 in 1000).

Participants were asked to rate their subjective perception of the magnitude of the probability presented in the scenario on a 7-point Likert-scale anchored to "extremely low" and "extremely high", with all points verbally labelled.

[82] This time the disease at stake was hepatitis A instead than malaria. Also, the attribution of the risk to a friend in the scenario was not necessary anymore as no ethical guideline imposed it for the type of participants involved in the study.

Table 1
Specialties of Medical Doctors taking part in the study

Specialty	Frequency	%	% correct
General Medicine	2	3.4	3.9
Orthopaedics	10	17.2	19.6
Cardiology	6	10.3	11.8
Neurology	2	3.4	3.9
Otolaryngology	3	5.2	5.9
Ophthalmology	1	1.7	2.0
Endocrinology	2	3,4	3.9
Geriatrics	2	3,4	3.9
Surgery	9	15.5	17.6
Gynaecology	4	7.9	6.8
Anaesthesia	4	7.9	6.8
Gastroenterology and Dig. Endoscopy	2	3.4	3.9
Rheumatology	2	3.4	3.9
Nephrology/Urology	1	1.7	2.0
Dermatology	1	0.9	1.0
Partial total	51	87.9	100.0
Not specified	7	12.1	
Total	58	100.0	

Procedure. Participants were randomly assigned to one of the two conditions constructed to follow the experimental design. They were allowed an unrestricted amount of time to complete the questionnaire.

The study did not require an ethical approval because it was not addressed to a special population (i.e., children, patients, and humans non capable of informed consent). Anonymity and confidentiality were ensured to participants, though.

Results and Discussion

Contrary to findings on general population, doctors' mean probability estimations did not differ significantly in the two considered experimental conditions, namely when the ratio format was either 1 in 200 or 5 in 1000 (M = 4.53, SD = 1.33), Z= -.01, p = .99. Thus, the hypothesis of equality of the means (same perceived probability irrespective of the format of the ratios) received confirmation in the data.

The absence of the effect observed in the other studies in general population could be linked to the fact that health professionals are used to perform operations on the mathematical formats they normally read on scientific sources, receive as results of technical tests, and have to communicate in clinical consultations. They might be used to execute similar operations across formats, and consequentially be more skilled than the average person in those operations. Thus, the superficial format conveying a statistical information would not matter very much to them, since as long as they had become experts, they might have adopted the habit of standardizing every numerical format to the one they “sense” best (e.g., percentages) before evaluating its magnitude. Nevertheless, this is only a conjecture, though very likely, as such ability was not tested in concomitance with practitioners’ evaluations.

Study 4.2

Goals

In risk communication to patients, Medical Doctors frequently use metaphors, similes, or analogies attempting to enhance the otherwise poor comprehension of small probabilities by laypeople (Edwards, 2003). It is generally thought, that analogies[83], through real (even thought imagined) quantitative sets of events, are conceived as ways to better convey the size of a risk or the probability of its occurrence, as well as to facilitate people's role in grasping the meaning of otherwise "cold" numerical values. For instance, the concept of chance is introduced by experts as rolling a dice or extracting a lottery ticket; other frequently used analogies are metaphorical histories or the "balls and jar" analogy (Edwards, 2003). The latter is a verbal technique where, in order to facilitate comprehension of the size of a probability, people are encouraged to imagine its magnitude as similar to the chance that an individual has of drawing a ball of some given desired features from an urn full of balls. For instance, a 1 in 100 probability of contracting a disease can be compared with the chance of extracting the only red ball from a hypothetical jar containing 99 white balls and 1 red ball.

Despite the undoubted great appealing of analogies as aiding communicative instruments, we were only aware of one study which tested empirically their influence on the subjective assessment of probability (Barilli, Savadori, Pighin, Bonalumi, Ferrari, Ferrari, et al., in press). In that study, a significant effect was found on the level of perceived probability when a balls and jar analogy was employed in risk communication. Thus, as it was suggested that verbal analogies 1) influence the degree of perceived probability of a phenomenon, and 2) help people better figuring out the magnitude of those probabilities, it was thought, that they could even influence the "1 in X" effect. Indeed, if thanks to the analogy, people would have better understood the numerical ratios proposed, it would also have been more probable that they could attribute two not significantly different evaluations to the same probability when in a "1 in X" or in an "N in NX" format. The analogy would, as such, work as a de-biasing technique. Hence, the hypothesis for the current study was:

Hypothesis: Using a verbal analogy like the balls and jar to communicate the probability eliminates the "1 in X" effect.

[83] Analogies are comparisons based on alignment and mapping of the underlying structure of two represented situations (Gentner, Holyoak, and Kokinov, 2001, chapt. 6).

Method

Participants. A total of 81 patients (50 women, 23 men, and 8 people who did not disclose their gender) of a hospital of North Italy, took part in the experiment voluntarily. The large majority of them (n = 62) were recruited at an out-patients' (gynaecologist), some of them (*n* = 15) at the Sterility Centre of the hospital, while four people in a maternity ward.

Mean age was 33.57 yrs. (*SD* = 4.83), varying between a minimum of 24 to a maximum of 47 (one participant did not disclose this information). Of the sample, the 51.4% (*n* = 36) had completed high school, the 32.9% (*n* = 23) had a degree, while 15.7% (*n* = 11) had the lowest education level in Italy (11 people did not disclose their education level). Of the sample, 40.5% (*n* = 30) had already one or more children at the moment of study completion, while the remaining 59.5% (*n* = 44) did not have children (7 people did not answer to this question). Indeed, the large majority of participants or their couple- if man- (namely the 84.3%, *n* = 59) was pregnant at the moment of survey completion (eleven people did not disclose this information).

Design and material. Participants read a scenario (see Study 1.1) encouraging them to imagine that one of their friends had just bought a trip to an exotic country. They were informed of the statistics on the risk of contracting malaria in that country, a probability expressed through a verbalized numerical ratio that could be expressed in one out of two formats, namely a "1 in X" format (i.e., 1 in 200) or a "N in NX" format (i.e., 5 in 1000). After that, they were asked to imagine a jar containing a number of balls equal to the denominator of either ratio (i.e., 200 or 1000); these balls were in two colours. A number of balls equivalent to the numerator of either ratio (i.e., 1 or 5) was described as in red colour, while the remaining number of balls (i.e., 199 or 995) was described as in white colour. People were encouraged to equate the probability they had read in the scenario to the probability of drawing a red ball from the jar.

The only dependent variable assessed in this study was the subjective magnitude of the probability of infection. It was measured by asking a personal assessment of the friends' probability of contracting malaria going to Kenya on a seven-point Likert scale anchored to "unimportant" and "almost certain" whose points were all verbally labelled.

Procedure. Each participant was randomly assigned to one of the two conditions constructed to follow the experimental design. Individuals could complete the questionnaire at their pace, without any time constraints.

Ethical approval. The study required an ethical approval because it was addressed to a special population (i.e., patients recruited in a hospital). Alike Study 1.1., under the requirement of the Clinical Ethical Committee, the questionnaire could not include any sensitive question that participants might feel unconformable answering. Indeed, the scenario employed in the questionnaire should have included, like others described in this dissertation, the Down syndrome disease as outcome, but that had to be substituted with another in order to cause a lower degree of anxiety (possibly none) in the already sensitive individuals. For the same reason, the risk mentioned in the scenario had to attributed to a participant's friend instead than to the participant her/himself. Anonymity and confidentiality in the treatment of the data gathered were still to be ensured to participants.

Results

A test on the equality of means showed that subjective probability perception was still greater when the probability was expressed as 1 in 200 (*M* = 3.73, *SD* = 1.30) than as 5 in 1000 even when the scenario included an analogy of the jar and ball type (*M* = 3.12, *SD* = 1.02), $t(79) = 2.33$; $p = .02$- see Fig.1. In Study 1.1, where the scenario was the same apart from the fact that visual analogy was not present, the means were very similar (i.e., *M* = 3.84; *SD* = 1.08 for 1 in 200, and *M* = 3.10; *SD* = 1.27 for 5 in 1000), thus the decrease in perceived probability expected for both formats did not take place. The "1 in X" effect, instead, was still present. Thus, it could not be affirmed that the hypothesis of this study was verified. The verbal analogy did not seem a good debiasing technique in helping the "1 in X" effect disappear.

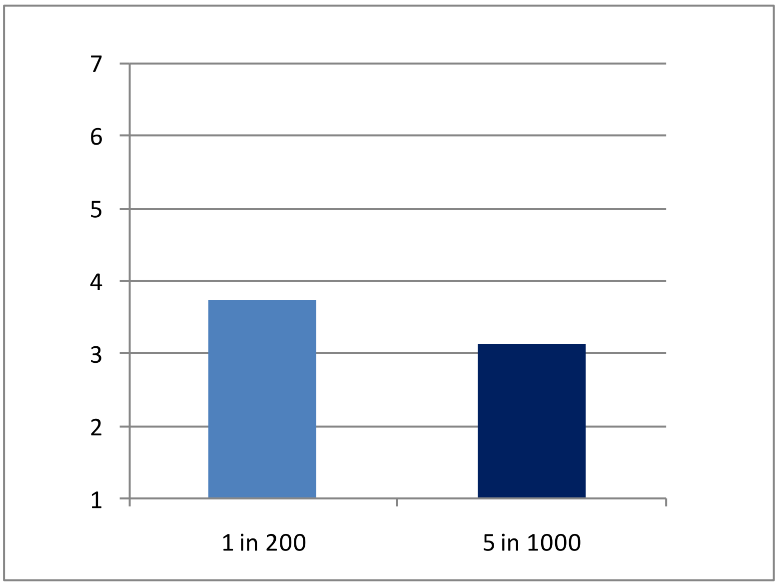


***Fig. 1.* Mean perceived probability in the two experimental conditions**

Discussion

Verbal analogies are techniques frequently used in risk communication to patients in the attempt to enhance the otherwise poor comprehension of small probabilities by laypeople (Edwards, 2003). The only study testing experimentally the efficacy of a verbal analogy in conveying the size of a risk or the probability of its occurrence (i.e., Barilli, Savadori, Pighin, Bonalumi, Ferrari, Ferrari, et al., in press) found that a ball-and-jar analogy could influence the degree of perceived risk for the probability of the outcome happening. The present study originated in the need to verify the possible debiasing effect of employing a similar analogy when the probability was expressed through a ratio in either a”1 in X” or an “N in NX” format. Such issue was interesting for two reasons: 1) analogies are among the suggested communicative interventions to ease the patient comprehension of a risk, thus their study had an ecologic value; and 2) since it had been showed that the “jar and ball” analogy could influence probability perception, it was reasonable to expect an effect of such technique when the probability was conveyed in ratios of different formats.

Results of this study showed that the “1 in X” effect was resistant to the introduction of a verbal analogy for the numerical probability. Indeed, people still showed a tendency to perceive a probability expressed in a “1 in X” format as larger than when expressed in an “N in NX” format. Contrary to what hypothesized, thus, the jar and balls analogy didn’t prove an useful mean to make the “1 in X” effect disappear.

Study 4.3

Goals

Visual displays are frequently employed in communication, and the health domain is not an exception. Indeed, beside classical visual means of representing probability (e.g., vertical bars, pie-charts, icon arrays), some researchers have even developed specific graphical aids to help doctors and health personnel communicating the probability of medical events, like, for instance, the Paling Palette (Paling, 2003).

Graphical displays can help people making decisions based on an accurate understanding of risk information, and they are useful in enhancing comprehension even of health messages (see, for instance, Fagerlin, Wang, and Ubel, 2005; Galesic, Garcia-Retamero, and Gigerenzer, 2009; Garcia-Retamero, Galesic, and Gigerenzer, in press; Lloyd and Reyna, 2001; Okan, Garcia-Retamero, Cokely, and Maldonado, in press; Paling, 2003; Zikmund-Fisher, Ubel, Smith, Derry, McClure, Stark, et al., 2008). Indeed, visual aids benefit of the positive features of visual information, namely saliency and being relatively easy to understand (Jarvenpaa; Woloshin, Schwartz, Byram, Fischoff, and Welch; both cited in Burkell, 2004), thus they improved people' degree of comprehension and recall of information.

As argued by Okan et al. (p. 16), "Visual aids can also improve understanding of risks associated with different medical treatments, screenings, and life-styles (Ancker, Senathirajaha, Kukafka, and Starren, 2006; Galesic et al., 2009; Garcia-Retamero and Galesic, 2010; Lipkus, 2007), and promote consideration of beneficial treatments that have side effects (Waters, Weinstein, Colditz, and Emmons, 2007)". Not only, they can eliminate errors induced by anecdotal narratives (Fagerlin, Wang, and Ubel, 2005), and, more importantly for the case of the present dissertation, reduce biases (see framing effect, Garcia-Retamero and Galesic, in press-a; or denominator neglect, Garcia-Retamero, Galesic, and Gigerenzer, in press; Garcia-Retamero and Galesic, 2009[84], and Stone, Sieck, Bull, Yates, Parks, and Rush, 2003). Even in the case of the "1 in X effect" reviewed in the present dissertation, it could have been that the employment of a visual aid might have helped reducing (or even eliminating) the bias at stake. Thus, it was decided to investigate effects of the use of icon array on probability perception. Icon arrays are graphical representations consisting of a number of circles or other icons symbolizing individuals who are affected by some risk. The choice of that particular aid was done for the following reasons:

1) As suggested by Hawley, Zikmund-Fisher, Ubel, Jancovic, Lucas, and Fagerlin (2008), in the wake of, among others, Price, Cameron, and Butow (2007), pictographs or icon arrays

[84] All these paper had still to be published at the moment of experiment design, data collection, and data analysis.

should be preferred by providers of probability information in shared decision making environments. Hawley and colleagues found, indeed, following an examination of six different formats of graphical presentation (bar graph, pictograph, modified pictograph- "sparkplug"- pie chart, modified pie-chart- "clock graph", and table) on health-related knowledge and treatment behaviour, that pictographs were the aids most effectively conveying both types of knowledge (i.e., the gist and the verbatim one) across numeracy levels; moreover, pictographs were trusted by participants who felt the nature of the information they conveyed as scientific. Furthermore, Feldman-Stewart, Brundage, and Zotov (2007) suggested, that icon arrays (i.e., the so-called "systematic ovals") are among the most easily processed visual aids, after vertical bars, thus they should be preferred in communications with patients, especially the low numerate individuals and the older ones.

