# Fostering Probabilistic Reasoning Away from Fallacies: Natural Information Formats and Interaction between School Levels

Francisco Vargas [1,2*], Tommaso Benincasa [1,3], Giuseppe Cian [3], Laura Martignon [2]

[1] Universidad el Bosque, COLOMBIA
[2] Ludwigsburg University of Education, GERMANY
[3] Leonardo da Vinci School, COLOMBIA

* CORRESPONDENCE: ✉ fvargasmancera@gmail.com

**ABSTRACT**

The article reports an empirical study on the introduction of elementary probabilistic concepts in school, focusing on tasks related to the psychological tradition of heuristics and biases. The concepts involved were studied using an extensional natural frequencies approach. We describe the school intervention conducted in an interaction across different school levels (5th and 9th grades) with the aim of promoting motivation and cooperation thereby strengthening learning. The different tests were assessed both qualitatively (based on argumentation analyses) and quantitatively. The results provide further evidence on the diversity of obstacles tied to probabilistic notions. More importantly, they exhibit an overall improvement in performance of students at both levels. This work confirms the efficacy of natural frequencies in eliciting the intended interpretation of probabilistic tasks and suggests that an appropriate interaction between different scholastic levels can be implemented as a fruitful learning arrangement.

**Keywords:** probabilistic thinking, natural frequencies, conjunction fallacy, cooperative learning, learning through teaching

## INTRODUCTION

According to the editors of the compendium Probabilistic Thinking: Presenting Plural Perspectives (Chernoff & Sriraman, 2014), research in mathematics education concerning probabilistic thought during the "Contemporary Period" (i.e., during the 1990s and 2000s) has been described as "investigating the teaching and learning of probability in classrooms and schools, which is due, in large part, to probability becoming a mainstream strand of worldwide curricula." Stochastic literacy is in fact of paramount importance both for informed citizenship (decision theory, data management, risk evaluation, etc.) and a required tool in a variety of disciplines.

Even so, educational research is still engaged with some open issues such as, for example, the discus-sion on the different interpretations of probability, the sources of obstacles and biases, and the possible strategies to be adopted in the teaching-learning process. These are main topics of interest in the "Assimilation Period" (Chernoff & Sriraman, 2014) which corresponds to the current phase. The present paper aims at contributing to these issues, focusing more specifically on two aspects:

(1) the analyses of obstacles that are responsible for deviations from standard norms, and

(2) the validation of didactic artifacts that can be implemented based on these analyses.

We base our work on the results of cognitive psychology on probabilistic thought. These are inherent to modern frameworks of rationality.

A main reference in this field has been the heuristics and biases program (Tversky & Kahneman, 1974; Kahneman, 2011). The possible interpretations/explanations of its experimental results have been the subject of heated debates. Concerning the "conjunction fallacy", for instance, there have been very different or even opposite accounts: from those based on our alleged incapacity to deal with it (the original position of Tversky and Kahneman), to the complete denial of its existence (e.g. in Hintikka, 2004). Another position identifies difficulties in the ambiguity of the terms involved, (Fiedler, 1988; Hertwig, 1995; Hertwig, et al., 2008; Hertwig & Gigerenzer, 1999) and propose information formats that elicit the intended interpretations.

Let us now briefly describe the position adopted by mathematics educators regarding reasoning biases. This position influences the methodology used when fostering students' probabilistic competencies. Two basic questions are: Do subjects' patterns of response change across time? Can cognitive biases be overcome by means of adequate learning arrangements or representation formats?

Here two lines of work from the developmental and the educational realms are fundamental:

In the piagetian perspective (Inhelder & Piaget, 1958; Piaget & Inhelder, 1975) when reaching the formal operational stage children have the maturity to deal with core logical and probabilistic notions. This perspective was challenged, among others, by Fischbein's research, conducted in the framework of his study on intuition. This research showed on the one hand, the presence of some skills and intuitions on probability even at an earlier stage, but also, on the other hand, reflected the presence of cognitive biases at different school levels, revealing that the incidence of some "misconceptions" is stable, and may even increase across different scholastic levels (Fischbein et al., 1997, see also Engel & Sedlmeier, 2005).

Some of Fischbein's findings may well be explained by those psychological analyses pointing at the ambiguity of the terms in the tasks themselves (here again, see Fiedler, 1988; Hertwig, 1995; Hertwig, et al., 2008; Hertwig & Gigerenzer, 1999). Besides shedding light on the nature of the biases, they provide instruments and methodologies for improving students' probabilistic performance. These instruments consist mainly of representation formats, i.e. specific semiotic registers in Duval's (2017) sense, namely natural frequencies and icon arrays (Galesic et al., 2009).

Along these lines our work investigates the components of adequate interventions for fostering probabilistic insight.

The value of the interventions we propose can be estimated considering some recent university-level studies which show that performance in problems inducing cognitive biases can be poor in spite of formal knowledge acquisition in more traditional statistic courses (Diaz & De La Fuente, 2007; Diaz & Batanero, 2009). From our perspective these results may be tied to traditional approaches to the subject.

The approach based on interventions, materials and activities we adopt here follows the line of (Martignon & Krauss, 2009). Here the advantages of natural frequency formats are consistently explored in a learning arrangement based on the interaction between groups of students of different ages (5th graders and 9th graders). In this learning-arrangement we let older students design hands-on activities to be carried out by younger ones.

The paper is organized as follows: first we introduce the theoretical background. After a general description of the design (2), we present the sequence of activities performed by the different groups. Then we report and analyze the results obtained. We conclude with a general discussion.

# PROBABILISTIC THOUGHT: SOME CONSIDERATIONS ON HEURISTIC APPROACHES, OBSTACLES AND STRATEGIES

## Representation Formats: The Natural Frequencies Approach

By natural frequencies formats in probabilistic reasoning we mean frequencies obtained by natural sampling preserving base rate and sub-sample information. Thus, samples are not standardized and sub-samples are not normalized at each step of reasoning. In a situation such as that of the mammography task (see subsection "Design/Materials" below), for instance, we do not talk about a probability of 1% of having the disease and a probability of 90% of having a positive test given that the patient is ill. Instead, we translate

this information into: "imagine 1000 women. Out of these 1000, 10 have the disease, and out of these 10, 9 have a positive test."

Natural frequency formats for probabilistic reasoning were introduced in the '90s (Gigerenzer & Hoffrage, 1995; Hertwig, 1995; Kleiter, 1994). They proved successful in eliciting normatively correct responses in probabilistic situations, notably in regard to Bayesian reasoning (Gigerenzer & Hoffrage, 1995) and probabilistic conjunctions (Hertwig, 1995). This success matches the premises of "ecological rationality" in the sense that humans' reasoning is adapted to information formats similar to those available in natural environments.

These results motivate an educational program which explores the possibility of improving students' skills at probabilistic reasoning by matching pedagogical strategies to cognitive processes. We adopt such a program even for early years in education.

We are also guided by Bruner's EIS-principle in the choice of materials: en-active and visual materials should be used before symbolic presentations: such materials have proven to be successful when combined with natural formats. Materials like "tinker cubes" are successful for illustrating and comparing different proportions and to introduce chance and other probabilistic notions (Martignon and Krauss 2007, 2009; Kurz-Milcke et al. 2008. See also other graphic and interactive devices at http://www.eeps.com/riskicon/).

## Engineering: from Psychology to Didactics

The so-called "rationality debate" in psychology leads us to search for the contexts and presentations which favor probabilistic reasoning. According to Meder and Gigerenzer (2014), in fact: "Instead of emphasizing human errors, the focus is shifted to human engineering: What can (and need) be done to help people with probabilistic inferences?"

The strategies followed in the ecological and bounded rationality traditions should be linked with the natural didactic aims of fostering the acquisition of mathematical competencies in students. The debates in psychology on the extent of biases and a different, more subtle treatment of "errors" or "mistakes", can in fact be approached from the mathematical didactics side using the conceptual tools proposed in the "theory of obstacles" (Brousseau, 2006). This theory presents a taxonomy of obstacles frequent in learning processes of any mathematical subject. We point to the relevant aspects in our study:

(1) *Epistemological obstacles*, those inherent to the concept itself. Probabilistic concepts, in fact, have been historically subject to foundational, conceptual and mathematical debates. This is reflected, for instance, in the different approaches to probability, whose very definition is far from straightforward.

(2) *Ontogenetic obstacles*, determined by the lack of developmental competencies necessary for the acquisitions of concepts. In our case, for instance, a certain level of numeracy, in particular understanding proportions, is required even for elementary probabilistic tasks. Furthermore, an obstacle for probabilistic thinking is the early lack of categorization strategies, e.g. coordinating extensional and intensional reasoning, see (Hertwig, 1995, pg. 3) [1]. Obviously, our conceptions of subjects' resources at a given stage depend on the position adopted in the rationality debate. Our position, as stated before, is guided by the tenants of ecological rationality.

(3) *Didactical obstacles*, which arise, for instance, in the communication between teacher and student, as is the case in typical misalignments between them. We focus not only on the already mentioned formats, registers and symbolisms tuned for a given purpose in a given moment, but also in the social roles established in the didactical situations proposed. In our study, students interact with each other adopting non-typical social roles (see 1.3).