2) Icon arrays had been previously employed in the study of phenomena similar to that investigated in the present dissertation, e.g., Ratio-bias (see Rudski and Volksdorf, 2002; but also Stone, Sieck, Bull, Yates, Parks, and Rush, 2003).

In line with what stated by Ancher, Senathirajaha, Kukafka, and Starren (2006, p. 616), i.e., that "Part-to-whole bar charts and part-to-whole sequentially arranged icon arrays probably invoke automatic visual area processing[10] and proportion judgments[11] and can be used to help viewers attend to the mathematical proportion.[34,53,54]", it was predicted that the visual representation could induce participants to properly consider both numerators and denominators of the ratios, and hence the usual "1 in X effect" would not materialize. Thus, the specific hypothesis for this study was the following:

Hypothesis: When a probability expressed through one out of two numerical ratios ("1 in X" or "N in NX") is further illustrated through a visual aid (e.g., a pictogram), the usual "1 in X" effect disappears.

Method

Participants. A total of a hundred and ninety-two individuals from the general population took part in the experiment voluntarily.

The 65.5% of the sample was constituted by women (n = 126), mean age was 25.87 yrs. (SD = 9.69), varying between a minimum of 18 to a maximum of 69 (one participant did not disclose his/her age).

Design and material. Questionnaires employed the usual scenario (see e.g., Study 2.4). Participants read instructions asking them to role-play and imagine themselves as Anna, the mother-to-be described in the scenario as at risk for a child affected by Down syndrome. The probability of having a child with that disease was expressed through a numerical ratio either in the "1 in X" or in the "N in NX" format (1 in 10 versus 10 in 100). Moreover, according to the experimental condition, the probability could be either stated using the numerical ratio only or the numerical ratio followed by a graphical representation (i.e., pictogram). In the ratio plus pictogram conditions, according to the magnitude of the ratio employed, the pictogram depicted either 10 or 100 dots symbolizing the denominator of the numerical ratio, with either 1 red dot or 10 red dots symbolizing the numerator of the ratio (i.e., affected children), and 9 or 90 white ones symbolizing healthy children[85]- see here below in Fig. 1 the pictogram for the 1 in 10 condition, while refer to the Appendix for both visual displays employed in the study. Thus, four different questionnaire versions were build up to follow the 2 x 2 between-subjects experimental design.

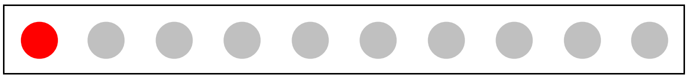


***Fig. 1.* The pictogram (i.e., icon array) used in the 1 in 10 condition**

Within this context, participants were asked a subjective evaluation of the magnitude of the stated probability of having a child affected by Down syndrome on an 11-point Likert scale from "extremely low" to "extremely high", with the half-way point labelled "moderate".

Procedure. Each participant was randomly assigned to one of the four conditions constructed to follow the experimental design. Individuals could complete the questionnaire at their pace, without any time constraints.

Results

Figure 2 displays the mean values of subjective probability in the four experimental conditions. As it can be anticipated from the figure, participants in the control condition (without visual aid) showed the classic "1 in X" effect, but this effect completely disappeared for participants who were provided with a visual aid, namely an equivalent probability perception was attributed to the two format conditions.

A 2 x 2 ANOVA on subjective probability returned no significant main effects of the ratio format, $F(1; 188) = 2.67$, $p = .10$, $\eta_p^2 = .01$, and of the representational format, $F(1; 188) = .71$, $p = .40$, $\eta_p^2 = .01$, on the perceived probability (and a non-significant interaction, $F(1; 188) = 2.83$, $p = .09$, $\eta_p^2 = .02$). Subsequent contrast analyses confirmed that participants in the control condition provided greater ratings for the 1 in 10 ratio ($M = 7.04$, $SD = 3.08$) than for the 10 in 100 ratio ($M = 5.65$, $SD = 2.56$), $t(94) = 2.41$, $p = .02$. However, participants who were given a

[85] The size of the pictogram was 9 x 0.9 cm for the small number ratio condition, and 9 x 9 cm for the large number ratio condition.

visual aid provided remarkably similar ratings for the 1 in 10 ratio (*M* = 5.98, *SD* = 3.08) and for the 10 in 100 ratio (*M* = 6.00, *SD* = 2.92), $t(94) = -.03$, $p = .97$. Thus, Hypothesis 1 was confirmed by data.

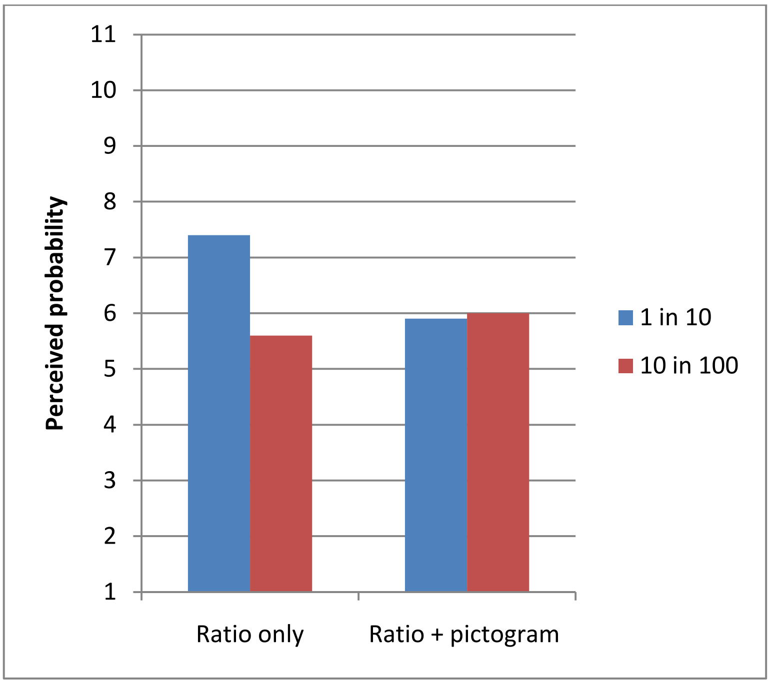


Fig. 2. **Subjective ratings of probability, with and without visual aid, for the two ratio format conditions**

Discussion

A simple visual aid like a pictogram where an instance was represented by a coloured dot made the "1 in X" effect disappear. It is not clear, whether the "1 in X" format yields an overestimation of the probability, or whether the "N in NX" yields an underestimation of the probability. The present study, though, suggested that the first of these propositions is correct. Participants who were provided with a visual aid gave similar assessments of the 1 in 10 and "10 in 100". It turned out that these assessments were significantly lower than that provided about 1 in 10 by participants in the control condition. If one assumes that the assessments given with a visual aid were better calibrated, then the results of the present study suggest that the 1 in X ratio leads to overestimate the probability it expresses.

Although further research will be needed to identify the boundary conditions of this intervention, its effectiveness might be due to the way it transformed probability ratios in readily identifiable, visualised natural frequencies (see Cosmides and Tooby, 1996; Gigerenzer, 1994). Also, as suggested by Ancher, Senathirajaha, Kukafka, and Starren (2006), as well as by Garcia-Retamero, Galesic, and Gigerenzer (in press), icon arrays were successful in eliminating the bias probably because, alike similar graphs, they made part-to-whole relations visually available, thus inducing individuals to properly consider both numerators and denominators of the probability that was depicted. Fuzzy-Trace theory account that "visual displays can help people represent superordinate classes (e.g., the overall number of patients who did and did not receive a treatment)." (Garcia-Retamero et al., in press, p. 2), received support in this study. Finally, pictographs demonstrated particularly useful in another situation not previously reviewed in the literature, thus broadening their relevance and usefulness in medical risk communication.

As per the case of neglect of the denominator, the recent set of investigations just cited has showed that the bias would be lessened when the information becomes easier to process, as when the probability is visually represented by means of icon arrays. Garcia-Retamero and colleagues found support for those accounts of the denominator neglect, like, for instance that of Reyna (1991), which stressed on the particular difficulty in ratio evaluation caused by the need of integrating information across multiple classes (i.e., the comparison of the number of people getting a disease to both those getting and those not getting the disease). Indeed, if usually individuals would simplify the judgment by focusing on the salient class (i.e., the numerator), instead when probability had been represented by means of icon arrays "people [would have] attend[ed] to the relationship between the numerator (i.e., the number of treated and nontreated patients who are affected) and the denominator (i.e., the entire population at risk; see also Lipkus, 2007)" (Garcia-Retamero, Galesic and Gigerenzer, in press, p. 10, information in squared brackets by the writer), thus showing a lessened bias.

General short discussion of the section

The question posed at the beginning of this section (whether or not the "1 in X" effect would be resistant to classic communicative interventions) did not show, as expected, an answer that could be generalised to different types of interventions. Indeed, while a verbal analogy based on urns and balls did not help participants to overcome the bias, a dot-type graphical aid made the effect disappear. It is not clear whether the "1 in X" format yields to an overestimation of the probability, or whether the "N in NX" yields an underestimation of the probability. The present study, though, suggested that the first of these propositions is correct. Participants who were provided with a visual aid gave similar assessments of the 1 in 10 and 10 in 100 ratios. It turned out, that these assessments were significantly lower than that provided about 1 in 10 by participants in the control condition, and broadly similar to that provided about 10 in 100 by participants in the control condition. If we assume that the assessments given with a visual aid were better calibrated, then the results of the third experiment suggest that the "1 in X" ratio leads to overestimate the probability it expresses. Very interestingly, doctors are not affected by the bias. Thus the bias seems to be domain-dependent (as expertise is domain dependent by definition) and part of the laypeople's perceptions of clinical conditions.

CHAPTER 4

Conclusions and General discussion

The present thesis was designed to investigate how the use of superficially different but mathematically equivalent ratio formats affects the magnitude perception of the probability that is conveyed. In particular, focus of investigation was the influence that those expressions, when employed in risk communication of prenatal screening test results, have on prospective parents' perceptions of the chance of having a Down syndrome-affected child. The issue was, in the choice of a ratio format to state the probability, would the magnitude of the numbers employed in the ratio (smaller, e.g., 1 in 10, or larger, e.g., 10 in 100) matter in terms of probability perception? Answering such question was the main aim of the present work.

Research originated from both a theoretical and a practical question. As it has been explained (see beginning of Chapter 3), from a theoretical standpoint studying format effects is one way to validate theories on how the mind processes information about risk, specifically dual process approaches postulating the existence of both an analytical and an experiential way for information-processing. From a practical side, instead, format effects have concrete consequences on individuals' judgments and decisions which are of particular relevance for life, especially in the case they deal with health issues. Hence, results of the investigation could be used to promote health care.

Close inspection of the literature on laypeople's comprehension of different ratio formats revealed individuals' tendency to make less mistakes in operations of the kind that might arise in medical decision making, when handling rates, namely, frequencies of events per unit of population exposed to the risk (e.g., 12 in 100), rather than proportions with a numerator of 1 and shifting denominators (e.g., 1 in 8). At the same time, though, evidence was also found of a habit diffused among health practitioners to privilege "1 in X" formats to "Y in 100/1000" formats in risk communications with patients. While it was apparent that both types of formats, namely rates and proportions, are normally employed in medical practice, it appeared less evident how they can impact laypeople's probability assessments.

With the aim of ascertaining such issue, a variety of approaches in the risk literature was examined- then reviewed in Chapter 2- which have specifically tackled the subject of probability perception as a function of its expression through ratios using smaller or larger numbers. Theories point in two diverging directions that can be summarised in a propensity to either, on the one side, attribute a higher probability judgment to ratios with larger rather than smaller numbers (i.e., denominator neglect, or Ratio-bias: 10 in 100 > 1 in 10), or vice- versa, on the other side, to judge probability as higher when expressed by ratios using larger rather than smaller numbers (i.e., group-diffusion effect: 1 in 10 > 10 in 100). However, implications of both groups of theories could not entirely be transferred to the domain of interest of the present work, the risk of having a Down syndrome-affected child, mainly because of problems of ecological validity. These problems were also those which prompted us to test empirically

the applicability of predictions derived from the reviewed approaches to the specific case under study, provided that the necessary adjustments to the original experimental designs be made.
Twelve main studies were conducted testing a total of 1673 individuals, employing a between-subjects rather than a within-subjects presentation of superficially different but substantially equivalent ratio formats contextualised in scenario paradigms; a separate rather than joint evaluation of the formats was required to participants, and numerator and denominator of the ratios were both explicitly stated. Main objectives of the investigation were the following: 1) ascertaining whether denominator neglect (i.e., Ratio-bias) or, on the opposite, numerator neglect (i.e., group-diffusion effect) would occur when the probability of having a Down syndrome-affected child was expressed by means of a ratio in single presentation conditions using (comparatively) smaller or larger numbers; 2) delineating boundaries of the bias; 3) testing an explanation for the bias stemming from dual-process theories, i.e., the affective hypothesis; 4) examining potentially corrective communication measures. Analyses focused mainly on the magnitude of perceived probability of the ratios as assessed on Likert scales by participants.

Summary of main empirical results

The first section of the chapter dedicated to the empirical investigations illustrated an irrational tendency found in study participants to evaluate the same .005 probability of a clinical condition (malaria, in Study 1.1, while Down syndrome, in Study 1.2) as higher when expressed as “1 in 200” rather than as “5 in 1000”. The higher subjective assessment found for ratios employing (comparatively) smaller rather than (comparatively) larger numbers was in line with results of group-diffusion studies (i.e., numerator neglect), hence, at the same time, it disconfirmed the applicability of an explanation in terms of Ratio-bias (or denominator neglect) to the case examined. Nevertheless, as the group-diffusion effect explanation had been validated with very different scenarios (see, for instance, bacteria, or carcinogen scenarios), it could well have been that such rationale for the effect did not hold in the condition at study.
The degree of worry for the condition showed the same trend of perceived probability, but since the former was showed to depend on the latter, it was decided that in the next studies the focus be mainly on perceived probability for the condition.

In the second section of the chapter, boundaries of the identified bias were delineated. Along studies, it was observed that laypeople perceived the probability of having a Down-syndrome-affected child as higher when it was expressed, respectively:

- as 1 in 200 than both as 5 in 1000, and as 50 in 10,000, while these last two formats did not evoke different probability assessments- Study 2.1;

- in formats using 1 at the numerator and a shifting denominator (i.e., "1 in X") rather than in formats using 1000 at the denominator and a shifting numerator (i.e., "Y in 1000")- Study 2.2;
- as 1 in 12 rather than as 10 in 120, i.e. in hundred-sized denominators (i.e., "Y in 100")- Study 2.3;
- as 1 in 25 rather than 0.4 in 10, and as 1 in 25 rather than 10 in 250- Study 2.6.

Altogether, these evidences seemed to point to some specific features of the ratio format using 1 at the numerator in respect to "N in NX" formats. The tendency to judge a probability as higher when expressed in a "1 in X" format rather than in a generic "N in NX" format could be dubbed the "*1 in X effect*". However, as it has been observed in the summary of Section 2, it was deemed necessary to exclude, that the higher probability perception found in "1 in X" formats reflected a more general focus on the denominator of the ratio. Thus, it was showed, that comparison of average probability perceptions of two equivalent ratios, none appearing in formats featuring directly a "1 in X", but with denominators of different magnitude, did not differ significantly.

Then, a short explorative study (i.e., 2.5) was carried out with the aim of potentially isolating one of those facets that were supposed to distinguish the "1 in X" format from the others, i.e., the degree to which they prompted individuals to operate mathematical transformation on their superficial appearance. A discrepancy between the two types of formats was found in the fact, that while the majority of participants declared to be used to transform "N in NX" formats, the majority also declared not to be used to that when dealing with "1 in X" expressions. Such idiosyncrasy of the "1 in X" format received additional confirmation in the following study, i.e. 2.6, which compared evaluation of the probability elicited by that ratio when in the usual Down syndrome risk scenario with other not previously tested "N in NX" formats. Results showed, that the presence of the "1 in X" format in the paired comparisons between different ratios was the *condicio sine qua non* for the occurrence of the bias in perceived probability.

Having established the effect, in the subsequent section of the chapter (i.e., the third one), we dealt with an affective explanation for the bias under study (see, e.g., Slovic and Peters, 2006) stemming from literature on dual-process approaches, namely that the "1 in X" format evokes a different degree of emotional response than any "N in NX" format. Neither an explicit (i.e., self-report, see Study 3.1) nor an implicit (i.e., elicitation of mental images, see Study 3.2) measure of the emotional answers prompted could support the affective hypothesis of those theorists. Data of the second of these studies, furthermore, evidenced a tendency opposite to that documented in Slovic, Monahan, and MacGregor (2000, see Chapter 2, paragraph 1.2.3.1), namely the theory of the "imaginability of the numerator". If on the one hand, then, confirmation of two-system views of risks received direct support in the study (i.e., 2.1) which showed a strong relation between appearance of the bias and a mainly experiential-way of reasoning, on the other hand, in this third section the ad hoc explanation elaborated by Slovic and colleagues to account for Ratio-bias did not show success. However, some methodological problems could have played a role in the absence of the detection of any difference at all between ratio formats in the emotional answer elicited.

An explanation for the "1 in X effect"

You may say, it's one in a hundred. ***But what if I'm that one?***
Van Steenkiste, Van der Weijden, Timmermans, Vaes, Stoffers, and Grol (2004, bold font added)

It is our impression, instead, that overall results could rather reflect the idea that the "1 in X" format conveys more identifiable information than any other format. Indeed, the very fact that people did not transform the "1 in X" format might mean that it is considered "satisfactory" to produce a judgment or to take a decision (in terms of "satisficing", Simon, 1955). One not tested hypothesis then, is that people do somehow "like" the "1 in X" format as it is, because they can easily "see" their case in the "1 case over X cases". It is possible in fact, that the presence of the number 1 at the numerator of a ratio expressing the probability of an event in single presentation conditions, as in the cases here analysed, becomes a highly salient feature at the eyes of individuals, given laypeople's difficulties in forming a judgment of probability when evaluating its numerical expression in isolation. As such, that facet would receive special attention and it would indirectly work as a warning sign for the person, denoting the possibility of being that "1" individual experiencing the bad consequences of the negative event described. It is apparent, that the effect postulated is opposite to CEST "experiential-learning" principle (e.g., Kirkpatrick and Epstein, 1992).

In light of these considerations, it could be reasonable to attend, that those participants to this thesis experiments who were assigned a "1 in X" format condition more easily imagined their future son/daughter identified with that single instance at the numerator than those participants who were assigned an "N in NX" format condition. That identification, in turn, could have resulted in an increase of the "experiential" sensation of risk in those persons (here, "experiential" has to be intended as the "experiential route vs. "analytical route", *à la* Slovic, Peters, Litchenstein, and McGregor, 2004). The augmented affective reaction, in turn, could have induced a higher perception of the probability, compared to the conditions where a 1 in the numerator was not present (i.e., "N in NX" conditions).

This causal relation between affective reaction and augmented perception of the probability has been explained by some scholars in slightly differing versions of what has then come to be known as the "identifiable victim effect" (Kogut and Ritov, 2005; Jenni and Loewenstein, 1997; Schelling, 1968; Small and Loewenstein, 2003).

According to this effect, people would react in a different way to identifiable rather than to statistical victims, an idea originally introduced by Schelling (1968). He noticed, that while the death of a particular individual invokes "[...] anxiety and sentiment, guilt and awe, responsibility and religion, most of this awesomeness disappears when we deal with statistic death". In his view, the explanation of this phenomenon had to be researched in the degree of emotional responses stimulated by identifiable rather than statistical victims. Small and Loewenstein (2003) demonstrated, that such different reaction to identifiable rather than statistical victims depends on the stronger affective reactions provoked by the identifiable target (not *a* victim, but *the* victim). In fact, this identifiability can provoke the adoption of the other's perspective, namely "imagining how that person perceives the situation and how he or she feels as a result (Davis, 1994, Batson, Klein, Highberger, and Shaw, 1995)" (Kogut and Ritov, 2005, p. 158). Indeed, the adoption of the other's perspective "[...] is likely to evoke, in the perceiver, feeling of empathy (sympathetic, moved, compassionate, tender, warm, softhearted, etc.) and also feelings of distress for that other person (Batson, Early, & Salvarani, 1997)" (Kogut and Ritov, 2005, p. 158)[86]. In the last resort, hence, arouse of empathic emotions is deemed as the cause of the higher judgment of probability for the single identifiable victim ("1 in X" format, in the present case) than for the statistical one ("N in NX" format) – see the schematic representation here below.

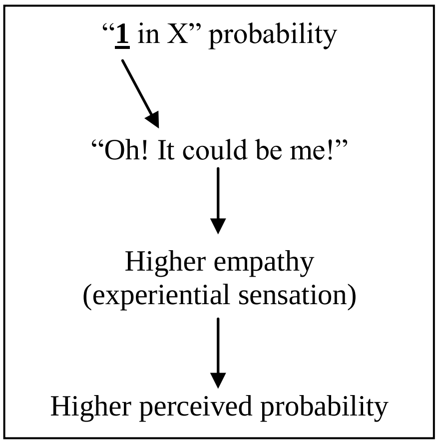


***Fig. 1.* The "1 in X effect" in perceived probability as a function of increased empathy with the "1" victim**

[86] Two rather different examples of empathy aroused by single identifiable victims and not by statistical victims are exemplified in the following two quotations, 1) Nobel laureate Albert Szent-Gyorgi (quoted in Featherstonehaugh, Slovic, Johnson, and Friedrich, 1997, p. 283) stated, "I am deeply moved if I see one man suffering and would risk my life for him. Then I talk impersonally about the possible pulverization of our big cities, with a hundred million dead. I am unable to multiply one man's suffering by a hundred million." and 2) Anne Dillard (quoted in Slovic, 2007) affirmed, "There are 1,1985,000 people alive in China. To get a *feel* for what this *means*, simply take yourself- in all your singularity, importance, complexity, and love- and multiply by 1,1985,000. See? Nothing to it."

Deahene and Melher (1992)'s affirmations (summarised here below) come in support of this hypothesis of a higher identifiability of the information conveyed by the "1 in X" format, provided that one spouses the view considering the Fechner' law (see Fechner, 1860) and the existence- even among numbers- of points of reference (see, Rosch, 1975; but also Milikovski and Elshout, 1994; 1995) as correctly characterising how the mind encodes numerical magnitudes[87]. Deahene and Melher found, that those number-words used for precise denotation of numerosity and to express those reference numbers suited for use in estimates (in particular, 1, 10, 12, 15, 20, 50, and 100), are more frequently used than the others in several different languages. Thus, we can think of these numbers-words as representing more available, and then more likely evoked, examples of some numerosities in people's mind. In particular, authors not only observed a decrease of frequency in the use of number-words with the increase in numerical magnitude (with local increase for the above mentioned reference numerals), but also, they highlighted, that number one is the most frequent in absolute of all numbers-words employed by people. Thus, it is our suggestion, that it could well be the case, that when people evaluate a ratio using a 1 as numerator, such number implicitly evokes the highest number of mental instances[88] compared to (almost) all other possible numerators of "N in NX" formats. This unconscious availability of more mental examples (similarly, to some extents, to Kahneman and Tversky's availability heuristic) arouses in turn the higher probability "sensation" -in terms of, for instance, Slovic, Peters, Litchenstein, and McGregor (2004)'s experiential feeling- and hence prompts individual to express the higher judgment of probability, as observed in the data.

Other explanations for the "1 in X effect"

On another vein, the bias might be the results of another factor as well, as we do not think a systematic tendency originates from only one factor, but there could be different co-occurring factors causing it. The erroneous attribution of different probability assessments to superficially different but substantially equivalent formats could arise in those individuals motivated to process the information. Indeed, much of the studies pointed to a motivational explanation of the bias, in other words it seems, that only those individuals that truly processed the numerical information (because they were familiar with it, because they were capable of, because they had children or were parents-to-be, for any other motivational reason) were subject

[87] There is no universal agreement on that explanation, as Peters, Slovic, Västfjäll1, and Mertz (2008, p. 620) have clearly expressed, "Although competing mathematical formulations of the precision of mental-number-line [MNL] representations exist (Dehaene & Changeux, 1993; Gallistel & Gelman, 2005) […]". Nevertheless, authors also observed (ibid.), that broadly speaking the two competing views on the MNL "[…] make similar behavioral predictions in number comparison studies and (we believe) in decision making."

[88] Our deduction, however, is different from the "experiential-learning principle" of CEST, which instead is more coherent with Deahane and Melher's findings, in that both addressed integer numbers.

to the bias. Therefore, it should be thought that there is a sort of processing threshold which needs to be exceeded in order for the bias to emerge. This threshold, anyway, needs effort to be overcome, as some of the literature on numerical processing seems to indicate (for implications in decision making, see Peters, Slovic, Västfjäll, and Mertz, 2008). Indeed, despite it is clear since time (see, e.g., Dehaene, 1996), that in the presence of numeric information nonverbal representations of numerosity (i.e., intuitive representations of the mental number line) are spontaneously activated regardless of format (e.g., dots, Arabic number, number word), “Number intuitions, however, are limited in their representational power and do not directly support concept of fractions, probabilities, or even the precise numbers important to many decisions” (Peters et al. 2008, p. 619). Given the particularly difficult nature of ratio concepts, thus, it might have well been the case that only those individuals motivated to process the ratio information (for many reasons, as seen) could overcome the necessary threshold to make the (biased) representation of the ratio possible, i.e. to somehow sense the numerical expression. The last affirmation could be tested in future studies manipulating, for instance, the motivation to elaborate the information by means of monetary reward (i.e., comparing the performance of individuals in a condition where they received payment with that of participants in a condition where they did not receive payment); or by means of different levels of incentives to participants in such a way that only some participants would be motivated to process information, hence would exceed the processing threshold. Also, since the issue of how the concepts of ratios and fractional numbers are represented in our mind has only very recently started to receive attention from researchers (see, for instance, Bonato, Fabbri, Umiltà, and Zorzi, 2007), it is possible that while plugging that dramatic gap existent in the literature more precious indications will become available for researchers in judgment and decision making.

Another result in the thesis was coherent with this threshold motivation hypothesis, although it should be replicated, given that several problems occurred in the assessment of the REI scale from which it follows. The result is, that the “1 in X” bias was associated with a prevalence of the experiential thinking style, though mainly with a low-experiential instead than high-experiential decision style. Indeed, results on lowly experientials, exactly because of their moderate degree of experentiality in making evaluations and taking decisions, denoted that they probably elaborated at least to some extent the ratio on which they were asked a magnitude assessment. On the contrary, results on highly experientials leaded to think that these individuals, who normally process information in an intuitive way, did not reach the above-cited threshold, in that they did not really process the numbers; in other words they seemed to base their evaluations on some gross clues in the ratios, like the presence of great-sized denominators.

A convergence of the two factors?

The two factors could also converge: The “1 in X” format conveys a more identifiable meaning than other formats (like the “N in NX”) for the individual which, hence, is more likely to process it, thus leading to higher judgments of probability.

Results of a recent study (i.e., Timmermans and Oudhoff, in press, personal communication) render this interpretation credible as authors would highlight the importance of the advantage that smaller denominators have over larger ones in ratios, in that the former are easier to visualize than the latter[89] (see, on this regard, the literature on the mental representation of integer numbers, becoming fuzzier as numbers get bigger i.e., Weber’s law; but also Moyer and Landauer, 1967). According to Timmermans and Oudhoff, since risk needs to be concrete (i.e. to refer to concrete events or people), if one wants to ease patient’s grasp of the probability meaning, smaller denominators are advantageous compared to larger ones. Not different is the take-home message of another very recent publication (i.e., Garcia-Retamero and Galesic, in press-b), in that authors urged health practitioners to avoid expressing probability in ratios using large denominators because the latter especially favour more biased interpretations (like denominator neglect); while, instead, they suggest to use ratios employing smaller, more plausible denominators.