Summing up, we see the necessity of designing appropriate activities and presentations. These are even more fundamental for concepts (such as the probabilistic ones) which can be dealt with, even theoretically, from multiple radically different approaches, and which can be unintuitive or even counterintuitive. We have to deal here with the problem of didactic transposition (Chevallard & Bosch, 2014): the scholarly knowledge

---

[1] The philosophical dichotomy "extensional-intensional" describes modes of defining collections, as proposed by the School of Port-Royal. Extensional refers to the (list of) members while intensional refers to the properties characterizing those elements. Gottlob Frege famously elaborated on the difference between the two approaches and, in his example, the intensional difference between "The evening star" and "the morning star", is not extensionally present: they are the same element (the planet Venus). The dichotomy is also reflected in mathematics, where a same set can be expressed extensionally (e.g. as {2, 4, 6, 8, 10}) or intensionally ({x | x is an even natural number and $1 < x < 11$}).

needs to be transposed into taught knowledge. Here, the proposed natural frequencies approach, is not entirely equivalent to the strictly probabilistic one. Yet, as it is claimed, it provides students (and humans in general) the possibility of handling probabilistic situations successfully and of capturing their meaning.

### Learning through Teaching: Interaction between Students of Different Levels

Students' active engagement, cooperative, team-based work has been extensively shown to be highly beneficial when appropriately implemented in learning environments (Slavin et al., 2003). Here we study an interaction between two levels, namely, 5th graders and 9th graders. This is conceived in order to benefit both scholarly levels. Part of the motivational expected success of the experience lay in the fact that probabilities were a marginal topic in their respective curricular trajectory, and that, being new at both levels, could be particularly appropriate for a "learning through teaching" design. In our case we analyze an interaction across levels in which both are approaching the subject at the same time, but, of course, with different tools/constraints posed by their specific level.

The condition for the selection of the scholastic level of the elementary school students was based on their mathematical skills and cognitive development. Following results from previous research in the same direction (Martignon & Krauss, 2009) and considering age and curricula, we selected 5th graders as an appropriate target group for the intervention. As for the older students we focused both on social and mathematical maturity.

## EXPERIMENTS/INTERVENTION DESIGN

The study was based on the interactions of two levels: 5th graders and 9th graders [2]. This took around 2 months (except for point 7, below). As shown in **Figure 1** the whole didactic sequence was developed according to the following stages:

(1) *9th graders pre-test.* It was conducted both in the target and the control groups. See next section.

(2) *Instruction.* After presenting the pre-test the tasks were discussed and the relevant mathematical concepts were introduced by teachers. The tasks were selected not just to reveal some problems or misconceptions, but as a means to introduce probabilistic concepts for the first time. The "fallacies" involved and their posterior analysis were intended to produce an "aha-moment" and to trigger the curiosity about the topic. The approach was based on natural frequencies.

(3) *Activities design.* Students were given 6 weeks to design activities which could be implemented in a 5th graders' class. This design was conducted in groups (usually formed of 4 students each) in 2 phases: first, a written proposal which explained the activity and its conceptual background, second the actual elaboration or choice of enactive materials as well as the performative action they would realize in the actual lessons. This action was simulated and video recorded. At each step students received feedback by their teachers.

(4) *5th graders pre-test.* Applied to a control group in a between-subjects design. See Section "5th Graders Pretest and Posttest."

(5) *Intervention.* At this stage 9th graders guided 5th graders during the realization of the designed activities. We considered that the intervention should be implemented also with the control group, because of its instructional value.

(6) *5th graders post-test.* After the activities the test applied to the control group was applied to the target group.

(7) *9th graders late post-test.* During the following school year (5 months later) "conjunction-fallacy"-style questions in contexts and formats different from the original one were posed in order to evaluate both transfer and sustainability of the investigated strategies. This test also included a questionnaire for evaluating students' reactions to and opinions on the experience.

We will next examine the different activities and results obtained.

---

[2] Students were members of Leonardo da Vinci School, a bi-cultural Colombian-Italian school at Bogotá, recognized by both Ministries of Education. Students are bilingual and mathematics is usually taught in Italian, the language we used in our tests.
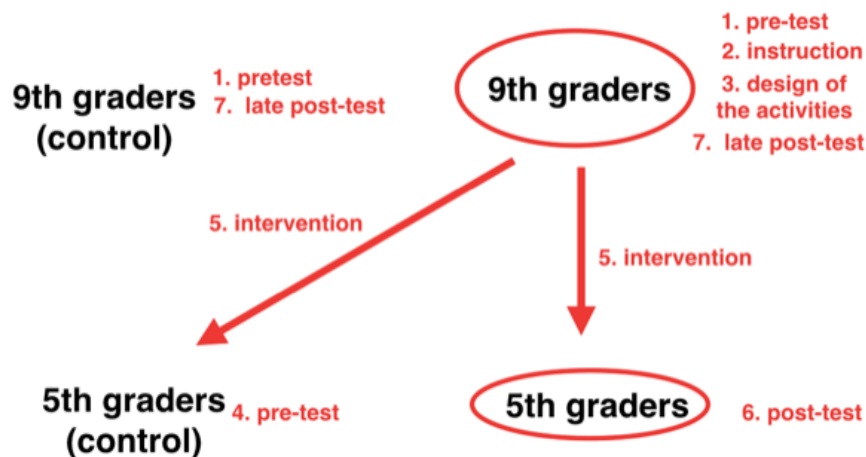
**Figure 1.** The didactic and experimental scheme

## 9th GRADERS' PRE-TEST

### Participants

48 9th graders: 24 males and 24 females. They were from three different groups. Students had some knowledge of statistics but no explicit study of probability theory. In fact, the test was presented as a first step for studying this topic.

### Procedure

Students were asked to answer a questionnaire during their mathematics class. Their time limit was the end of the class yet all of them finished in less than 25 minutes. Even if the questions were closed, students were invited to write any additional comment or justification they wanted.

### Design/Materials

We worked with written tests with 3 questions which we describe next:

(1) The classical Linda task (Tversky and Kahneman, 1983) with two options:

*Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?*

*[A] Linda is a bank teller.*

*[B] Linda is a bank teller and is active in the feminist movement.*

(2) Mammography (mainly based on Gigerenzer, 2008, pg.16): participants were asked to choose between 4 intervals of equal size. The exact phrasing of the question was as follows:

*Let's consider a woman who has just got a positive result in a mammogram test. Knowing the result of the test, she asks the doctor: "Is it certain that I am ill?" The doctor answers, providing the following data on the spread of the disease and the reliability of the test:*

- *The probability that a woman has the disease is 1%.*
- *If a woman is ill, the probability of the test being positive is 90%.*
- *If a woman isn't ill, the probability that the test is positive is 9%.*

*In your opinion, what is the probability that the woman has contracted the disease?*

*Choose from the following four options:*

*[A] 0–25% [B] 26–50% [C] 51–75% [D] 76–100%*

**Table 1.** Answers to the Linda question

| | Type of group | | |
|---|---|---|---|
| **Type of response** | **Treatment** | **Control** | **Total** |
| A | 8 | 6 | 14 |
| B | 25 | 9 | 34 |

According to standard terminology, the first piece of information conveys the *prevalence* (the *base rate*) of the test, the second one its *sensitivity*, and the third one its *false positives rate*.

We introduced a variant from typical literature: 24 of the subjects were presented the question without the 3rd piece of information, i.e., the "specificity". The other 24 were asked to answer the question with all 3 pieces of information.

Our working hypothesis was: if the results of the two groups were roughly the same this would point to the view that the 3rd piece of information is not taken into account for the answer. In other words, besides the "base rate neglect", there may be also a "false positives rate neglect" at play. This was confirmed by our results.

(3) A question which typically illustrates the phenomenon known as Falk's effect:

*Giovanni and Marco receive two boxes, each one containing 2 white balls and 2 black balls. Giovanni extracts a ball from his box and finds it is white. Without putting it back in the box, he extracts a second ball. The probability that this second ball is also white is less, equal or greater than the probability of it being black? Marco extracts out a ball from his box and puts it aside without looking at it. Then he extracts a second ball and sees that it is white. The probability that the first ball extracted is white is less, equal, or greater than the probability of it being black*

Falk's phenomenon is seen by Fischbein and Schnarch (1997) as showing the presence of the intuitive principle according to which "an event cannot act retroactively on its cause".

This question was included because it is interesting to compare the very different effects at play in question 2 and question 3: in 2, $P(C|E)$ and $P(E|C)$ tend to be interchanged. In contrast, in question 3 the situation is completely asymmetrical for subjects: in $P(E|C)$ E is dependent on C, whereas in $P(C|E)$, C is judged as independent of E. In other words, causal intuitions are stronger than those represented by the formal rules of probability.

Questions 1 and 3 have been studied with young students before (Fischbein & Schnarch, 1997). Question 2 has been the subject of a number of tests with adult subjects and with young students in different contexts and formats, see e.g. (Wassner et al., 2004, Zhu & Gigerenzer 2006).

# RESULTS

## Question 1

As expected, we obtained an overall prevalence of the B's (violating the conjunction rule). There was also a remarkably high number of A's. Can this be directly understood as a tendency to answer more correctly according to extensional, inclusion criteria? From the written comments that the participants added to their answers we can see that most of the A answers were not justified by set-theoretic inclusion reasoning, but by other reasons (see the corresponding analysis below).