Further studies should then address the easiness to process the ratio information in isolation when in smaller and larger numbers trough standard two task paradigms, namely making use of cognitive load, for instance, or inducing participants in conditions of time pressure.

Implications for practice

It is apparent, that present findings have a particular relevance for doctors and health personnel in charge of risk communication. These experts should be warned, in light of these results, that their choice of a ratio format rather than another to express the probability of a health-related outcome (here, having a Down syndrome-affected child) does not come without consequences on patients’ assessments. In line with literature on expertise, Study 4.1 showed that medical doctors were not to prone to the bias, probably exactly because of their expertise in the field of application, and/or because they are individuals more used to perform operations on mathematical formats, as their profession requires on a daily basis. The same expertise that seems to prevent them from the bias could however be the one responsible of making them insensitive to issues of formats effects in risk communication. Therefore, the decision to test communicative interventions able to attenuate or even eliminate the “1 in X” effect.

[89] Obviously, authors must start from the assumption (empirically proven, or only theoretical, it is not known) that individuals mainly focus on denominators in the ratios to draw some clues for intuitively assessing their magnitude.

Results of Study 4.2 showed that the bias persisted when adding the balls and jar verbal analogy to the numerical risk presentation, a technique that is common practice in risk communication and that consists in encouraging people to build a visual representation of the probability. Instead, the bias disappeared when individuals were showed a visual depiction of the probability by means of a pictogram where an instance was represented by a dot (Study 4.3). Effectiveness of the latter intervention might reside in the fact, that icon arrays- as suggested by Ancher, Senathirajaha, Kukafka, and Starren (2006), as well as by Garcia-Retamero, Galesic, and Gigerenzer (in press) in the wake of Fuzzy-Trace theory (see, e.g., Reyna and Brainerd, 1995)- alike similar visual aids, make part-to-whole relations visually available. Thus, they would induce individuals to consider both numerator and denominator of the ratio. Despite the need of more evidences, on the basis of these results it can be tentatively argued, that providing icon arrays in addition to numerical information would be an effective method of preventing difficulties in understanding health relevant risk communications.

Limitations

The most evident concern for the studies included in this thesis is the fact, that the bias is not a systematic tendency, as small changes in superficial details could easily attenuate or remove the effect; in other words, the disclosed tendency is rather weak. Anyway, we think that this is intrinsic in a bias of that type, in that, indeed, it is based on simple changes in the magnitude of numbers, namely a facet that can be reached by definition in an unlimited number of even minuscule modifications. This is not a justification of course; however it does testify the difficulty of implementing a conclusive research in this area. In an effort to give a little contribution, with no claim of being complete, we decided to follow some avenues, thus obviously leaving other ones unexplored. For instance, it could have been useful to study whether the bias is present with higher probability values that those examined, like 30%, 50%, 70%, 90% (in the wake of, for instance, Pacini and Epstein’s study, 1999-b). That would have allowed to define whether the tendency is for some reasons unique to low and very low probabilities like the ones tested in this research work (i.e., 10%, 1%, or lower), or instead extends to other probability magnitudes. Hence, it goes without saying, that there is still a vivid need for further investigation that can explore, describe, and account for the bias studied.

A second concern for the studies included in this thesis is, that only one methodology was used, namely questionnaires or paper-and-pencil tasks[90] focused mainly on one dependent measure, i.e., perceived probability. While it is surely true that the number of studies performed and that of the individuals tested made results obtained from that methodology appear very strong, it is also true that crossed verification by means of different methodological paradigms

[90] Two of these studies, indeed, were collected on the Web, but the type of task required was the same of real paper-and-pencil tasks (i.e., answering to a set of questions).

could as well have strengthened the validity of findings. For instance, as already affirmed, using time-pressure paradigms could have helped, as well as the employment of eye-tracking techniques, to detect which is the focus of patient's attention. Similarly, other measures (i.e., dependent variables, for instance Willingness to Pay or Willingness To Risk) could have offered interesting hints on the trend of the bias. Nevertheless, it has to be argued that, for the specific aims of the present investigation, analyses of judgment on the level of probability of occurrence of the event appeared as the most appropriate, also in line with literature on decision making, and on risk in particular.

Another concern for the studies included in this thesis regarded the type of sample adopted. Indeed, while it is true that participants of possibly different ages and occupational sectors were recruited, nevertheless it has to be recognized, that they still were convenience samples. It can be said, however, that 1) since practically all adult individuals in the population until a certain age (different for men and women) are possible prospective parents, and since no main individual differences related to demographics have been found, it can be affirmed that the present findings are not misleading of the general population's way of judging; 2) importantly, patients in the specific condition at stake (i.e., parents-to-be, or individuals that have just become mother/father, hence namely those who have to evaluate/ have just evaluated screening test results on the risk of having a Down syndrome-affected child) have been tested in two studies.

As a final consideration, future research could aim to enhance the validity of findings of the present work. It has to be noticed that, apart from one study (i.e., the first) which addressed the perception of probability referred to contracting malaria, all others studies analyzed the risk of having a Down syndrome-affected child, and as such the instrument and the materials used focused on that medical domain. Hence, it is not clear a) whether the "1 in X effect" occurs mainly for the clinical condition analyzed (i.e., the Down-syndrome risk), and b) if the bias can influence risky decision making competence in other important domains. For what regards the first point, one can ask, in other words, whether the "1 in X effect" always occurs when the probability is referred to medical conditions other than the risk of having a Down syndrome-affected child. Are there diseases, that for some intrinsic features, lead individuals to judge in a way that disconfirms the "1 in X effect"? Thus, another avenue for future research would be to replicate the finding of this thesis on other medical domains.

Not only, as for what concerns the second point, it would be extremely interesting to investigate even non-medical outcomes, both negative and positive ones- one promising avenue is, for sure, the field of marketing/advertising/promotion, that of protective measures/insurance/life saving, but maybe also that of investments. In such a way, boundaries of the bias would be further delineated and the way in which individuals evaluate ratio formats would be further disclosed.

BIBLIOGRAPHIC REFERENCES

Acredolo, C., O' Connor, J., Banks, L., and Horobin, K. (1989). Children's ability to make probability estimates: skills revealed through application of Anderson Functional Measurement Methodology. *Child Development*, *60*, 933-945.

Adams, A. and Smith, T. (2003). Risk communication and anesthesia. In J. A. Lack, A. Rollin, G. Thoms, L. White, and C. Williamson (Eds.), *The Royal College of Anesthetists. Raising the standard: information for patients. Principles, samples of current practice and new text* (pp. 77-86). Alden Press Ltd: Oxford (UK).

Alaszewski, A. and Horlick-Jones, T. (2003). How can doctors communicate information about risk more effectively? *British Medical Journal*, *327*, 728-731.

Alhakami, A. S. and Slovic, P. (1994). A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis*, *14*(6), 1085-1096.

Alonso, D. and Fernández-Berrocal, P. (2003). Irrational decisions: Attending to numbers rather than ratios. *Personality and Individual Differences*, *35*(7), 1537-1547.

Amsel, E., Close, J., Sadler, E., and Klaczynski, P. (2009). College students' awareness of irrational judgements on gambling tasks: A dual process account. *The Journal of Psychology*, *143*(3), 293-317.

Ancker, J. S., Senathirajaha, Y., Kukafka, R., and Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*(6), 608-618.

Anderson, H. H. (1980). Information integration theory in developmental psychology. In F. Wilkening, J. Becker, and T. Trabasso (Eds.), *Information integration by children* (pp.1-45). Hillsdale, NJ: Erlbaum.

Antonarakis, S. E., Petersen, M. B., McInnis, M. G., Adelsberger, P. A., Schinzel, A. A., Binkert, F., et al. (1992). The meiotic stage of nondisjunction in trisomy 21: Determination by using DNA polymorphisms. *American Journal of Human Genetics*, *50*(3), 544-550.

Appelt, K. C., Milch, K. F., Handgraaf, M. J. J., and Weber, E. U. (2010). Decision Making Individual Differences Inventory and review: Guidelines and tools for the use of individual differences in judgment and decision-making research. Manuscript submitted for publication.

Appelt, K., Milch, K., Handgraaf, M., and Weber, E. DMIDI (Decision Making Individual Differences Inventory). Accessed March, 12, 2010, from http://www.dmidi.net/

Ayal, S. and Zakay, D. (2009). The perceived diversity heuristic: The case of pseudodiversity. *Journal of Personality and Social Psychology*, *96*(3), 559-573.

Barbey, A. K. and Sloman, A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavior and Brain Sciences*, *30*, 241-297.

Barilli, E., Savadori, L., Pighin, S., Bonalumi, S., Ferrari, A., Ferrari, M., et al. (in press). From chance to choice: the use of a verbal analogy in the communication of risk. *Health, Risk, and Society.*

Baron, J. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty*, *14*, 301-309.

Benthin, A., Slovic, P., Moran, P., Severson, H., Mertz, C. K., and Gerrard, M. (1995). Adolescents' health-threatening and health-enhancing behaviours: A study of word association and imagery. *Journal of Adolescent Health, 17*, 143-152.

Bjorklund, F. and Backstrom, M. (2008). Individual differences in processing styles: Validity of the Rational–Experiential Inventory. *Scandinavian Journal of Psychology*, *49*, 439-446.

Black, W. C., Nease, R. F., and Tosteson, A. N. (1995). Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *Journal of the National Cancer Institute*, *87*, 720-731.

Bonato, M., Fabbri, S., Umiltà, C., and Zorzi, M. (2007). The mental representation of numerical fractions: Real or integer? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1410-1419.

Bonini, N., Del Missier, F., and Rumiati, R. (2008). Decisione degli esperti e differenze individuali nella decisione. In N. Bonini, F. Del Missier, and R. Rumiati (Eds.), *Psicologia del giudizio e della decisione* (pp. 225-243). Il Mulino, Bologna.

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

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

Brase, G. L. (2002). Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making*, *15*, 381-401.

Brun, W. and Teigen, K., (1988). Verbal probabilities: Ambiguous, context dependent or both? *Organizational Behaviour and Human Decision Processes*, *41*, 390-404.

Burkell, J. (2004). What are the chances? Evaluating risk and benefit information in consumer health materials. *Journal of the Medical Library Association*, *92*, 200-208.

Cacioppo, J. T., Petty, R. E., and Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*, 306-307.

Callahan, P. (1989). *Learning and development of probability concepts. Effects of computer-assisted instruction and diagnosis*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Calman, K. C., Bennett, P. G., and Coles, D. G. (1999). Risks to health: some key issues in management, regulation and communication. *Health, Risk and Society*, *1*, 107-116.

Chaiken, S. and Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.

Coombs, C. H. (1975). Portfolio theory and the measurement of risk. In M. F. Kaplan and S. Schwartz (Eds.), *Human judgment and decision processes* (pp. 63–85). New York: Academic Press.

Collins COBUILD English Language Dictionary (10^{th} ed.), Bungay, Suffolk, UK: Richard Clay Ltd.

Cosmides, L. and Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1-73.

Cuite, C. L., Weinstein, N. D., Emmons, K., Colditz, C. (2008). A test of numerical formats for communicating risk probabilities. *Medical Decision Making, 28,* 377-384.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain.* New York: Putnam.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience*, *1*, 47-68.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S. and Melher, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43,* 1-29.

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

Dewhurst, M. E., McCarthy Veach, P., Lampman, C., Petraitis, J., Kao, J., and LeRoy, B. (2007). Probability biases in genetic problem solving: A comparison of undergraduates, genetic counseling graduate students, and genetic counselors. *Journal of Genetic Counseling*, *16*(2), 157-163.

Dodero, N., Baroncini, P., and Manfredi R. (2001). *Moduli di lineamenti di matematica-triennio scientifico c.* Ghisetti e Corvi Editori, Bergamo.

Dohle, S., Keller, C., and Siegrist, M. (2010). Examining the relationship between affect and implicit associations: Implications for risk perception. *Risk Analysis*, *30*(7), 1116-1128.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioural Brain Science*, *16*, 681-735.

Edwards, A. (2003). Communicating risks through analogies. *British Medical Journal*, *327*(7417), 749-749.

Edwards, A. G. K. and Elwyn, G. (2001). I. Communicating and understanding risk. Understanding risk and lessons for clinical risk communication about treatment preferences. *Quality in Health Care*, *10*, i9-i13.

Edwards, A. G. K., Elwyn, G., and Mulley, A. (2002). Explaining risks: Turning numerical data into meaningful pictures. British Medical Journal, *324,* 827-830.

Eiser, J. (1998). Communication and interpretation of risk. *British Medical Bulletin*, *54,* 779-790.

Emanuel, E. J. and Emanuel, L. L. (1992). Four models of the physician-patient relationship. *The Journal of the American Medical Association*, *267*(16), 2221-2226.

Epley, N. and Gilovich, T. (2006). The anchoring-and-adjustment heuristic. *Psychological Science*, *17*(4), 311-318.

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 Press.

Epstein, S. (1993). Implications of cognitive-experiential self-theory for personality and developmental psychology. In D. Funder, R. Parke, C. Tomlinson-Keasey, and K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 399-438). Washington, DC: American Psychological Association.

Epstein, S. (2003). Cognitive-experiential self-theory of personality. In Millon, T., and Lerner, M. J. (Eds.), *Comprehensive Handbook of Psychology, Volume 5: Personality and Social Psychology* (pp. 159-184). Hoboken, NJ: Wiley & Sons.

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

Epstein, S. and Pacini, R. (2000-2001). The influence of visualization on intuitive and analytical information processing. *Imagination, Cognition, and Personality, 20*, 195-216.

Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1995). *Individual differences in heuristic processing.* Unpublished manuscript, University of Massachusetts at Amherst.

Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390-405.

Estrada, C., Martin-Hryniewicz, M., Peek, B. T., Collins, C., and Byrd, J. C. (2004). Literacy and numeracy skills and anticoagulation control. *The American Journal of the Medical Sciences, 328*, 88-93.

Fagerlin, A., Wang, C., and Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making*, *25*(4), 398-405.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., and Smith, D. M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale (SNS). *Medical Decision Making*, *27*(5), 672-680.

Fechner, G. T. (1860). *Elemente der Psychophysik I*. Leipzig.

Feldman-Stewart, D., Brundage, M. D., Van Manen, L., and Svenson, O. (2004). Patient-focused decision-making in early-stage prostate cancer: Insights from a cognitively based decision aid. *Health Expectations: An International Journal of Public Participation in Health Care & Health Policy*, *7*(2), 126-141.

Feldman-Stewart, D. Brundage, M. D., and Zotov, V. (2007). Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. *Medical Decision Making, 27*, 34-43.

Feldman-Stewart, D., Kocovski, N., McConnell, B. A., Brundage, M. D., and Mackillop, W. J. (2000). Perception of quantitative information for treatment decisions. *Medical Decision Making*, *20*, 228-238.

Fetherstonhaugh, D., Slovic, P., Johnson, S. M., and Friedrich, J. (1997). Insensitivity to the value of human life: A Study of psychophisical numbing. *Journal of Risk and Uncertainty*, *14* , 283- 300.

Finucane, M. L., Alhakami, A., Slovic, P., and Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*(1), 1-17.

Finucane, M. L., Peters, E., and Slovic, P. (2003). Judgment and decision making: The dance of affect and reason. In S. L. Schneider and J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 327-364). Cambridge: Cambridge University Gates.

Finucane, M. L., Slovic, P., and Mertz, C. K. (2000). Public perception of the risk of blood transfusion. *Transfusion*, *40*, 1017-1022.

Fischbein, E., Pampu, I., and Manzat, I. (1970). Comparisons of ratios and the chance concept in children. *Child Development*, *41*(2), 377-389.

Fischoff, B., Lichtenstein, S., Slovic, P., Derby, S., and Keeney, R. (1981). *Acceptable risk*. New York: Cambridge University Press.

Fischoff, B., Slovic, P., and Lichtenstein, S. (1982). Lay foibles and expert fables in judgments about risk issue. *The American Statistician*, *36*(3), 240-255.

Galesic, M. and Garcia-Retamero, R. (2010). Statistical numeracy for health. A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*, *170*(5), 462-468.

Galesic, M. and Garcia-Retamero, R. (in press). Graph literacy: a cross-cultural comparison. *Medical Decision Making*.

Galesic, M., Garcia-Retamero, R., and Gigerenzer, G. (2009). Using icon arrays to communicate medical risks to low-numeracy people. *Health Psychology*, *28*, 210-216.

Garcia, S. M. and Tor, A. (2009). The N-effect. More competitors, less completion. *Psychological Science*, *20*(7), 871-877.

Garcia-Retamero, R. and Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: A cross-cultural comparison. *American Journal of Public Health*, *99*(12), 2196-2202.

Garcia-Retamero, R. and Galesic, M. (2010). Who profits from visual aids: Overcoming challenges in people's understanding of risks. *Social Science and Medicine*, *70*, 1019-1025.

Garcia-Retamero, R. and Galesic, M. (in press-a). How to Reduce the Effect of Framing on Messages About Health. *Journal of General Internal Medicine.*

Garcia-Retamero, R. and Galesic, M. (in press-b). Using plausible group sizes to communicate information about medical risks. *Patient Education and Counseling*.

Garcia-Retamero, R., Galesic, M., and Gigerenzer, G. (in press). Do icon arrays help reduce denominator neglect? *Medical Decision Making.*

Gates, E. (2004). Communicating risk in prenatal genetic testing. *Journal of Midwifery and Women's Health*, *49*(3), 220-227.

Gentner, D., Holyoak, K. J., and Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science*. Cambridge, MA: MIT Press.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright and P. Ayton (Eds.), *Subjective probability* (pp. 129-161). New York: Wiley.

Gigerenzer, G. (2002). *Quando i numeri ingannano*. Raffaello Cortina Editore, Milano.

Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty.* New York, NY, US: Oxford University Press.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science, 3*(1), 20-29.

Gigerenzer, G. and Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *British Medical Journal*, *327*, 741-744.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., and Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8,* 53-96.

Gigerenzer, G. and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684-704.

Gigerenzer, G., Todd, P. M., and the ABC Group (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Goyder, E., Barratt, A., and Irwig, L. M. (2000). Telling people about screening programmes and screening test results: How can we do it better? *Journal of Medical Screening*, *7*, 123-126.

Green, J. M., Hewison, J., Bekker, H. L., Bryant, L. D., and Cuckle, H. (2004). Psychosocial aspects of genetic screening of pregnant women and newborns: A systematic review. *Health Technology Assessment, 8*(33).

Green, J. and Statham, H. (1993). Psychological aspects of genetic counselling. In T. Marteau and M. Richards (Eds.), *The troubled helix: Social and Psychological aspects of the new human genetics* (pp. 140-163). Cambridge, UK: Cambridge University Press.

Grimes, D. A. and Snively, G. R. (1999). Patients' understanding of medical risks: Implications for genetic counseling. *Obstetrics and Gynecology*, *93*, 910-914.

Gurmankin, A. D., Baron, J., and Armstrong, K. (2004). The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. *Medical Decision Making*, *24,* 265-271.

Hall, S., Chitty, L., Dormandy, E., Hollywood, A., Wildshut, H. I. J., Fortuny, A., et al. (2007). Undergoing prenatal screening for Down's syndrome: presentation of choice and information in Europe and Asia. *European Journal of Human Genetics, 15,* 563-569.

Halpern, D. F., Blackman, S., and Salzman, B. (1989). Using statistical information to assess oral contraceptive safety. *Applied Cognitive Psychology*, *3*, 251-260.

Hamm, R. M (1988). Clinical intuition and clinical analysis: Expertise and the cognitive continuum. In J. Dowie and A. Elstein (Eds.), *Professional judgment: A reader in clinical decision making*. Cambridge, UK: Cambridge University Press.

Hawley, S. T., Zikmund-Fisher, B., Ubel, P., Jancovic, A., Lucas, T., and Fagerlin, A. (2008). The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling*, *73*, 448-455.

Hecht, S. A., Close, L., and Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology*, *86*(4), 277-302.

Hertwig, R., Zangerl, M. A., Biedert, E., and Margraf, J. (2008). The public's probabilistic numeracy: How the tasks, education, and exposure to games of chance shape it. *Journal of Behavioral Decision Making*, *21*, 457-470.

Ho, A. S. Y. and Leung, K. (1998). Group size effects on risk perception: A test of several hypothesis. *Asian Journal of Social Psychology*, *1*, 133-145.

Hoemann, H. W. and Ross, B. M. (1982). Children's concepts of chance and probability. In C. Brainerd (Ed.), *Children's logical and mathematical cognition: Progress in cognitive developmental research* (pp. 93-121). New York: Springer-Verlag.

Hoffrage, U. and Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*(5), 538-540.

Hook, B. E. (1981). Rates of chromosome abnormalities at different maternal ages. *Obstetrics & Gynecology*, *58*(3), 267-400.

Hook, B. E., Cross, P. K., and Schreinemachers, D. M. (1983). Chromosomal abnormality rates at amniocentesis and in live-born Infants. *The Journal of the American Medical Association*, *249*, 2034-2038.

Ischebeck, A., Schocke, M., and Delazer, M. (2009). The processing and representation of fractions within the brain. An fMRI investigation. *Neuroimage, 47,* 403-413.

Jenni, K. E. and Loewenstein, G. (1997). Explaining the Identifiable Victim Effect. *Journal of Risk and Uncertainty*, *14*, 235-257.

Jones-Lee, M. W., Hammerton, M., and Philips, P. R. (1985). The value of safety: Results of a national sample survey. *The Economical Journal*, *95*(377), 49-72.

Jones-Lee, M. W., Loomes, G., and Philips, P. R. (1995). Valuing the prevention of non-fatal road injuries: Contingent valuation vs. standard gambles. *Oxford Economic Papers (New series)*, *47*(4), 676-695.

Kahn, B. E. and Lemhann, D. R. (1991). Modeling choice among assortments. *Journal of Retailing*, *67*, 274-299.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*. *58*(9), 697-720.

Kahneman, D. and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and biases. The psychology of intuitive judgment* (pp. 49-81). Cambridge, UK: Cambridge University Press.

Kahneman, D. and Miller, D. T. (1986). Norm Theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136-153.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237-251.

Kahneman, D. and Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-31.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.

Kahneman, D. and Tversky, A. (1981). On the study of statistical intuitions. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 493–508). New York: Cambridge University Press.

Keller, C., Siegrist, M., and Visschers, V. (2009). Effect of risk ladder format on risk perception in high- and low-numerate individuals. *Risk Analysis*, *29*(9), 1255-1264.

Kirkpatrick, L. A. and 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.

Koehler, J. J. (1997). One in millions, billions, and trillions: Lessons from people v. Collins (1968) for people v. Simpson (1995). *Journal of Legal Education*, *47*, 214-223.

Koehler, J. J. (2001-a). The psychology of numbers in the courtroom: How to make DNA match statistics seem impressive or insufficient. Unpublished manuscript, 1-41.

Koehler, J. J. (2001-b). When are people persuaded by DNA match statistics? *Law and Human Behavior*, *25*(5), 493-513.

Koehler, D. J., Brenner, L., and Griffin, D. (2002). The calibration of expert judgment: Heuristic and biases beyond the laboratory. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and Biases. The psychology of intuitive judgment.* (pp. 686-715). Cambridge, UK: Cambridge University Press.

Koehler, J. J. and Macchi, L. (2004). Thinking about low-probability events. *Psychological Science*, *15*(8), 540-546.

Kogut, T. and Ritov, I. (2005). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behaviour and Human Decision Processes*, *9*, 106-116.

Kunreuther, H., Novemsky, N., and Kahneman, D. (2001). Making low probabilities useful. *The Journal of Risk and Uncertainty*, *23*(2), 103-120.

Lee, J., Grigg, W., and Dion, G. (2007). *The nation's report card: Mathematics 2007* (NCES Report No. 2007-494). Washington, DC: U.S. Department of Education, National Centre for Education Statistics. Retrieved March 26, 2007, from http://nces.ed.gov/nationsreportcard/mathematics.

Levy, A. G. and Baron, J. (2005). How bad is a 10% chance of losing a toe? Judgments of probabilistic conditions by doctors and laypeople. *Memory and Cognition*, *33*(8), 1399-1406.

Lin, C-T. J. and Milon, J. W. (1995). Contingent valuation of health risk reduction for shellfish products. In J. . Caswell (Eds.), *Valuing Food Safety and Nutrition* (pp. 83-114). Westview Press, Boulder: CO.

Lipkus, I. S. (2007). Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Medical Decision Making*, *27*, 696-713.

Lipkus, I. M. and Peters, E. (2009). Understanding the role of numeracy in health: Proposed theoretical framework and practical insights. *Health Education & Behaviour*, *36*(6), 1065-1081.

Lipkus, I. M., Samsa, G., and Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37-44.

Lichtenstein, S. and Slovic, P. (2006). *The construction of preference*. Cambridge, UK: Cambridge University Press.

Lloyd, F. J. and Reyna, V. F. (2001). Genetic Testing and Medical Decision Making. *Archives of Internal Medicine*, *161*, 2406-2408.

Lowenstein, G. F., Weber, E. U., Hsee, C. K., and Welch, N. (2001). Risk as Feelings. *Psychological Bulletin*, *127*(2), 267-286.

Macchi, L. (2000). Partitive formulation of information in probabilistic problems: Beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes*, *82*(2), 217-236.

MacGregor, D. G., Slovic, P., Dreman, D., and Berry, M. (2000). Imagery, affect, and financial judgment. *Journal of Behavioral Finance*, *1*(2), 104-110.

MacInnis, D. J., and Price, L. (1987). The role of imagery in information processing: Review and extensions. *Journal of Consumer Research*, *13*, 473-491.

Markowitz, H. M. (1952). Portfolio-selection. *Journal of Finance*, *7*, 77-91.

Marks, A. D. G., Hine, D. W., Blore, R. L., and Phillips, W. J. (2008). Assessing individual differences in adolescents' preference for rational and experiential cognition. *Personality and Individual Differences*, *44*, 42-52.

Marteau, T. M., Kidd, J., Cook, R., Michie, S., Johnston, M., Slack, J., et al. (1991). Perceived risk not actual risk predicts uptake of amniocentesis. *BJOG: An International Journal of Obstetrics & Gynaecology*, *98*(3), 282-286.

Marteau, T. M., Saidi, G., Goodburn, S., Lawton, J., Michie, S., and Bobrow, M. (2000). Numbers or words? A randomized controlled trial of presenting screen negative results to pregnant women. *Prenatal Diagnosis*, *20*, 714-718.

Mazur, D. J. (1990). Interpretation of graphic data by patients in a general medical clinic. *Journal of General Internal Medicine*, *5*, 402-405.

Mazur, D. J. and Merz, J. F. (1994). How age, outcome severity, and scale influence general medicine clinic patients' interpretations of verbal probability terms. *Journal of General Internal Medicine*, *9*, 268-271.

Michie, S., Lester, K., Pinto, J., and Marteau, T. M. (2005). Communicating Risk Information in Genetic Counselling: An observational Study. *Health, Education, and Behaviour, 32*(5), 589-598.

Milikovski, M. and Elshout, J. J. (1994). Differences in the semantic integration of tabled and non-tabled numbers? *British Journal of Psychology*, *85*, 29-39.

Milikovski, M. and Elshout, J. J. (1995). What makes a number easy to remember? *British Journal of Psychology*, *86*, 537-547.

Miller, D. T., Turnbull, W., and McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology*, *57*, 581-589.

Miron-Shatz, T., Hanoch, Y., Graef, D., and Sagi, M. (2009). Presentation format affects comprehension and risk assessment: The case of prenatal screening. *Journal of Health Communication*, *14*, 439-450.

Motterlini, M. and Crupi, V. (2006). Introduzione. Errori e decisioni in medicina. In V. Crupi, G. F. Gensini, and M. Motterlini (Eds.), *La dimensione cognitiva dell'errore in medicina* (pp. 11-42). Franco Angeli, Milano.

Moyer, R. S. and Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519-1520.