In any case, the results are as expected, according to the literature. Fischbein and Schnarch (1997), in particular, report an 80% of incidence of the conjunction fallacy for 9th graders.

## Question 2

Separating the two conditions (including and not including the false positives rate of the test) we obtain the results shown in **Table 3**. According to the literature we are prone to neglect base rates. This is very dominant even among doctors and other specialists. A further question is: do participants provide the same amount of D-type answers with or without the explicit mention of the false positives rate? The present results not only show a similar high-D/low-A trend for both conditions, but that in the "without False Positives Rate" condition we have even a (small) increase in the A answers.

**Table 2.** Overall answers to the mammography question

| | Type of group | | |
|---|---|---|---|
| Type of response | Treatment | Control | Total |
| A | 8 | 4 | 12 |
| B | 2 | 2 | 4 |
| C | 4 | 3 | 7 |
| D | 18 | 6 | 24 |
| NA | 1 | 0 | 1 |

**Table 3.** Endorsement frequencies for the mammography question including and not including the false positives rate (FPR)

| | Type of group | |
|---|---|---|
| Type of response | With FPR | Without FPR |
| A | 4 | 8 |
| B | 2 | 2 |
| C | 5 | 2 |
| D | 13 | 11 |
| NA | 0 | 1 |
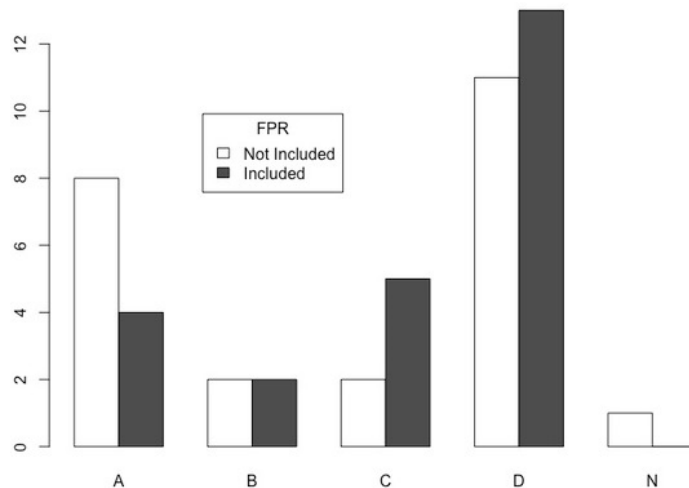


**Figure 2.** Endorsement rates for the mammography question with the false positives rate (FPR) not included and included

**Table 4.** Compared frequencies of the answers in the the two cases (Giovanni's and Marco's) asked for in Question 3

| | Type of group | | | | | |
|---|---|---|---|---|---|---|
| | Treatment | | Control | | Total | |
| | Giovanni | Marco | Giovanni | Marco | Giovanni | Marco |
| Larger | 5 | 1 | 2 | 1 | 7 | 2 |
| Equal | 4 | 20 | 2 | 14 | 6 | 34 |
| Smaller | 21 | 11 | 14 | 0 | 35 | 11 |
| NA | 0 | 1 | 0 | 0 | 0 | 1 |

## Question 3

The results in **Table 4** confirm the robustness of Falk's Phenomenon and its presence (for the same subjects) together with base rate bias, a phenomenon which apparently goes in the opposite direction.

**Figure 3.** "B is more probable because of her interest in themes of social discrimination and her participation in antinuclear demonstrations."

## RESULTS: HIGH SCHOOL STUDENTS' EXPLANATIONS

Going beyond the quantitative results we just presented, students' explanations, included in their answers, provide hunches about underlying processes at play. Explicitly externalized argumentation may not faithfully reflect these processes, yet it may provide some empirical indication about interpretation and reasoning mechanisms. This is relevant, as the same answer may be obtained by different students through entirely different processes, as we shall see. In what follows, we present a selection of explanations which strike us as interesting because they illustrate a diversity of interpretations.

### Question 1

The analysis of explanations reveals phenomena mostly already reported in the literature for adults, e.g. in the "think aloud" protocols in Hertwig (1995).

We present first some argumentations given by students in favor of their choice "B", i.e., of answers violating the conjunction rule. They reflect, in general, the use of communicative attitudes and assumptions tuned to a collaborative (as opposed to adversarial) interpretation of the information given.

This is present, in particular, in the use of pragmatic implicatures, in the sense of (Grice, 1975), that go beyond the strict (i.e., classical logical) consequences of the information given.

We move then to the analysis of the "A" answers and do not find radically different types of explanation from those of the "B's". The same kind of phenomena will often appear. We therefore conclude that even if these answers are consistent with an "extensional" interpretation of the task, in reality the actual reasoning appears to be guided by principles different from an extensional representation of it.

Some of the "B" answers must be seen, according to this, far from being just incorrect or fallacious. Instead of this, justifications show us in many cases some very reasonable ways to cope with the information given.

### *The expected answers*

A very common justification for violation of the conjunction rule answers (B in our case) makes use of the the "typicality" judgment. This is an important case in how the term "probability" may be understood in many different senses. It has been shown (Hertwig, 1995), that when there is a disambiguation about probability and typicality judgments in the same task, violations of the conjunction rule decrease substantially. Examples of justifications like "B is more probable because of her interest in themes of social discrimination and her participation in antinuclear demonstrations" (**Figure 3**) can be understood from this perspective.

It is difficult in some cases to discriminate the presence of tipicality judgments from other communication phenomena in action, with which it can be intertwined. Making explicit reference in argumentation to communicative pragmatic principles implies a higher level of discourse which goes beyond the information content, passing to the level of reasoning and communication principles involved. Even so, this meta-level of discourse is present in some of the justifications as we illustrate next.

Among the Gricean conversational principles we find that the information given should not be superfluous: it must be relevant. As stated in one of Grice's maxims: "Do not make your contribution more informative than is required". In **Figure 4** the student makes explicit use of this meta-discursive principle when he says: "It seems to me that the second possibility is more probable because if it were the other one, the given information about the interest in discrimination issues would be useless." As seen in the figure, the student even underlines some of the information as if saying: "Why do you give me this information, if you don't want me to use it?" This may even be stronger with traditional school tasks, where mathematical problems assigned usually do not provide useless information.

**Figure 4.** Grice collaborative relevance maxim exemplified



**Figure 5.** A and B interpreted as exclusive events



**Figure 6.** "Because her being a feminist is not mentioned in any place; even if she is very interested in social justice this doesn't mean she is an activist."

On the other hand, in **Figure 5** the student argues: "She is a woman who struggles for her ideals, therefore I don't think that she limits herself to be just a bank-teller, but she attempts also to change society in some way, and this is so if she is active in the feminist movement." We see another manifestation of the relevance principle: the "exclusion" implicature (see Hertwig 1995, p.45). According to this, the two options A and B are interpreted as exclusive events, as can be seen here by the opposition "but", and by expressions as "she limits herself" and "just a bank-teller". Based on this interpretation, the information provided about Linda would continue to be relevant, in contrast to the conjunction rule interpretation.

### Non fallacious but non-extensional

In different accounts of the Conjunction Fallacy, the answer "A" is assumed to be the "correct" one according to the inclusion A ∩ B ⊆ A and the monotonicity of the probability operator. In this sense, "A" answers are assumed to indicate an extensional interpretation of the situation.

Nevertheless, a completely different view emerges from the analyses of students' justifications of this choice. In fact, in our group, even if 29 % (14 students) of our participants made the "A" choice, none of them made an argument along these lines. In contrast our participants made reference to other kinds of argumentations using the information given in Linda's profile (which, according to the extensional interpretation, turns out to be superfluous).

The more predominant line of reasoning (it is present in 9 of the students) is based on the observation that in Linda's profile, her being a feminist is neither explicitly stated, nor a consequence of the information provided (**Figure 6**).

We do not consider this rationale to be extensional because the inclusion A∩B ⊆ A is in fact independent of B being or not being mentioned/implied. Even so, this line of argumentation is radically different from the

**Figure 7.** "Pseudo-extensional" justification



**Figure 8.** Arguing against stereotypes

**Table 5.** Frequencies of participants invoking pieces of information and their combination with BR = base rate, S = sensitivity and FPR= false positives rate. We separate participants in the standard condition (including the three pieces of information) and the condition with only BR and S (no FPR)

| | Type of response | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | BR+S+FPR | BR+FPR | S+FPR | BR+S | BR | S | NONE | TOTAL |
| STANDARD | 2 | 2 | 6 | 2 | 1 | 2 | 4 | 19 |
| NO FPR | 0 | 0 | 0 | 7 | 4 | 6 | 4 | 21 |

ones discussed above because participants present here a skeptical, critical position with respect to the link between the information given and the hypothesis "Linda bank teller".

Among our students, only one exhibited "pseudo-extensional reasoning" (Hertwig, 1995), i.e., based on the fact that "Linda bank teller" (BT) and "Linda feminist" (F) are taken as incompatible.

In other words, BT is assumed to be in contradiction to F, hence the choice A is the only option, as for the student in **Figure 7**: "A, because the stereotype of a bank teller is not that of taking part of demonstrations…"

Curiously, option A answers admit argumentations both based on stereotypes, as with the previous one, but also others justifying the choice against stereotypes (**Figure 8**): "Why should a leftish woman obligatorily be a feminist? Or, why if a woman doesn't follow the the mass media's ideals about what she is supposed to do when 30, is she necessarily a feminist? This is definitely a stereotype."