Newell, B. R., Mitchell, C. J., and Hayes, B. K. (2008). Getting scarred and winning lotteries: Effects of exemplar cuing and statistical format on imagining low-probability events. *Journal of Behavioural Decision-Making*, *21*, 317-335.

Nisbett, R. E. and Ross, L. D. (1980). *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Norris, P., Pacini, R., and Epstein, S. (1998). The Rational-Experiential inventory, short form. Unpublished inventory. University of Massachusetts at Amherst.

Nunnally, J. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Ohnishi, M., Fukui, T., Matsui, K., Hira, K., Shinozuka, M., Ezaki, H., et al. (2002). Interpretation of and a preference for probability expressions among Japanese patients and physicians. *Family practice*, *19*, 7-11.

Okan, Y., Garcia-Retamero, R., Cokely, E. T., and Maldonado, A. Individual differences in graph Literacy: Overcoming biases in risk comprehension. Manuscript submitted for publication.

Oxford English Dictionary (2006). *Frequency*. Retrieved June 7, 2010, from Oxford English Dictionary Online, http://www.oed.com.

Pacini, R. and Epstein, S. (1999-a). The interaction of three facets of concrete thinking in a game of chance. *Thinking and Reasoning*, *5*(4), 303-325.

Pacini, R. and Epstein, S. (1999-b). 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*(6), 972-987.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255-287.

Paling, J. (2003). Strategies to help patients understand risks. *British Medical Journal*, *327*, 745-748.

Pascal, B. (1967). In P. Serini (A cura di), Pensieri (pp. 58-59, n° 146). Einaudi, Torino.

Pelham, B., Sumarta, T. T., and Myakovsky, L. (1994). The easy path from many to much: The numerosity heuristic. *Cognitive Psychology*, *26*, 103-133.

Peters, E. and Levin, I. P. (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision Making*, *3*(6), 435-448.

Peters, E., Lipkus, I. M., and Diefenbach, M. A. (2006). The functions of affect in health communications and in the construction of health preferences. *Journal of Communication*, *56*, 140-162.

Peters, E., Slovic, P., Västfjall, D., and Mertz, C. K. (2008). Intuitive numbers guide decisions. *Judgment and Decision Making*, *3*, 619-635.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., and Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*, 407-413.

Piaget, J. and Inhelder, B. (1975). *The origin of the idea of chance in children*. New York: Norton (originally published 1951).

Pighin, S., Savadori, L., Barilli, E., Cremonesi, L., Ferrari, M., and Bonnefon, J.-F. (in press). The ‘1 in X effect’ on the subjective assessment of medical probabilities. *Medical Decision Making.*

Pinto-Prades, J. L., Martinez-Perez, J., Abellán-Perpiñán, J. (2006). The influence of the ratio bias phenomenon on the elicitation of health states utilities. *Judgment and Decision Making*, *1*(2), 118-133.

Price, M., Cameron, R., and Butow, P. (2007). Communicating risk information: The influence of graphical display format on quantitative information perception-accuracy. *Patient Education and Counseling*, *69*, 121-128.

Price, P. C. and 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.

Reyna, V. F. (1991). Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review*, *11*, 317-336.

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. and Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioural Decision Making*, *4*, 249-262.

Reyna, V. F., and Brainerd, C. (1994). The origins of probability judgment: A review of data and theories. In G. Wright and P. Ayton (Eds.), *Subjective Probability* (pp. 239-272). New York, NY: Wiley.

Reyna, V. F. and Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1-75.

Reyna, V. F. and 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. and 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., Nelson, W.L., Han, P. K., and Dieckmann, N. F. (2009). How numeracy Influences risk comprehension and medical decision making. *Psychological Bulletin, 135*(6), 943-973.

Rosch, E. (1975). Cognitive Reference Points. *Cognitive Psychology, 7,* 532-547.

Rosenstock, I. M. (1966). Why people use health services. *The Milbank Memorial Fund Quarterly, 94,* 94-127.

Rothman, R. L., Housam, R., Weiss, H., Davis, D., Gregory, R., Gebretsadik, T., et al. (2006). Patient understanding of food labels: The role of literacy and numeracy. *American Journal of preventive medicine*, *31*, 391-398.

Rottenstreich, Y., and Hsee, C. K. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological Science*, *12*, 185-190.

Rubaltelli, E., Pasini, G., Rumiati, R., Olsen, R. A., and Slovic, P. (2010). The influence of affective reactions on investment decisions. *Journal of Behavioral Finance, 11*, 168-176.

Rudski, J. M. and Volksdorf, J. (2002). Pictorial versus textual information and the ratio-bias effect. *Perceptual and Motor Skills*, *95*(2), 547-554.

Santuososso, A. (1996). *Il consenso informato. Tra giustificazione per il medico e diritto del paziente.* Raffaello Cortina Editore, Milano.

Saraph, J., Benson, P, and Schroeder, R. (1989). An instrument for measuring the critical factors of quality management. *Decision Sciences, 20*(4), 910-829.

Savadori, L. and Rubaltelli, E. (2008). Rischio e decisione. In N. Bonini, F. Del Missier, and R. Rumiati (A cura di), *Psicologia del giudizio e della decisione* (pp. 73-96). Bologna: Il Mulino Editore.

Schelling, T. C. (1968). The life you save may be your own. In S. Chase (Ed.), *Problems in public expenditure analysis*. Washington, DC: The Brookings Institute.

Schwartz, L. M., Woloshin, S., Black, W. C., and Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*(11), 966-972.

Schwartz, L. M., Woloshin, S., and Welch, H. G. (1999). Risk communication in clinical practice: Putting cancer in context. *Journal of the National Cancer Institute Monographs*, *25*, 124-133.

Schwarz, N. and Hippler, H.-J. (1987). What response scales may tell your respondents: Informative functions of response alternatives. In H.-J. Hippler, N. Schwarz, and S. Sudman (Eds.), *Social Information Processing and Survey Methodology* (pp. 163-178). New York: Springer Verlag.

Schwarz, L. M., Strack, F., Muller, G., and Chassein, B. (1988). The range of response alternatives may determine the meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, *6*, 107-117.

Shanteau, J., Weiss, D. J., Thomas, J., and Pounds, J. (2003). How can you tell if someone is an expert? Performance-based assessment of expertise. In S. Schneider and J. Shanteau (Eds.), *Emerging perspectives on judgment and decision-making* (pp. 620-639). Cambridge, UK: Cambridge University Press.

Shiloh, S., Salton, E., and 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. *Monographs of the Society for Research in Child Development*, *46*(2), 1-84.

Simon, H. A. (1955). A behavioural model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118.

Simon, H. A. (1979). Rational decision-making in business organizations. *American Economic Review*, *69*, 493–513.

Sjöberg, L. (1998). Worry and Risk Perception. *Risk Analysis*, *18*(1), 85–93.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3-22.

Slovic, P. (2004). Perception of risk from asteroid impact. *Preliminary Draft of a Paper prepared for the ICSU Workshop Comet/ asteroid Impacts and Human Society.*

Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbers and genocide. *Judgment and Decision Making*, *2*(2), 1-17.

Slovic, P., Finucane, M. L., Peters, E., and MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, *24*(2), 311-322.

Slovic, P., Kraus, N., and Covello, V. T. (1990). What Should We Know About Making Risk Comparisons? *Risk Analysis*, *10*(3), 389-392.

Slovic, P., Layman, M., Kraus, N., Flynn, J., Chalmers, J., and Gesell, G. (1991). Perceived risk, stigma, and potential economic impacts on a high-level nuclear waste repository in Nevada. *Risk Analysis, 11*(4), 683-696.

Slovic, P., Monahan, J., and MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases providing instruction, and employing probability versus frequency formats. *Law and Human Behaviour, 24,* 271-296.

Slovic, P. and Peters, E. (2006). Risk perception and affect. *Current Directions in Psychological Science*, *15*(6), 322-325.

Small, D. and Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty*, *26*(1), 5-16.

Small, D., Loewenstein, G., and Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behaviour and Human Decision Processes*, *102*, 143-153.

Smith, M. U. and Good, R. (1984). Problem solving and classical genetics: Successful versus unsuccessful performance. *Journal of Research in Science Teaching, 21*(9), 895-912.

Stone, E. R., Sieck, W. R., Bull, B. E., Yates, J. F., Parks, S. C., and Rush, C. J. (2003). Foreground: background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behaviour and Human Decision Processes*, *90*(1), 19-36.

Sunstein, C. R. (2003). Terrorism and probability neglect. *The Journal of Risk and Uncertainty*, *26*, 121-136.

Sunstein, C. R. and Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *Working Paper 03-2*, 1-52. AEI-Brookings Joint Center for Regulatory Studies.

Szalay, L. B. and Deese, J. (1978). *Subjective meaning and culture: An assessment through word associations*. Lawrence Erlbaum Associates: Hillsdale, N. J. and New York.

Thornton, J. G., Hewison, J., Lilford, R. J., and Vail, A. (1995). A randomised trial of three methods of giving information about prenatal testing. *British Medical Journal*, *311*, 1127.

Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., and Lucca, N. (1988). Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology*, *54*, 323-338.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.

Tversky, A. and Kahneman, D. (1981). The framing of decision and the psychology of choice. *Science*, *211*, 453-458.

van den Berg, M., Timmermans, D. R. M., Knol, D. L., van Eijk, J. T. M., de Smit, D. J., van Vugt, J. M. G., et al. (2008). Understanding pregnant women's decision making concerning prenatal screening. *Health Psychology*, *27*(4), 430-437.

Van Steenkiste, B., Van der Weijden, T., Timmermans, D., Vaes, J., Stoffers, J., and Grol, R. (2004). Patients' ideas, fears and expectations of their coronary risk: Barriers for primary prevention. *Patient Education and Counseling, 55*, 301–307.

van Vliet, A. A. M., Grimes, D. A., Popkin, B., and Smith, U. (2001). Lay persons' understanding of the risk of Down's Syndrome in genetic counseling. *British Journal of Obstetrics and Gynecology*, *108*, 649-650.

Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., and de Vries, N. N. K. (2009). Probability information in risk communication: A review of the research literature. *Risk Analysis*, *29*(2), 267-287.

Von Neumann, J. and Morgenstern, O. (1947). *Theory of games and economic behavior*. NJ: Princeton University Press.

Wallsten, T., Budescu, D., and Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, *39*, 176-190.

Waters, E. A., Weinstein, N. D., Colditz, G. A., and Emmons, K. (2006). Formats for improving risk communication in medical tradeoff decisions. *Journal of Health Communication*, *11*, 167-182.

Weber, E. H. (1834). *Depulsu, resorptione, auditu et tactu*. Leipzig: Koehler.

Weinstein, N. D. (1999). What does it mean to understand a risk? Evaluating risk comprehension. *Journal of National Cancer Institute Monographs*, *25*, 15-20.

Weinstein, N. D., Sandman, P. M., and Hallman, W. H. (1994). Testing a visual display to explain small probabilities. *Risk Analysis*, *14*, 895-897.

Wertz, D. C., Sorenson, J. R., and Heeren, T. C. (1986). Client's interpretation of risk provided in genetic counseling. *American Journal of Human Genetics*, *39*, 253-264.

Witteman, C., van den Bercken, J., Claes, L., and Godoy, A. (2009). Assessing rational and intuitive thinking styles. *European Journal of Psychological Assessment*, *25*(1), 39-47.

Wolfe, C. R. (1995). Information seeking on Bayesian conditional probability problems: A fuzzy-trace theory account. *Journal of Behavioral Decision Making*, *8*, 85-108.

Woloshin, S., Schwartz, L. M., Byram, S., Fischhoff, B., and Welch, G. (2000). A new scale for assessing perceptions of chance. A validation study. *Medical Decision Making, 20*(3), 298-307.

Yamagishi, K. (1994-a). Consistencies and biases in risk perception: I. Anchoring process and response-range effect. *Perceptual and Motor Skills, 79*, 651-656.

Yamagishi, K. (1994-b). Consistencies and biases in risk perception: II. Individual subjects performance. *Perceptual and Motor Skills, 79*, 659-663.

Yamagishi, K. (1997-a). When a 12.86 % mortality is more dangerous than 24.14%: Implications for risk communication. *Applied Cognitive Psychology, 11*, 496-506.

Yamagishi, K. (1997-b). Upward versus downward anchoring in frequency judgments of social facts. *Japanese Psychological Research*, *39*(2), 124-129.

Yamaguchi, S. (1998). Biased risk perceptions among Japanese: Illusion of interdependence among risk companions. *Asian Journal of Social Psychology*, *1*, 117-131.

Yamaguchi, S., Okumura, T., Chua, H. F., Morio, H., and Yates, J. F. (2008). Levels of control across cultures: Conceptual and empirical analysis. In F. J. R. van de Vijver, D. A. van Hemert, and Y. H. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp.123-144). NewYork: Erlbaum.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151-175.

Zikmund-Fisher, B. J., Smith, D. M., Ubel, P. A., and Fagerlin, A. (2007). Validation of the Subjective Numeracy Scale: Effects of low numeracy on comprehension of risk communications and utility elicitations. *Medical Decision Making*, *27*(5), 663-671.

Zikmund-Fisher, B. J., Ubel, B. A., Smith, D. M., Derry, H. A., McClure, J. B., Stark, A., et al. (2008). Communicating side effect risks in a tamoxifen prophylaxis decision aid: The debiasing influence of pictographs. *Patient Education and Counseling*, *73*(2), 209-214.

APPENDIX

Experimental material employed in the studies

SECTION 1.

Study 1.1

Legga lo scenario e risponda alle domande:

Immagini che un suo amico abbia comprato una vacanza per il Kenya. Lei ha appena letto che in Kenya il rischio di contrarre la malaria è di 1 su 200 [5 su 1000][91].

1) Secondo lei, se il suo amico va in Kenya, la probabilità che contragga la malaria è:

☐-----------☐-----------☐----------☐----------☐-----------☐-----------☐
assolutamente trascurabile | molto bassa | bassa | media | alta | molto alta | prossima alla certezza

2) Secondo lei la malaria è una malattia:

☐----------------☐-----------------☐---------------☐---------------☐
Gravissima | molto grave | abbastanza grave | poco grave | per nulla grave

3) Se lei andasse in Kenya, quanto la preoccuperebbe la possibilità di contrarre la malaria?