## Question 2

As a first step here we consider the information explicitly integrated in the justifications. We recall that there are three pieces of information involved:

· The base rate (BR) or prevalence: the prior probability that a woman has the disease, 1% in our question.

· The sensitivity of the test (S): the probability that the test is positive given that a woman has the disease, here 90%.

· The false positives rate (FPR): the probability that the test is positive even if a woman does not have the disease (9%).

Results in **Table 5** show that the most frequently invoked piece of information is sensitivity: 12 out of 19 in the first condition (against 7 and 10 for BR and FPR respectively) and 13 out of 21 in the second one (against 11 for BR). This is consistent with results in **Table 3** which are higher for answer D.

We also observe, that only 2 out of 19 students make reference to all three pieces of information, as Bayes' rule would require.

In the following we illustrate four salient misconceptions apparent in our students' argumentations.

[A] 0 – 25% [B] 26 – 50% [C] 51 – 75% ☒ 76 – 100%

Ha la pió alta possibilita perchè il dottor ha detto che la posibilita di avere la malatia con il test positivo è il 90% e lei ha il test positivo.

**Figure 9.** "Pseudo-extensional" justification

Scegli tra le seguenti quattro opzioni:

[A] 0 – 25% [B] 26 – 50% [C] 51 – 75% (D) 76 – 100%

Perche so il test è positivo quindi 90% è che lla la malattia e 9% che non ha la malattia

**Figure 10.** Sensitivity and false positive rate as alternative possibilities

[A] 0 – 25% [B] 26 – 50% [C] 51 – 75% [D] 76 – 100%

D. 100% perchè hoi riceuuio il risutato positivo è quindi lue la molatia

**Figure 11.** Deterministic understanding of an uncertainty situation

2. A non ci stanno fornendo dati sufficienti in fatti ci stanno fornendo dati delle donne in generale non della donna in particolare

**Figure 12.** Single case probability

### *Interchanging conditional probabilities*

As mentioned above a prominent phenomenon when dealing with conditional probabilities is interchanging P(X|Y) with P(Y|X). This appears in our results which show a great amount of high probability (D) answers. This "fallacy of the transposed conditional" (Falk, 1986) is even manifested explicitly by some students, as in **Figure 9**: "The doctor said that the probability of having the disease with the positive test is of 90%, and she received a positive test." This conveys the precise converse of the information given: "If a woman has the disease, the probability that the test is positive is 90%."

Others go even further and look at the sensitivity and the false positive rate and see these as making up a whole which actually makes up the positive tests. They are mistaking sensitivity with positive diagnosticity and at the same time, the false positive rate with the probability of not having the disease testing positive. 6 out of our 19 students presented this kind of answer. In **Figure 10**: "76-100%, because if the test is positive there is a 90% (probability) that she has the disease, and 9% that she doesn't."

### *Deterministic attitudes under uncertainty*

Again, as for Question 1, the uncertainty inherent in the notion of probability is difficult to grasp. In some of the answers the explanations exhibit a denial of uncertainty per se, and provide deterministic factual answers. In **Figure 11**, for instance, the student says that the probability is "100% because she had a positive result, and therefore *has* the disease" (our emphasis).

The interference between single case probabilities and statistical information is another problematic issue: Student in **Figure 12** answers "A" without conviction arguing that "they are not giving us enough data: in fact they are giving us data about women in general, not about this woman in particular." In an analogous way, the student in **Figure 13**, after providing support for his choice "D", concludes: "Anyway, if the doctor reports these statistics, the probability that the woman has the disease is a different one."

**Figure 13.** Statistics vs. single case probability



**Figure 14.** Trying to operate the numbers given



**Figure 15.** Mean value



**Figure 16.** Explicit base neglect

### The didactic contract in play

In some of the previous cases we can see students trying to overcome the need to operate and giving an apparently formal procedure with the given numbers. This is part of the established didactic contract (Brousseau, 2006). According to this author, in fact, students try to give "an explanation that the teacher wanted to hear." In many cases "the subjects produce the answer least incompatible with their knowledge, even when they see very well that it is false: the obligation of answering is stronger than that of answering correctly" (Brousseau, 2006, see also the "clause of the formal proxy" in D'Amore, 1999). If our students knew Bayes' formula they would probably try to apply it in a formal way. Not having this tool, they struggle to provide some calculation that helps them come up with "the correct" answer, no matter how nonsensical the operations are. We see this, e.g., in **Figure 14** where the student probably tried to remember something about how he used to deal with percentages.

Yet another example is given in **Figure 15**: there are three numbers given, so in order to find the answer he finds the mean value: "26-50% because the mean is 33.3%." This is even more salient here because the student first selected the intuitive option 76-100% and only later provided his final formal answer.

### Base rate acknowledged and yet wrong answers

The base rate neglect may occur not just as an unconscious phenomenon. Base rate may be acknowledged and yet considered useless. The student in **Figure 16** even crosses out this part of information in saying: "This datum seems to me superfluous, the important thing is the test reliability."

---

**Figure 17.** Base rate: contrary effect



**Figure 18.** "Giovanni: The possibility that it is white is less than before because there are two black ones and one white one. Marco: It's equal because when the first ball has been extracted we still had two black and two white ones."

Furthermore, as seen before (**Table 5**), almost half of our students invoked the base rate (prevalence) in their justifications. However, in two cases the low base rate information (1%) was not taken as an argument against a high probability of the woman having the disease given a positive test (as is really the case) but as an argument in favor of that high probability: "...therefore the woman has a big possibility of being ill, because, given that the disease is so rare, almost inevitably the test must be correct" (**Figure 17**).

Another student argues: "(She) has a probability of 75-100% because there are few cases of women with cancer, so the test has less margin to be wrong."

## Question 3

### *"Time axis fallacy"*

Some of the justifications make it very explicit that the two situations are completely asymmetrical exhibiting chronologist and causalist conceptions of conditional probabilities (Gras & Totohasina, 1995).

In **Figure 18** this is persuasively explained: the surrounding line frame establishes how the situation in fact is before the extraction of the ball for Marco and Giovanni. This apparent factual reality prevents the student from incorporating the (apparently irrelevant) information about the second extraction, which is in fact represented in the crossing of the white ball.

**Figure 19.** Conflict between (alleged) formality and (correct) intuitions



**Figure 20.** Uncertainty



**Figure 21.** Transferable probability

### Formality vs. intuition

In the answer given in **Figure 19** the student struggles between the previous position and taking into account the information given about the second extraction in Marco's question. She feels the conflict between what she supposes to be the correct, formal answer (which is in fact precisely the "fallacy" here) and the intuitions given by her "instinct": "...The extraction of a white ball in the second turn doesn't change the initial possibilities, even if instinct may lead you to think that the first one was a black one."

### Uncertainty and probability

As for the previous questions, some of the difficulties for our students were not about conditional probabilities themselves and the conflict with time or causality. In some cases students argue using a reason analogous to the laplacian "principle of indifference": in view of uncertainty the possibilities are equally likely. For the student in **Figure 20**, "For Giovanni (the probability) is equal because it is unknown where he put his hand extracting the ball, and the same is true for Marco."

Here, degrees of uncertainty are not considered: if we don't know "where the hand was" extracting the ball, we are in the condition of total indifference.

The student in **Figure 21**, makes use of this same "principle", but, incorporating the information acquired, makes use of a notion of probability as transferable: "In the case of Giovanni, after extracting the first white ball, he has a probability of 1/3 of extracting a white ball another time, but the probability of this ball is of 50% and not of 33.3%, as could be thought, because when we had the 4 balls ...each of them had the 25% of being extracted. But when a white one is extracted, the 25% of this unites with the 25% of the other white one, and for this reason it is more probable to extract another white one..."

Here probability is conceived as something objectively existing in the balls themselves which can be transferred from one of them to another becoming "united".

### Observations

Concluding, we have analyzed students' explanations and observed that simply classifying their choices as right or wrong is insufficient. The next step will be to show how these students prepared and implemented

interventions for 5th graders and how they changed their mental models through this process. Our next section describes the principles on which the intervention was based.

# THE INTERVENTION

9th graders were asked to work in a team collaborative setting. They had to prepare materials and activities designed to foster children's intuitions about probabilities. These activities were required to conform to the following guidelines:

- Introduce the expressions "more probable"/ "less probable" based on proportion comparisons.
- Use situations in which properties can be combined to form conjunctions and their conjoined probabilities can be assessed.
- Use proportions like "*m* out of *n*" when referring to favorable out of possible cases.
- *Avoid* using conjunction fallacy-type tasks, that is, of the Linda kind.

We introduced this last requirement because the conjunction fallacy is an important component of of the test used (see below). We wanted to avoid training effects: we did not want 5th graders to learn how to answer such specific questions, but to acquire an extensional training with conjunctions which could trigger an extensional treatment even for conjunctions of events described intentionally.

There were also some methodological requirements:

- The materials used should have enactive or iconic character (Bruner, 1966) facilitating interaction and visualization.
- All the members of each team should engage in an active role during the activities.
- Activities should to be guided by a Socratic interaction between 9th graders and 5th graders.
- Each activity was expected to last 10-15 minutes.

During the activities, 5th graders were divided into groups of 5-8 in different locations of the classroom, and 9th graders rotated around them. Teachers, both of the 5th and the 9th level, were inside the classroom as observers. Some of the activities were video recorded. We will briefly outline them next. Since we had two 9th grade classes, each of these groups intervened in one of the two 5th grade classes. For this reason only half of the following activities were performed in each of them [3].

We acknowledge that letting elementary school students be taught by their higher level and older companions is not something exempt from risks, possible perplexities or even mistakes (either conceptual or didactic). Here we tried to minimize these factors: (1) by proposing activities about a very specific topic and presented/represented with a number of constraints, as explained above, (2) by asking to 9th graders to write down an explicit rationale for the choices made and to simulate in advance the activities (video recordings) (3) from this, giving them feedback and preventing possible inadequate or incorrect features in their proposals. This is clearly organizationally demanding for their teachers and in traditional schools it is difficult to systematically repeat interventions like this one [4].

# 5th GRADERS' PRE-TEST AND POST-TEST

## Participants

84 5th graders, Ages between 10 and 12 years. Mean age=10.52 years 42 males, 42 females. They had no knowledge of probability theory. In fact, the intervention was a first approximation to the topic.

## Procedure

The 5th grade students were asked to answer a 4 page written questionnaire during their math class. All of them finished in less than 20 minutes. 40 of the participants took the questionnaire before the intervention and the other 44 after a 1 hour intervention.

---

[3] A detailed description of the activities is offered as an Appendix.

[4] Nevertheless, the pay-offs of this kind of designed interaction may make it worth exploring in connection with other mathematical subjects as well, especially if they are faced with a problem-solving approach, something to be done perhaps in interactions more extended in time than the one presented here.

**Figure 22.**

## Design/Materials

The 5th grade students were asked to answer a 4 pages written questionnaire during their math class. All of them finished in less than 20 minutes. Tests with 5 tasks each.

The tasks of the test are inspired by experiments in (Multmeier 2012; Massini, 2018). Here the tasks were slightly modified and adapted. The most salient feature in all of them is the use of natural frequency formats and therefore an extensional representation of the situations. This includes, in particular, an interpretation of the conjunction in accordance with set theoretic intersection.

For example:

*Thirty men live in a small village, you can see them represented in the figure* (**Figure 22**)

*Look at all the men in the village: are there more men with moustaches who are wearing a hat or more men with moustaches who are wearing a hat?*

*Your answer:*

• *more men with a moustache wearing a hat*

• *more men with a moustache who are not wearing a hat*

*Note: In the following questions, you always must write the total number of men on the right side!*

*"How many men are wearing a hat?" Your answer: ......... out of ........*

*"How many men have a moustache? Your answer: ........ out of ........*

*If you randomly choose a man in the village, is it more likely that he is wearing a hat or that he has a moustache and is wearing a hat?*

*Your answer:*

A crucial question in the test is an adapted version of the Linda problem.

*Marco lives in Bogota and is a 7th grade student. He likes to solve sudokus and play chess in the championships of his city. If you have to make a bet, what would you choose?*

*Marco is good at basketball*

*Marco is good at basketball and at math.*

## Results

Here we focus on a comparison between the two groups. The percentage of correct answers for each of the questions is represented in the graph.
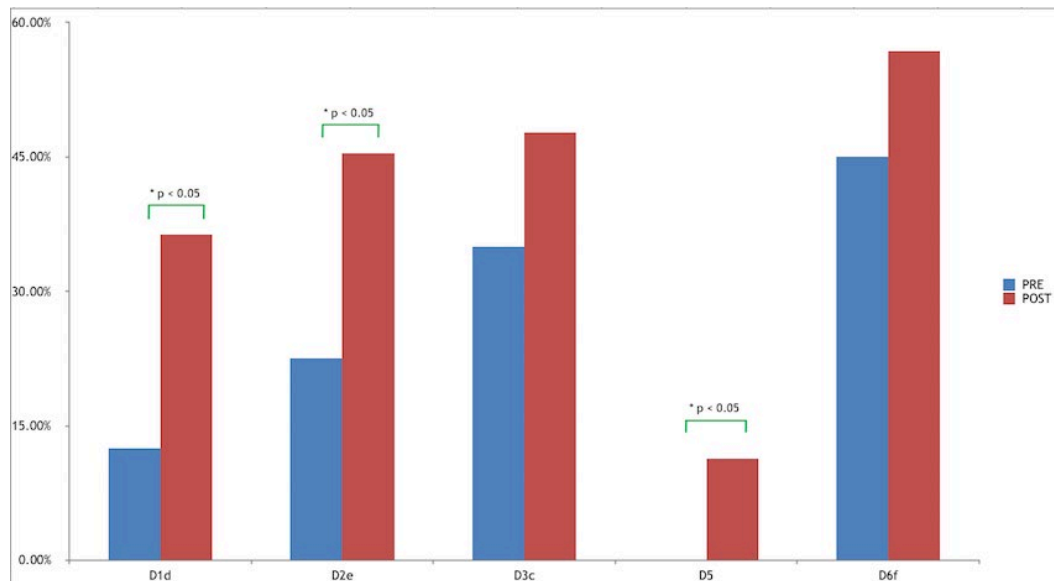
**Figure 23.** Pre-test/Post-test comparison in 5 questions about frequencies involving conjunction of characteristics
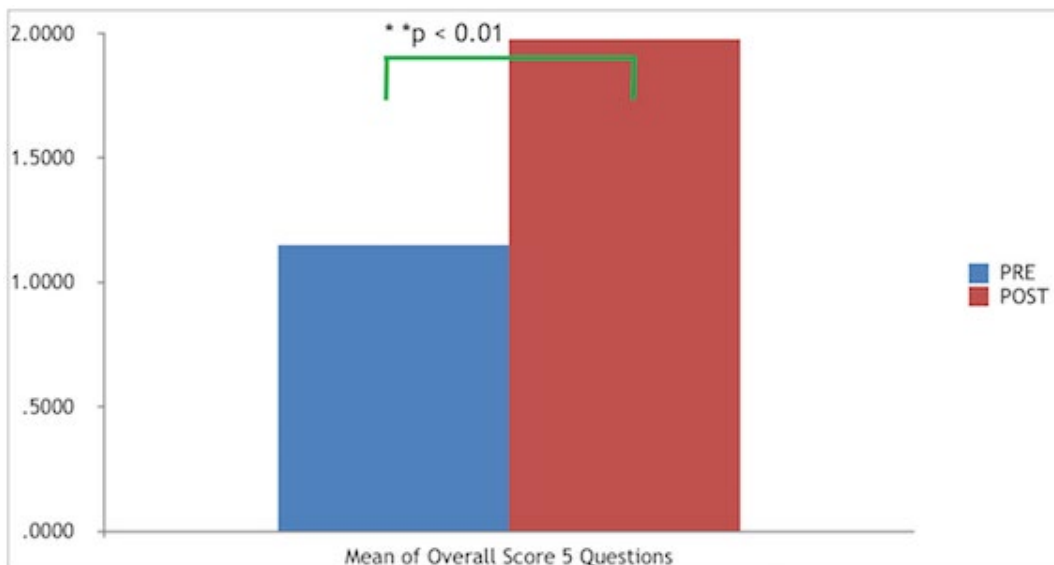


**Figure 24.** Mean performance in the 5 conjunction questions (total score=5)

We can see a general improvement in the performance on the test. Comparing the means, we see a change from the 52 % to the 68% percent.

Nevertheless, the really crucial questions, from the logical and probabilistic point of view, are considered separately in **Figure 23**. It is in these answers, in fact, that we can see a shift in the interpretation of the conjunction towards the intended one.

As we can observe, there is a consistent improvement in all the answers. The mean score almost doubles from 23% to 39.5%. Question D5, in particular, passes from being not answered "correctly" by any of the participants to a percentage of 11% (5 participants). This shows a transfer effect in this problem, even after a very short intervention, as in this case.

# 9th GRADERS' POST-TEST

## Participants

The same 9th grade students that participated in the pre-test (in Section "9th Graders Pretest"). The post-test was conducted 5 months later than the intervention, so actually the students were at this point at grade 10. For this reason, the same 3 classes participated but some of the students were no longer in the groups. For reasons that we explain next, the number of subjects who participated in the pre-test and that provided complete answers was 22 in the treatment group, and 10 in the control group.

## Procedure

At the beginning of grade 10, students presented a diagnostic proof about the mathematical contents of the previous grade, among them some questions on probabilities (which made up our post-test). Our purpose was to assess the assimilation of the concepts and techniques described above. We did not want to let students take an ad hoc post-test on biases. We chose to ask them to solve tasks included among others on different mathematical topics in a regular setting. Given that some of the participants concentrated on other questions, there was an important decrease in the number of answers provided to our tasks.

## Design/Materials

We presented in the test two "conjunction fallacy" questions, analogous to some of those included in (Tverski & Kahneman, 1983)

It would have been too obvious if we presented a Linda-style question, so we presented 2 questions in a different setting and with no explicit use of the word "and".