☐-----------☐-----------☐----------☐----------☐-----------☐-----------☐
per niente | molto poco | poco | mediamente | abbastanza | molto | moltissimo

Legga e risponda alle seguenti domande:

5) Immagini di lanciare 1000 volte un dado con sei facce non truccato. Facendo 1000 lanci, quante volte ritieni che il dado potrebbe dare come esito un numero pari (2, 4, 6)? |________| *(500 su 1000)*

6) Nella BIG BUCKS LOTTERY, le probabilità di vincere un premio di $10 sono pari all'1%. Secondo te quante persone potrebbero vincere $10 se 1000 persone comprano un singolo biglietto della lotteria a testa? |________| *(10 persone su 1000)*

7) Nell'ACME PUBLISHING SWEEPSTAKES la probabilità di vincere una macchina è pari a 1 su 1000. Quale percentuale di biglietti dell'ACME PUBLISHING SWEEPSTAKES vince una macchina? |_________| % *(0.1)*

8) Quale dei seguenti numeri rappresenta la probabilità più alta?
☐ 1 su 100
☐ 1 su 1000
☐ 1 su 10
(1 su 10)

[91] In all studies, the expression/s employed in experimental condition/s alternative to that written in black font appears/appear in light blue font and enclosed in rectangular brackets; also, correct answers (when existent) to questions are in italics and enclosed in round brackets.

9) Quale dei seguenti numeri rappresenta la probabilità più alta?
☐ 1%
☐ 10%
☐ 5%
(10%)

10) Il rischio della persona A di contrarre una malattia nei prossimi 10 anni è pari all'1%, e quello della persona B è il doppio di quello della persona A; qual è il rischio della persona B di contrarre una malattia? |_________| % *(2)*

11) Il rischio della persona A di contrarre una malattia nei prossimi 10 anni è pari a 1 su 100, e quello della persona B è il doppio di quello della persona A; qual è il rischio della persona B di contrarre una malattia? |_________| su |__________| *(2 su 100)*

12) Se la probabilità di contrarre una malattia è pari al 10%, quante persone in un campione di 100 ci si deve aspettare che contrarranno la malattia? |__________| *(10)*

13) Se la probabilità di contrarre una malattia è pari al 10%, quante persone in un campione di 1000 ci si deve aspettare che contrarranno la malattia? |__________| *(100)*

14) La probabilità di contrarre una malattia è pari a 20 su 100; ciò dovrebbe essere uguale ad attendersi una probabilità pari al |__________| % di contrarre la malattia. *(20)*

15) La probabilità di contrarre un'infezione virale è .0005. Su 10 mila persone, quante di esse ci si deve aspettare che contrarranno l'infezione? |__________| *(5 persone)*

Study 1.2

Legga il seguente scenario, provando ad immedesimarsi in uno dei due protagonisti (cioè Anna, se Lei è una donna, oppure Luca, se Lei è un uomo):

Anna e Luca sono una coppia che aspetta un bambino.
I due si recano insieme alla prima visita medica di controllo. Il ginecologo conferma lo stato di gravidanza di Anna e li informa che, considerata l'età di Anna, la loro probabilità di avere un figlio affetto dalla sindrome di Down è approssimativamente di 1 su 200 [5 su 1000].

Ora risponda alla seguente domanda:

Se fosse nei panni di Anna (oppure di Luca, se Lei è un uomo) riterrebbe la probabilità di 1 su 200 [5 su 1000] di avere un figlio affetto dalla sindrome di Down

❑	❑	❑	❑	❑	❑	❑
Estrema-mente bassa	Molto bassa	Bassa	Moderata	Alta	Molto alta	Estrema-mente alta

(on the following page)

Secondo Lei, la sindrome di Down è una malattia genetica

❑	❑	❑	❑	❑	❑
Per nulla grave	Un po' grave	Abbastanza grave	Molto grave	Gravissima	Non so

(on the following page)

Quale tra queste affermazioni La rappresenta meglio? (indichi una sola risposta)
❑ Non ho mai sentito parlare della sindrome di Down
❑ Ho sentito parlare/letto della sindrome di Down, ma non ho mai visto direttamente persone affette da tale malattia
❑ Ho visto direttamente persone affette da sindrome di Down, ma non ho mai interagito con loro
❑ Ho interagito con persone affette da sindrome di Down, ma solo saltuariamente
❑ Ho interagito spesso con persone affette da sindrome di Down

SECTION 2.

Study 2.1

Legga il seguente scenario, provando ad immedesimarsi in uno dei due protagonisti (cioè Anna, se Lei è una donna, oppure Luca, se Lei è un uomo):

Anna e Luca sono una coppia che aspetta un bambino.
I due si recano insieme alla prima visita medica di controllo. Il ginecologo conferma lo stato di gravidanza di Anna e li informa che, considerata l'età di Anna, la loro probabilità di avere un figlio affetto dalla sindrome di Down è di 1 su 200 [5 su 1000, 50 su 10000].

Ora risponda alla seguente domanda:

Se fosse nei panni di Anna (oppure di Luca, se Lei è un uomo) riterrebbe la probabilità di 1 su 200 [5 su 1000, 50 su 10,000] di avere un figlio affetto dalla sindrome di Down:

❑	❑	❑	❑	❑	❑	❑
Estrema-mente bassa	Molto bassa	Bassa	Moderata	Alta	Molto alta	Estrema-mente alta

(on the following page)

Per ognuna delle affermazioni sulla sinistra, indichi, per cortesia, in che misura La descrive segnando una crocetta nella casella corrispondente.

Cerco di evitare le situazioni in cui è necessario riflettere a fondo
Non sono molto bravo a risolvere problemi complicati
Mi piace basarmi sulle mie impressioni intuitive
Mi piacciono le sfide intellettuali
Non possiedo un gran intuito
Nel risolvere i problemi della mia vita di solito mi trovo bene ad usare l'istinto
Non sono molto bravo a risolvere problemi che richiedono un'attenta analisi logica
Credo che sia importante fidarsi delle proprie intuizioni
Usare l'intuito può essere molto utile per risolvere i problemi
Non mi piace dovere riflettere molto
Spesso mi baso sulle mie intuizioni nel decidere il corso di un'azione
Mi piace risolvere problemi che richiedono di pensare molto
Mi fido delle mie prime impressioni sulle persone
Pensare non è il mio ideale di attività piacevole
Non sono una persona che riflette in modo molto razionale
Quando si tratta di fidarsi delle persone, posso di solito basarmi sulle mie sensazioni istintive
Se dovessi basarmi sui miei sentimenti istintivi, commetterei spesso degli errori
Ragionare con attenzione sulle cose non è uno dei miei punti forti
Preferisco i problemi complessi a quelli semplici
Non mi piacciono le situazioni nelle quali devo basarmi sull'intuito
Non ragiono bene quando sono sotto pressione
Riflettere a fondo e a lungo su un problema mi dà poca soddisfazione
Penso che ci siano delle situazioni nelle quali bisognerebbe fidarsi del proprio istinto
Sono molto più bravo a risolvere un problema in modo logico rispetto alla maggior parte delle persone
Penso che sia sciocco prendere decisioni importanti sulla base delle impressioni
Ho una mente logica
Non penso sia una buona idea affidarsi alle proprie intuizioni per prendere decisioni importanti
Mi piace pensare in termini astratti
Generalmente non faccio affidamento sulle mie sensazioni per aiutarmi a prendere una decisione
Difficilmente mi sbaglio quando ascolto i miei sentimenti istintivi più profondi per trovare una risposta
Non ho problemi a ponderare le cose con attenzione
Non vorrei dipendere da nessuno che si descriva come una persona intuitiva
Nel risolvere i problemi della mia vita di solito mi trovo bene ad usare la logica
Per me è sufficiente conoscere la risposta ad un problema senza dovere capire il ragionamento che ci sta dietro
Probabilmente i miei giudizi istantanei non sono buoni quanto quelli della maggior parte della gente
Tendo ad usare il mio cuore come guida per le mie azioni
Di solito le mie decisioni hanno ragioni chiare e comprensibili
Di solito riesco a percepire se una persona ha ragione o ha torto, anche se non so spiegare come lo avverto
Sarebbe interessante per me imparare nuovi modi di pensare
Sospetto che le mie intuizioni siano inesatte tanto spesso quanto esatte

The 40-items of the REI scale used in the study translated into Italian and readapted from those of Pacini and Epstein (1999-b)- here in order 1.

The scale available for response to each item of the 40-item REI scale above:

❑ Assolutamente falso per me	❑ Piuttosto falso per me	❑ Né vero né falso per me	❑ Piuttosto vero per me	❑ Assolutamente vero per me

(on the following page) The 8 items of the SNS scale used in the study, translated into Italian and readapted from those of Fagerlin et al. (2007):

Quanto si ritiene brava/o a fare operazioni con le frazioni?

❑ Per niente	❑ Un po’	❑ Abbastanza	❑ Molto	❑ Estremamente

Quanto si ritiene brava/o a fare operazioni con le percentuali?

❑ Per niente	❑ Un po’	❑ Abbastanza	❑ Molto	❑ Estremamente

Quanto si ritiene brava/o a calcolare una mancia pari al 15% del conto?

❑ Per niente	❑ Un po’	❑ Abbastanza	❑ Molto	❑ Estremamente

Quanto si ritiene brava/o a calcolare il prezzo finale di una maglia su cui va applicato lo sconto del 25%?

❑ Per niente	❑ Un po’	❑ Abbastanza	❑ Molto	❑ Estremamente

Quando legge il giornale, quanto considera utili le tabelle con i numeri ed i grafici che fanno parte di un articolo?

❑ Per niente	❑ Un po’	❑ Abbastanza	❑ Molto	❑ Estremamente

Quando una persona Le vuole spiegare che qualcosa ha una certa possibilità di accadere, preferisce che quella persona usi le parole (“succede raramente”) oppure i numeri (“esiste l’1% di possibilità”)?

❑ Preferisco sempre le parole	❑ Preferisco le parole, nella maggior parte dei casi	❑ Metà delle volte preferisco le parole e metà i numeri	❑ Preferisco i numeri, nella maggior parte dei casi	❑ Preferisco sempre i numeri

Quando ascolta le previsioni del tempo, preferisce quelle che usano le parole (“oggi c’è una bassa probabilità che piova”) oppure quelle che usano le percentuali (“oggi c’è un 20% di rischio che piova”)?

❑ Preferisco sempre le parole	❑ Preferisco le parole, nella maggior parte dei casi	❑ Metà delle volte preferisco le parole e metà i numeri	❑ Preferisco i numeri, nella maggior parte dei casi	❑ Preferisco sempre i numeri

In generale, quanto considera utile l’informazione numerica?

❑ Per niente	❑ Un po’	❑ Abbastanza	❑ Molto	❑ Estremamente

(on the following page) Among socio-demographic variables, the two questions used to create the measure of familiarity with the outcome:

Ha figli? ❑ Sì ❑ No

Se sì, Le è mai stata comunicata da uno specialista medico la probabilità che il nascituro fosse affetto da sindrome di Down?

❑ Sì ❑ No

Study 2.2

Nella tabella presentata qui di seguito, è indicato il rischio di avere un figlio affetto dalla sindrome di Down in base all'età materna. Le chiediamo cortesemente di cerchiare il rischio relativo alla sua età.

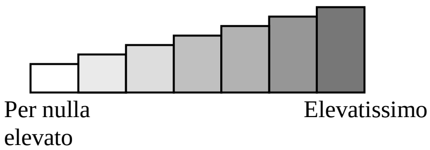
Table 1
("1 in X" format)

Età materna	Rischio
18-20	1 ogni 1540 nascite
21-23	1 ogni 1480 nascite
24-26	1 ogni 1350 nascite
27-29	1 ogni 1120 nascite
30-32	1 ogni 795 nascite
33-35	1 ogni 475 nascite
36-38	1 ogni 240 nascite
39-41	1 ogni 110 nascite
42-44	1 ogni 49 nascite
45-47	1 ogni 21 nascite
48-50	1 ogni 8 nascite
51 e oltre	> 1 ogni 6 nascite

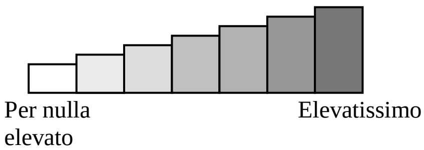
Table 2
("Y in 1000" format)

Età materna	Rischio
18-20	0.6 ogni 1000 nascite
21-23	0.7 ogni 1000 nascite
24-26	0.8 ogni 1000 nascite
27-29	0.9 ogni 1000 nascite
30-32	1.3 ogni 1000 nascite
33-35	2.1 ogni 1000 nascite
36-38	4.2 ogni 1000 nascite
39-41	9.1 ogni 1000 nascite
42-44	20.4 ogni 1000 nascite
45-47	47.6 ogni 1000 nascite
48-50	125 ogni 1000 nascite
51 e oltre	> 166.7 ogni 1000 nascite

D1. Quanto giudica elevato il rischio da lei cerchiato di avere un figlio affetto dalla sindrome di Down? Segni una x all'interno della casella corrispondente alla sua valutazione.



D2. Quanto la preoccupa tale rischio? Segni una x all'interno della casella corrispondente alla sua valutazione.



D3. Il rischio da lei cerchiato è:
- ☐ superiore a 1 su 250
- ☐ inferiore a 1 su 250

D4. Rispetto al rischio di una donna di 23 anni, il rischio da lei cerchiato è:
☐ superiore
☐ inferiore
☐ uguale

D5. Il rischio da lei cerchiato significa che la probabilità di avere un figlio affetto dalla sindrome di Down è:
☐ maggiore del 50%
☐ minore del 50%

D6. Immagini che il rischio di un'altra donna in gravidanza sia il doppio rispetto a quello da lei cerchiato. Quanto sarebbe questo rischio? ________ su _______.

D7. Il rischio da lei cerchiato equivale ad un rischio di _______ su 1000.

D8. Qual è la probabilità di avere un figlio NON affetto dalla sindrome di Down considerata la sua età? __________ su _____________.