**Question 1**

*Next November 12, there will be a football match between Lichtenstein and Italy for qualifying for the World Cup. Sort the following results from the least likely to the most likely:*

*a) Italy wins the match*

*b) Italy loses the first half of the match*

*c) Italy loses the first half but wins the match*

*d) Italy wins the first half but loses the match*

**Question 2**

*Consider a regular 6-sided die with 4 green faces and 2 red faces. Which of the following sequences is most likely to come out when rolled? Why?*

*1) RGRRR*

*2) GRGRRR*

*3) GRRRRR*

*4) RRRRRR*

## Results

Students were not able to make the transfer of the concepts studied in the setting of the second question. Here the misuse of proportional thinking and interference with the concept of randomness were so strong that all the answers given were for option 2.

As for the first question, many of the students succeeded in answering it respecting the conjunction rule. Here actually, there were two cases in which it had to be applied:

Event a) is more probable than the conjunction c)

Event b) is more probable than the conjunction c)

These were the only restrictions that we considered for classifying an answer as respecting the conjunction rule. The second is really the harder one, for according to typicality criteria c) is designed to be more appealing than b).

**Table 6.** Pre-test/Post-test comparison of answers presenting conjunction rule violations (CRV), or respecting the conjunction rule (CR)

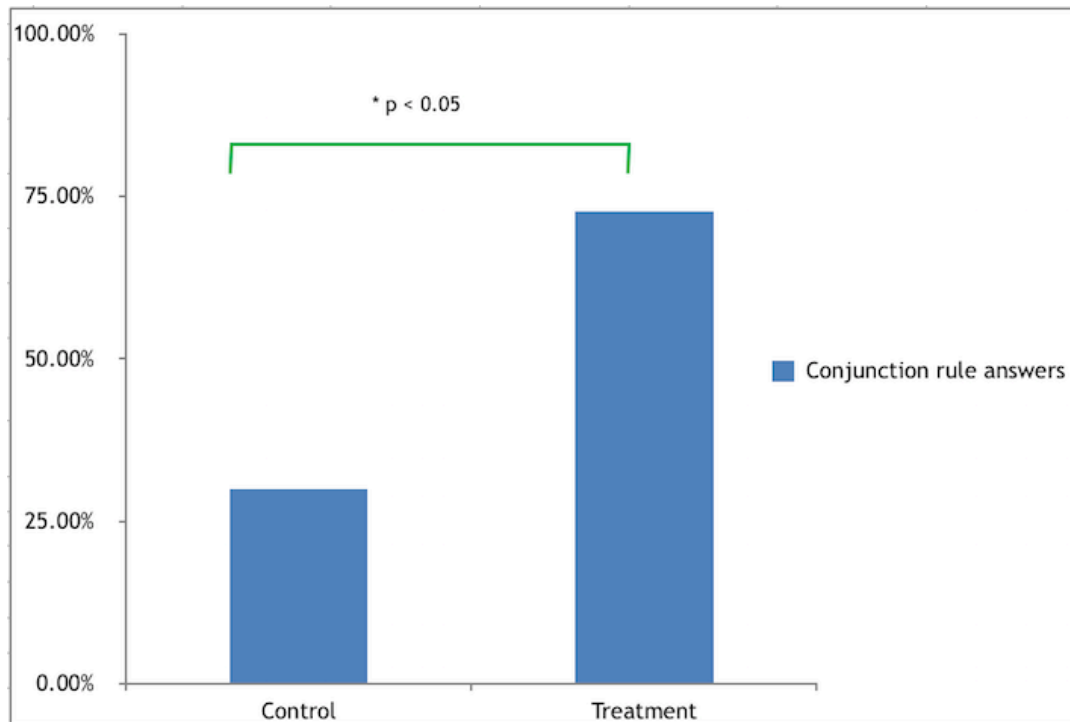| | Type of response | | |
| --- | --- | --- | --- |
| | **CRV (Post-test)** | **CR (Post-test)** | **TOTAL** |
| CRV (Pre-test) | 6 | 11 | 15 |
| CR (Pre-test) | 0 | 5 | 5 |
| TOTAL | 6 | 16 | 22 |



**Figure 25.** Treatment vs. control comparison

### Pre-test vs. post-test comparison

The conjunction fallacy question of the pre-test and this question of the post-test are not directly comparable even if they focus on the same phenomenon. This is so because of format reasons: in the pre-test we had only 2 options, whereas in Question 1 here, there are 4!=24 possible orderings for the 4 events described. Out of these, only 8 respect consistently the conjunction rule. This factor represents an increase of difficulty in the question posed.

We obtained the results reported in the **Table 6**. According to these, 11 out of the 22 subjects who answered these questions in both tests, moved from a choice violating the conjunction rule in the pre-test, to an answer consistent with it.

### Treatment vs. Control Comparison

Performance in the two groups is shown in **Figure 25**. We applied Fisher's exact test which indicates that the difference is significant (two-sided Fisher's exact test; $p = 0.049$; Cramer's V = 0.40). This indicates that, the odds of answering the question correctly was 6.22 times higher in the treatment group than in the control group.

These results suggest that the whole process of design and intervention with activities did have an effect in triggering an extensional treatment of conjunction fallacy tasks. This effect cannot be attributed only to the correction and explanation after the initial pre-test, which was provided also to the control group.

| QUESTION | MEAN |
|---|---|
| A) The fact of doing the initial test helped me to understand the difficulties present in the different probabilistic notions. | 4.2 |
| B) I think that having to do a research on a topic and propose a way to develop activities in other class groups is a strategy that foster understanding in mathematical topics | 4.4 |
| C) I felt that this way of developing topics reinforced my curiosity and motivation. | 4 |
| D) I think that this way of developing topics reinforced the curiosity and motivation of 5th graders. | 4.1 |
| E) I think that 5th graders were able to approach some basic concepts of probability and that the proposed activities gave suitable elements for their comprehension. | 4.5 |
| F) I endeavored to make the selected activities clear and entertaining and to make the experience a constructive one for all of us participating in it. | 4.7 |
| G) The fact of having to present and teach these themes helped me to reflect on the difficulties of teaching and the difficulties of learning. | 4.5 |
| H) I would like to repeat experiences of this kind. | 4.4 |

**Figure 26.** Questions of the survey, evaluated on a scale from 1 to 5

## A SURVEY

Beyond the previous results, we were interested in evaluating 9th graders' own perception of different aspects of the didactic experience. We assessed their perceptions by means of a survey (**Figure 26**).

The survey considered three main dimensions:

- *Cognitive and metacognitive.* In question A we focused on the use of the pre-test instrument in order to elicit students' understanding of their starting point. Question B and E focused on the cognitive byproducts of having to prepare a topic in order to teach it, and in the suitability of the activities themselves.

- *Motivation/engagement.* This aspect was one of the aims of the whole design. We assessed it from the point of view of the 9th graders' attitude towards their own curiosity, motivation and effort (questions C and F), from their point of view of 5th graders' attitude towards their own curiosity and motivation (question D), and of the possibility of repeating similar experiences in the future (question H).

- *Didactic awareness.* We finally focused on the benefits of having better insights into the different obstacles faced by themselves and the younger students. This is inquired into, as mentioned, by question A; in question G we assessed the students' view on their "learning by teaching".

In all of these aspects, the students' perception was positive, as can be seen in the table in **Figure 26**. All the evaluations were in mean either 4 or above in a scale from 1 to 5. This is relevant, since many of the participants did not usually have a very positive perception and attitude towards mathematics, in general.

## GENERAL DISCUSSION

We summarize here the main achievements of our study covering three aspects: (1) the relevance of analysing students' argumentations which provide insights on their reasoning; (2) the advantages of working with natural frequency formats like "... out of ..." for fostering children's competencies in probabilistic thinking and for triggering extensional reasoning in conjunction tasks; (3) the success of "learning by teaching" and "learning from peers" procedures in school levels interactions.

## Argumentation analysis and the psychology of reasoning

It is clear, from our analyses of Section "Results High School Students' Explanations", that multiple choice questionnaires are not enough for understanding the reasoning processes that lead to some of the answers or to classify a given option as a "bias", as if it were a well-determined phenomenon. The three questions examined show that a wide spectrum of reasoning processes may lead to the same choice.

Similarly, the notion of "error" in math education, may turn out to be a simplistic one, and may predispose us to ignore common sense or reasonable principles, which may be appropriate in habitual circumstances. Our aim was to substantiate the thesis that the tasks analysed are heavily dependent on interpretational issues. Interpretation and reasoning are in fact two processes intimately connected and what we see in practice is a continuous back and forth between these two stances ("reasoning to an interpretation" and "reasoning from an interpretation", in (Stenning & van Lambalgen, 2008) terminology.

In this context, the notion of "obstacle" seems to be more neutral than those of "bias" and "error" and leads us to the design/engineering problem of endowing our non-idealized students, with adequate representation tools which can help them cope with counter-intuitive notions. We also highlight that among the obstacles found, the understanding of the very notion of probability is a remarkable one: already Piaget considered the crucial passage from not being able to distinguish necessary from chance phenomena, to developing the concept of chance (Piaget & Inhelder, 1975). This transition would occur during the formal operational stage, yet we still see remnants of its lack, for example in **Figure 11** and **Figure 20**.

We also observe in the examples provided other conflicts regarding the concept of probability: it may be understood as something objectively in the world or something dependent on the information available, it may depend on the statistics of a whole population or refer just to single cases…

These examples support the fact that the term "probable" does not have a unique meaning as shown by (Hertwig, 1995).

## Ecological Rationality and Math Education

In general terms, the results obtained here confirm the existent literature, on the facilitating effects of natural frequencies for probabilistic reasoning at different developmental stages. This may be reinforced by pictorial/enactive representations and interactive/socially engaging activities. Here we place ourselves in the tradition of ecological rationality: probabilities are not per-se inaccessible to our minds, but they can be grasped if translated into appropriate representation formats (here, natural frequencies). This suitability is a consequence of how information was available in the environments that shaped our cognition.

## Social Aspects, Engagement and Meta-cognition

Sociocultural aspects connected to mathematics learning and teaching are known to be anything but minor factors in the didactic process. In the activities and interactions which we described above these were central in at least two instances:

(1) in the interaction between members of the teams formed by 9th graders in order to design, prepare and perform the activities, and

(2) in the interaction between 5th graders and 9th graders.

In the first instance, students had to cooperate with each other. Their teachers noticed that many of them were more committed than usual, possibly due to their common purpose. This can also be seen in their video recordings. Each of them had a specific role in the team. More importantly, they also had to play, as a team, the new "game" or role of having to teach somebody else. This made them approach the subject matter in a different way that was "fun" and worth exploring. In fact, several students expressed that they would like to repeat the experience (see question H of the survey). This goes in line with the perspective of the "engagement structures" (Goldin et al., 2011) in action. Even if here the situation was not spontaneously generated in the interior of the classroom, what 9th graders did and how they did it was certainly in accordance with the "Let Me Teach You" structure described by Goldin and his collegues.

Furthermore, the expected benefits for 5th graders were inspired by the possibility of a partial suspension of the constraints inherent in the didactic contract (Brousseau). Observing their engaged older school-friends and the games these prepared for them was appealing and inspirational. These practices can become a valuable complement to traditional teaching as suggested, e.g., by Vigotsky's Sociocultural Theory. As J. Harrys claims in her theory on "the nurture assumption", children's behavior is not shaped primarily by that

of their parents or teachers, but by that of their peers: "Children identify with a group of others like themselves and take on the norms of the group" (Harris, 2011).

## ACKNOWLEDGEMENTS

We thank Lars Uhlig for his statistical support and Dr. Peter Dines for reading and improving linguistic aspects of this paper.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

**Francisco Vargas** – Universidad el Bosque, Colombia and Ludwigsburg University of Education, Germany.

**Tommaso Benincasa** – Universidad el Bosque, Colombia and Leonardo da Vinci School, Colombia.

**Giuseppe Cian** – Leonardo da Vinci School, Colombia.

**Laura Martignon** – Ludwigsburg University of Education, Germany.

## REFERENCES

Brousseau, G. (2006). *Theory of didactical situations in mathematics: Didactique des mathématiques*, 1970–1990 (Vol. 19). Springer Science Business Media.

Bruner, J. S. (1966). *Toward a theory of instruction*, Cambridge, Mass.: Belkapp Press.

Chernoff, E., & Sriraman, B. (2014). Probabilistic thinking. *AMC*, *10*, 12. https://doi.org/10.1007/978-94-007-7155-0

Chevallard, Y., & Bosch, M. (2014). Didactic transposition in mathematics education. In *Encyclopedia of mathematics education* (pp. 170-174). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_48

D'Amore, B. (1999). *Elementi di didattica della matematica*. Bologna: Pitagora.

Diaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education*, *4*(3), 131-162.

Diaz, C., & De La Fuente, I. (2007). Assessing students' difficulties with conditional probability and Bayesian reasoning. *International Electronic Journal of Mathematics Education*, *2*(3), 128-148.

Duval, R. (2017). *Understanding the Mathematical Way of Thinking–The Registers of Semiotic Representations*. Springer International Publishing. https://doi.org/10.1007/978-3-319-56910-9

Engel, J., & Sedlmeier, P. (2005). On middle-school students' comprehension of randomness and chance variability in data. *ZDM*, *37*(3), 168-177.

Falk, R. (1986). Conditional probabilities: insights and difficulties. In *Proceedings of the Second International Conference on Teaching Statistics* (pp. 292-297).

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological research*, *50*(2), 123-129. https://doi.org/10.1007/BF00309212

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational studies in mathematics*, *15*(1), 1-24. https://doi.org/10.1007/BF00380436

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for research in mathematics education*, 96-105. https://doi.org/10.2307/749665

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology*, *28*(2), 210. https://doi.org/10.1037/a0014474

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

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological review*, *102*(4), 684. https://doi.org/10.1037/0033-295X.102.4.684

Goldin, G. A., Epstein, Y. M., Schorr, R. Y., & Warner, L. B. (2011). Beliefs and engagement structures: Behind the affective dimension of mathematical learning. *ZDM, 43*(4), 547. https://doi.org/10.1007/s11858-011-0348-z

Gould, S. J. (1992). Bully for Brontosaurus: Reflections in natural history, New York: Norton.

Gras, R., & Totohasina, A. (1995). Conceptions d'élèves sur la notion de probabilité conditionnelle révélées par une méthode d'analyse des données: Implication-similarité-corrélation. *Educational Studies in Mathematics*, *28*(4), 337-363. https://doi.org/10.1007/BF01274078

Grice, H. P. (1975). Logic and conversation, in Cole, P. and Morgan, J. L. (eds.), *Syntax and Semantics 3: Speech acts (*pp. 41-58), New York: Academic Press.

Harris, J. R. (2011). *The nurture assumption: Why children turn out the way they do.* Simon and Schuster.

Hertwig, R. (1995). *Why Dr. Gould's homunculus doesn't think like Dr. Gould: The "conjunction fallacy" reconsidered.* Konstanz: Hartung-Gorre.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of behavioral decision making*, *12*(4), 275. https://doi.org/10.1002/(SICI)1099-0771(199912)12:4<275::AID-BDM323>3.0.CO;2-M

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of and. *Cognition*, *108*(3), 740-753. https://doi.org/10.1016/j.cognition.2008.06.008

Hintikka, J. (2004). A fallacious fallacy? *Synthese*, *140*, 25–35. https://doi.org/10.1023/B:SYNT.0000029938.17953.10

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures.* New York: Basic Books. https://doi.org/10.1037/10034-000

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: heuristics and biases.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809477

Kleiter, G. D. (1994). Natural sampling: Rationality without base rates. In *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375-388). Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4308-3_27

Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication. *Annals of the New York Academy of Sciences*, *1128*(1), 18-28. https://doi.org/10.1196/annals.1399.004

Martignon, L. (2014). Fostering children's probabilistic reasoning and first elements of risk evaluation. In *Probabilistic Thinking* (pp. 149-160). Springer Netherlands. https://doi.org/10.1007/978-94-007-7155-0_9

Martignon, L., & Krauss, S. (2009). Hands-on activities for fourth graders: A tool box for decision-making and reckoning with risk. *International Electronic Journal of Mathematics Education*, 227–258

Massini, J. (2018). *How to foster the coordination of intensional and extensional reasoning in young children Analyzing the Conjunction Fallacy in 4th class.*, Master thesis im Masterstudiengang (M.A.): Paedagogischen Hochschule Ludwigsburg.

Meder, B., & Gigerenzer, G. (2014). Statistical thinking: no one left behind. *Probabilistic Thinking* (pp. 127-148). Springer Netherlands. https://doi.org/10.1007/978-94-007-7155-0_8

Multmeier, J. (2012). *Representations facilitate Bayesian reasoning: computational facilitation and ecological design revisited* (Doctoral dissertation, Freie Universität Berlin).

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children.* Norton.

Saldanha, L., & Liu, Y. (2014). Challenges of developing coherent probabilistic reasoning: rethinking randomness and probability from a stochastic perspective. In *Probabilistic Thinking* (pp. 367-396). Springer Netherlands. https://doi.org/10.1007/978-94-007-7155-0_20

Slavin, R. E., Hurley, E. A., & Chamberlain, A. (2003). Cooperative learning and achievement: Theory and research. *Handbook of psychology*, 177-198. https://doi.org/10.1002/0471264385.wei0709

Stenning, K., & van Lambalgen, M. (2008). *Human Reasoning and Cognitive Science.* Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/7964.001.0001

Tversky, A., & Kahneman, D. (1974). *Judgment under uncertainty: Heuristics and biases.* science, 185(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1983): Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293-315. https://doi.org/10.1037/0033-295X.90.4.293

Wassner, C., Martignon, L., & Biehler, R. (2004). Bayesianisches Denken in der Schule. *Unterrichtswissenschaft*, *32*(1), 58-96.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*(3), 287-308. https://doi.org/10.1016/j.cognition.2004.12.003