Study 2.3

Legga il seguente scenario e risponda alle domande.

Anna è una donna di 48 anni, sposata da quando ne aveva 29. Lei e il marito da tempo desideravano un figlio ma problemi di fertilità dovuti a cause non evidenti hanno impossibilitato l'inizio di una gravidanza. Nel momento in cui Anna e Luca si erano rassegnati alla situazione, Anna scopre di essere incinta. Si reca dal ginecologo per una visita di controllo. Il medico le conferma lo stato di gravidanza e la informa che, data la sua età, il rischio di avere un figlio affetto dalla sindrome di Down è di circa 1 su 12 [10 su 120].

1. Se lei fosse nei panni di Anna, come valuterebbe la probabilità di 1 su 12 [10 su 120] di avere un figlio affetto dalla sindrome di Down?

❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑

Estremamente bassa — Media — Estremamente alta

2. Per sapere con certezza se il figlio di Anna ha un'anomalia cromosomica è necessario che Anna svolga un test gratuito, l'amniocentesi, che però comporta un rischio aggiuntivo di aborto di circa 1 caso su 100. Se lei fosse nei panni di Anna, farebbe l'amniocentesi?

❑ sì ❑ no

(on the following page). **Provi ora a rispondere alle seguenti domande il più velocemente possibile:**

3. A quale delle seguenti percentuali corrisponde la probabilità di 1 su 12 [10 su 120] ? *(circa 10%)*
   - ❑ Circa 1%
   - ❑ Circa 10%
   - ❑ Circa 100%

4. Quale dei seguenti numeri rappresenta la probabilità più alta? *(10%)*
   - ❑ 0.1%
   - ❑ 1%
   - ❑ 10%

5. La probabilità di 1 su 12 è maggiore o minore della probabilità del 50%? *(minore)*
   - ❑ maggiore
   - ❑ minore

6. La probabilità di 1 su 12 [10 su 120] corrisponde circa a: *(10 su 100/ 1 su 10)*
   - ❑ 10 su 100
   - ❑ 10 su 1000
   - ❑ 1 su 100
   - [❑ 1 su 10
   - ❑ 10 su 1000
   - ❑ 1 su 100]

7. Se la donna A ha la probabilità di 1 su 12 [10 su 120] di avere un figlio affetto dalla sindrome di Down e la donna B ha una probabilità doppia rispetto alla donna A, qual è la probabilità della donna B? *(2 su 12/ 20 su 120)*
   - ❑ 2 su 6
   - ❑ 1 su 24
   - ❑ 2 su 12
   - [❑ 2 su 60
   - ❑ 10 su 240
   - ❑ 20 su 120]

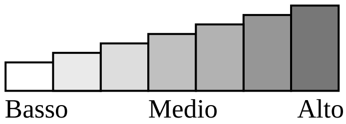
Study 2.4

Immagini il seguente scenario:

Anna e Luigi stanno aspettando un bambino. Anna ha 45 anni. Si recano dal ginecologo per la prima visita ed il medico li informa che data l'età di Anna, il rischio per la coppia di avere un figlio affetto dalla sindrome di Down è di circa 3 su 48 [10 su 160].

Si metta nei panni di Anna (se lei è una donna) o nei panni di Luigi (se lei è un uomo).

D1. Quanto giudica elevato il rischio di 3 su 48 [10 su 160] di avere un figlio affetto dalla sindrome di Down? Segni una x all'interno della casella corrispondente alla sua valutazione.



D2. Per sapere con certezza se il figlio di Anna e Luigi ha la sindrome di Down è necessario che Anna svolga un test, l'amniocentesi, che però comporta un rischio aggiuntivo di aborto di circa 1 caso su 100. Se lei fosse nei panni di Anna e Luigi, farebbe l'amniocentesi?

☐ sì, certamente
☐ tendenzialmente sì
☐ tendenzialmente no
☐ assolutamente no
☐ non so

Study 2.5

Alcune statistiche sono semplici e altre più complesse. Quelle più semplici tendiamo a non trasformarle. Per esempio "1 su 10" normalmente non viene trasformato, mentre "3 su 30" tende ad essere trasformato in "1 su 10".

(here in order 1)

0.4 su 10
è una statistica che Lei tende a trasformare in un altro formato, oppure che tende a considerare così come Le viene presentata?

❑ Tendenzialmente la trasformo	❑ Tendenzialmente la considero così com'è

1 su 25
è una statistica che Lei tende a trasformare in un altro formato, oppure che tende a considerare così come Le viene presentata?

❑ Tendenzialmente la trasformo	❑ Tendenzialmente la considero così com'è

2 su 50
è una statistica che Lei tende a trasformare in un altro formato, oppure che tende a considerare così come Le viene presentata?

❑ Tendenzialmente la trasformo	❑ Tendenzialmente la considero così com'è

10 su 250
è una statistica che Lei tende a trasformare in un altro formato, oppure che tende a considerare così come Le viene presentata?

❑ Tendenzialmente la trasformo	❑ Tendenzialmente la considero così com'è

Study 2.6

Legga il seguente scenario, provando ad immedesimarsi in uno dei due protagonisti (cioè Anna, se Lei è una donna, oppure Luca, se Lei è un uomo):

Anna e Luca sono una coppia che aspetta un bambino.
I due si recano insieme alla prima visita medica di controllo. Il ginecologo conferma lo stato di gravidanza di Anna e li informa che, considerata l'età di Anna, la loro probabilità di avere un figlio affetto dalla sindrome di Down è di 0.4 su 10 [1 su 25, 2 su 50, 10 su 250].

Ora risponda alla seguente domanda:

Se fosse nei panni di Anna (oppure di Luca, se Lei è un uomo) riterrebbe la probabilità di 0.4 su 10 [1 su 25, 2 su 50, 10 su 250] di avere un figlio affetto dalla sindrome di Down:

❑	❑	❑	❑	❑	❑	❑
Estrema-mente bassa	Molto bassa	Bassa	Moderata	Alta	Molto alta	Estrema-mente alta

(on the following page)

Alcune statistiche sono semplici e altre più complesse. Quelle più semplici tendiamo a non trasformarle. Per esempio "1 su 10" normalmente non viene trasformato, mentre "3 su 30" tende ad essere trasformato in "1 su 10".

Ripensi alla statistica "0.4 su 10" ["1 su 25", "2 su 50", "10 su 250"] **che ha incontrato nello scenario.**
Nel rispondere alla domanda, l'ha trasformata in un altro formato oppure l'ha considerata così come Le era stata presentata?

❑	❑
L'ho trasformata in un altro formato	L'ho considerata così com'era

SECTION 3.

Study 3.1

Legga il seguente scenario e risponda alle domande.

Anna è una donna di 48 anni, sposata da quando ne aveva 29. Lei e il marito da tempo desideravano un figlio ma problemi di fertilità dovuti a cause non evidenti hanno impossibilitato l'inizio di una gravidanza. Nel momento in cui Anna e Luca si erano rassegnati alla situazione, Anna scopre di essere incinta. Si reca dal ginecologo per una visita di controllo. Il medico le conferma lo stato di gravidanza e la informa che, data la sua età, il rischio di avere un figlio affetto dalla Sindrome di Down è di circa 1 su 12 [10 su 120].

Provi a mettersi nei panni di Anna.

Quanto è negativo il sentimento che prova pensando alla probabilità di 1 su 12 [10 su 120] di avere un figlio affetto dalla sindrome di Down?

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Per nulla negativo					Neutro					Estremamente negativo

Quanto è positivo il sentimento che prova pensando alla probabilità di 1 su 12 [10 su 120] di avere un figlio affetto dalla sindrome di Down?

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Per nulla positivo					Neutro					Estremamente positivo

Following Study (with a single bipolar scale)

Legga il seguente scenario e risponda alla domanda.

Anna è una donna di 45 anni che scopre di essere incinta. Si reca dal ginecologo per una visita di controllo. Il medico le conferma lo stato di gravidanza e la informa che, data la sua età, il rischio di avere un figlio affetto dalla Sindrome di Down è di circa 1 su 10 [10 su 100].

Che tipo di emozione prova pensando alla probabilità di 1 su 10 [10 su 100] di avere un figlio affetto dalla sindrome di Down?

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Estremam. negativa					Neutra					Estremam. positiva

Study 3.2

Legga il seguente scenario, provando ad immedesimarsi in uno dei due protagonisti (cioè Anna, se Lei è una donna, oppure Luca, se Lei è un uomo):

Anna e Luca sono una coppia che aspetta un bambino.
I due si recano insieme alla prima visita medica di controllo. Il ginecologo conferma lo stato di gravidanza di Anna e li informa che, considerata l'età di Anna, la loro probabilità di avere un figlio affetto dalla sindrome di Down è di 1 su 200 [5 su 1000].

(on the following page)

Qual è la prima immagine* che le viene in mente quando pensa alla probabilità di 1 su 200 [5 su 1000] **di avere un figlio affetto dalla sindrome di Down?**

Immagine ____ -2 -1 0 +1 +2

*** Per "immagine" intendiamo i pensieri o le associazioni mentali che le vengono in mente. Per esempio, se qualcuno menziona la parola "calcio", Le potrebbero venire in mente i Mondiali, Francesco Totti, il *derby*, o anche un prato verde. A noi interessa la prima immagine che Le viene in mente.**

(on the following page)

Come avrà notato, sulla destra della riga su cui ha scritto i Suoi pensieri c'era una scala da -2 a +2. Per favore, ora torni indietro e, usando la scala, valuti le emozioni ed i sentimenti che prova nei confronti dell'immagine che ha scritto. Fornisca la Sua valutazione facendo una "X" su uno dei punti della scala, ricordando che ciascun valore corrisponde alle etichette qui sotto:

Assolutamente negativo	Abbastanza negativo	Neutro	Abbastanza positivo	Assolutamente positivo
-2	**-1**	**0**	**+1**	**+2**

SECTION 4.

Study 4.1

Immagini il seguente scenario: lei sta valutando la possibilità di fare un viaggio in Kenya e la informano che la probabilità di contrarre l'epatite A per chi viaggia in Kenya è di 1 su 200 [5 su 1000].

Secondo lei, questa probabilità è:

-

❑	❑	❑	❑	❑	❑	❑
estremamente bassa	molto bassa	bassa	media	alta	molto alta	estremamente alta

Study 4.2

Legga lo scenario e risponda alla domanda:

Immagini che un suo amico abbia comprato una vacanza per il Kenya. Lei ha appena letto che in Kenya il rischio di contrarre la malaria è di 1 su 200 [5 su 1000].
Per comprendere meglio il rischio, provi ad immaginare di avere un contenitore con 200 [1000] palline di cui 199 [995] bianche e 1 rossa [5 rosse]. Immagini di infilare una mano nel contenitore e di estrarre una pallina. La probabilità di contrarre la malaria equivale alla probabilità di estrarre la [una] pallina rossa.

Secondo lei, se il suo amico va in Kenya, la probabilità che contragga la malaria è:

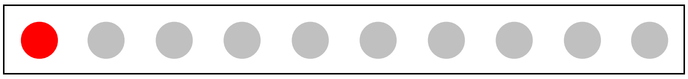
☐	☐	☐	☐	☐	☐	☐
assolutamente trascurabile	molto bassa	bassa	media	alta	molto alta	prossima alla certezza

Study 4.3

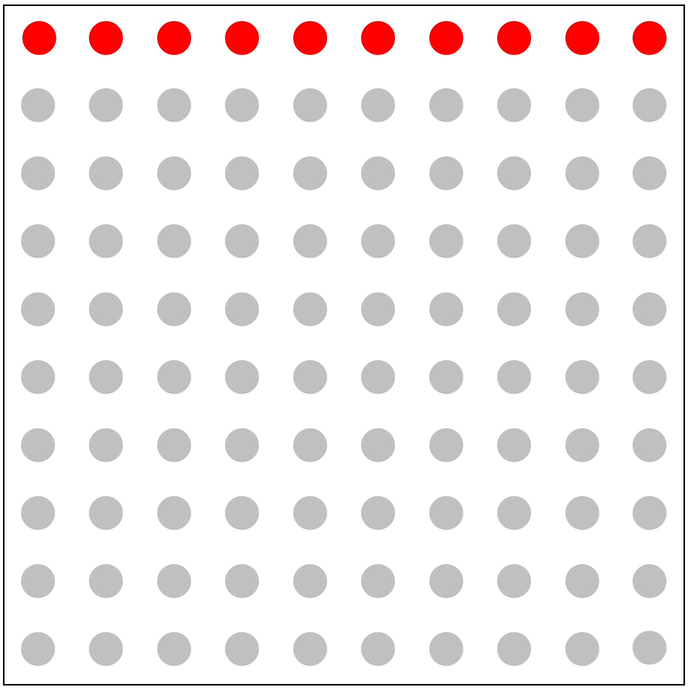
Legga il seguente scenario e risponda alle domande.

Anna è una donna di 45 anni che scopre di essere incinta. Si reca dal ginecologo per una visita di controllo. Il medico le conferma lo stato di gravidanza e la informa che, data la sua età, il rischio di avere un figlio affetto dalla Sindrome di Down è di circa 1 su 10 [10 su 100].

(In the graphical condition, the following textual part was added)
Volendo agevolare la comprensione di tale valore di rischio, il medico fornisce ad Anna la seguente figura: *(either a or b, according to the experimental condition)*



a. The pictogram used in the 1 in 10 condition



b. The pictogram used in the 10 in 100 condition

Provi a mettersi nei panni di Anna e a rispondere alle seguenti domande:

D1. Come valuterebbe la probabilità di 1 su 10 [10 su 100] di avere un figlio affetto dalla sindrome di Down?

❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑

Estremamente bassa — Media mente — Estremamente alta

D2. Quanto la preoccuperebbe la probabilità di 1 su 10 [10 su 100] di avere un figlio affetto dalla sindrome di Down?

❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑

Per nulla — Mediamente — Moltissimo

D3. Che tipo di emozione prova pensando alla probabilità di 1 su 10 [10 su 100] di avere un figlio affetto dalla sindrome di Down?

❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑

Estremamente negativa — Neutra — Estremamente positiva