# APPENDIX: DESCRIPTION OF THE ACTIVITIES

## The Die

9th graders built a 20-faced die, whose faces had numbers from 1 to 20. The faces could be colored either blue or orange. The idea was to have, for example, a different number of even faces colored with blue and with orange. 9th graders could interview their younger schoolmates asking questions like: "How many orange faces do the die have?", "How many orange faces have even numbers?" "How many blue faces have even numbers?" or "If you bet by rolling the dice, what would you bet: will an even and blue number come out or an even and orange one?".
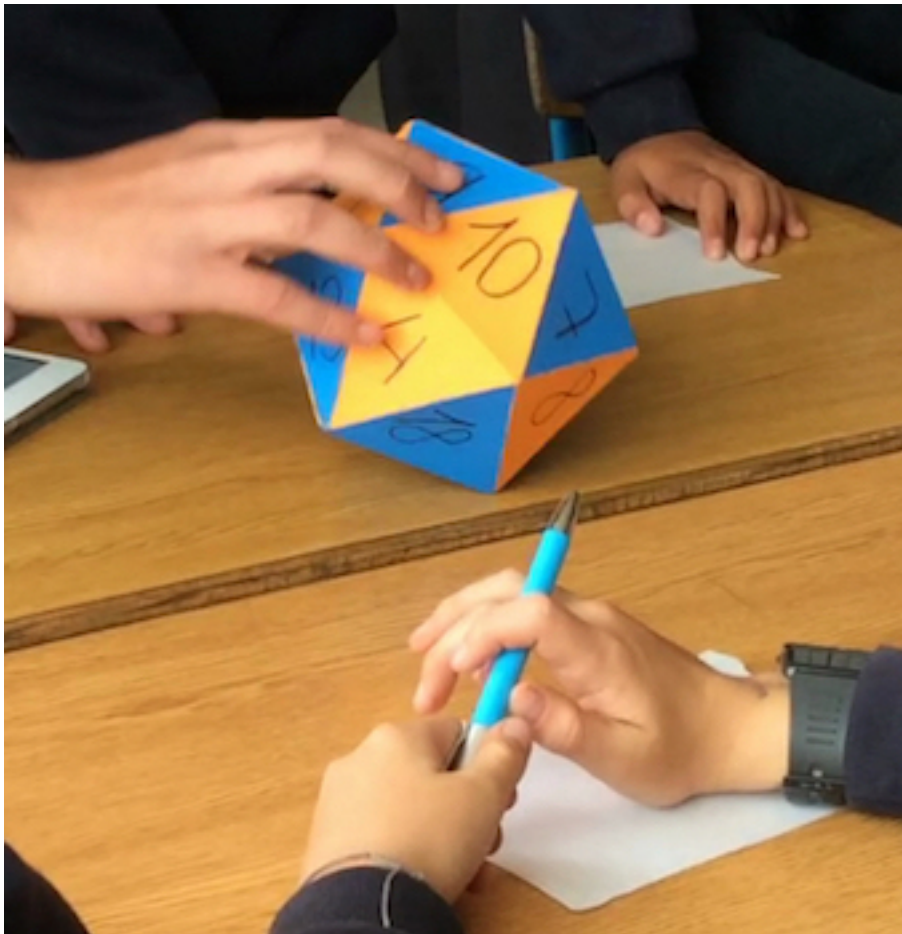


**Figure 27.** The die

## The Stadium and the Monopoly Game-board

The students built a polystyrene football stadium. The idea was to simulate the distribution of fans in the stands during a football match between two football teams. They distributed thumbtacks to 5th graders representing the supporters of the two teams, either blue or red. They asked the students to identify in the model of the stadium the different grandstands. Next, they placed the thumbtacks on them. Here the
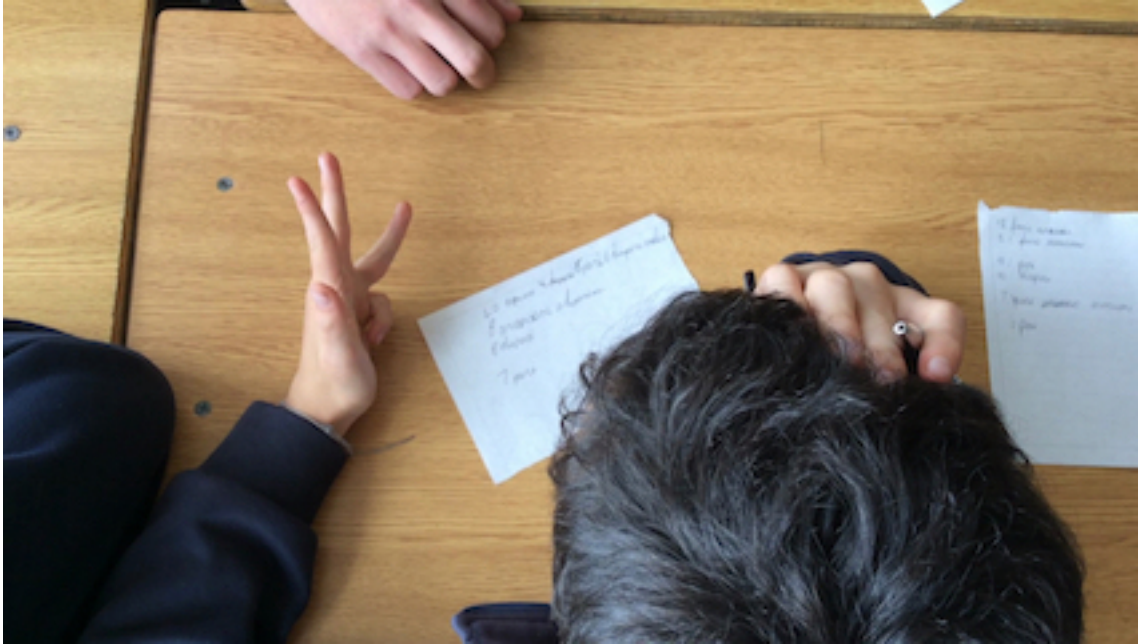
**Figure 28.** A student taking notes



**Figure 29.** The stadium

conjunction was obtained taking into consideration which team the fans supported and where they were placed.

The activity with the Monopoly game-board was similar. 9th graders let their younger asked schoolmates to place some house and hotels over the different properties. The conjunction of events has been achieved by considering two different colors properties and on the number of houses and hotels over them.

## Playing Cards

Among the games prepared by 9th graders, three made use of cards: two of them with French playing cards and the other one with pokemon playing cards. 9th graders prepared an easy game to start the activities. Using the French playing cards they selected, from the whole deck, a ten card subset, which was composed of nine black cards and a red one. The game consisted of several rounds: 5th graders were challenged to find a black card without seeing the deck. At the end of each stage, the selected card was discarded and the next round began with one card less. Students were also asked to stop the game whenever they found it to be too risky. After few turns students learned that it was easier to lose the game when the set of cards became smaller, since the likelihood of choosing the red card grew steadily. In the other activity the students had to consider whether the card was a figure or number, and if it was of clubs or hearts. In the pokemon card case the students had some fire or water pokemon cards with a power higher or lower than 100 PV.



**Figure 30.** Pokemon cards

## The Simpsons

9th graders printed more than twenty images representing the faces of the characters of "The Simpsons" and hung them on the white-board. Younger students sat in front of them and were asked to play a game like 'Who am I?'. Older students simulated a TV quiz, choosing a character and challenging younger schoolmates to guess who he was. They started to ask several questions while presenters removed the figures that had to be discarded. Older students used Euler diagrams in order to represents sets discarded and remaining after each question thus emphasizing, among other aspects, the inclusion relation and the effects of conjunctions and negations.

**Figure 31.** The Simpsons

## The Lego Bricks and Football Clubs T-shirts

This activity was composed of two successive stages. At first, 9th graders, assuming the teacher's role, asked their 5th graders to divide themselves into two groups (boys and girls), to count the number of members of each group and to make a note of the results. They then created three other groups; in this case, the characteristic to take into account was the hair color (blond, brown, red), also in this case the students made a note of the number of elements of each set. Finally the conjunction of events was introduced making students reflect on the number of girls with blond hair and girls with brown hair in the classroom by comparing the results to the total number of girls and students with blond or brown hair.

In a second stage, 9th graders took some different colored Lego bricks (blue, pink, brown, yellow and red). They represented the boys in the class with blue bricks, the girls with pink ones, brown for students with brown hair etc. Then 5th graders counted the blocks of different colors. In order to represent the status of the class using conjunctions, 9th graders joined the bricks (a blue and a brown block to represent a boy with brown hair, a pink and yellow one to represent a blond girl etc.). Again, the students made a note of the results of the various counts and 9th graders asked questions to make younger students reflect on the conjunction of events:

"Taking by chance a student in your class, what would you bet: that she is a girl or she is a girl with blond hair?"

At the end, 9th graders stressed that in the Lego bricks task the students obtained the same results as with real boys and girls.

Another activity that took advantage of the division between boys and girls in the class was achieved by using the football preferences of 5th grade students. In fact, the 9th grade students have brought T-shirts of two popular football clubs. They handed out T-shirts among 5th grade students depending on their football preferences. In this case, the event was the conjunction between being a boy or a girl and cheer one or another of the teams.

## Popcorns and Sweets

Another simple game was prepared by 9th grade students: they cooked different kinds of popcorns, each one with a different color and taste: sweet ones (1/3) and two salted ones: normal (1/3) and with cheese (1/3). 9th graders put an equal number of each type in a little bag and asked the younger ones some questions like: "If you had to, would you bet that you would randomly choose a salted one or a sweet one?" or "Would you bet that in a random choice, you would select a salted one or one with cheese?" (having in this last question a set inclusion).

A similar activity was performed with candy bars. In this case 9th graders bought biscuits and chocolates, the other feature to consider was if the candy bars had a vanilla or cocoa taste.

Random selections were actually done mixing the candy bars in a bag.



**Figure 32.** Sweets

http://www.iejme.com