Probability in reasoning: A developmental test on conditionals

Barrouillet, Pierre Noël; Gauffroy, Caroline

Cognition 137 (2015) 22–39

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Probability in reasoning: A developmental test on conditionals

Pierre Barrouillet *, Caroline Gauffroy

Université de Genève, Switzerland

ARTICLE INFO

Article history:
Received 28 May 2014
Revised 8 December 2014
Accepted 10 December 2014
Available online 12 January 2015

Keywords:
Reasoning
Conditional
Development of reasoning
Mental models
Probabilities

ABSTRACT

Probabilistic theories have been claimed to constitute a *new paradigm* for the psychology of reasoning. A key assumption of these theories is captured by what they call *the Equation*, the hypothesis that the meaning of the conditional is probabilistic in nature and that the probability of *If p then q* is the conditional probability, in such a way that P(*if p then q*) = P($q|p$). Using the probabilistic truth-table task in which participants are required to evaluate the probability of *If p then q* sentences, the present study explored the pervasiveness of *the Equation* through ages (from early adolescence to adulthood), types of conditionals (basic, causal, and inducements) and contents. The results reveal that *the Equation* is a late developmental achievement only endorsed by a narrow majority of educated adults for certain types of conditionals depending on the content they involve. Age-related changes in evaluating the probability of all the conditionals studied closely mirror the development of truth-value judgements observed in previous studies with traditional truth-table tasks. We argue that our modified mental model theory can account for this development, and hence for the findings related with the probability task, which do not consequently support the probabilistic approach of human reasoning over alternative theories.

1. Introduction

The capacity to reason is of paramount importance for members of the Homo sapiens species and, not surprisingly, understanding how human beings reason and how this capacity develops with age have been among the main aims of psychology. This enquiry has for a long time been connected with the questions of rationality and logic. Accordingly, Piaget described intellectual development as a progress toward rationality through the construction of mental operations structured in a logical way (Inhelder & Piaget, 1958; Piaget & Inhelder, 1959). More recently, prominent theories suggested the existence in human mind of logical rules constituting a form of mental logic (Braine & O'Brien, 1998; Rips, 1994). Alternative accounts were proposed that denied the existence of such rules, assuming that people reason by constructing and manipulating mental models of the state of affairs the available premises refer to (Johnson-Laird & Byrne, 1991). Despite their divergences, these approaches shared common conceptions about both the reasoning processes that deserve investigation, and the normative theory to which human reasoning should be compared. Theories rooted in this tradition focused on the processes of deduction and truth preservation based on binary distinctions between truth and falsity or validity and invalidity. How logical thinking is possible in humans and how far people conform to logical standards were questions of main interest, with the key discovery that human reasoning is prone to biases and often relies on heuristics instead of analytic thinking (Evans, 1982; Kahneman, 2003, 2011; Kahneman & Tversky, 1972, 1973).

* Corresponding author at: Université de Genève, Faculté de Psychologie et des Sciences de l'Education, 40 bd du Pont d'Arve, 1211 Genève 4, Switzerland. Tel.: +41 22 379 92 51.

E-mail address: Pierre.Barrouillet@unige.ch (P. Barrouillet).

http://dx.doi.org/10.1016/j.cognition.2014.12.002

These divergences between the actual human reasoning performance and formal logic progressively turned the attention of psychologists to notions like degrees of belief, subjective probability, and utility (see Oaksford & Chater, 1994, for one of the first examples of this turn). This resulted in the emergence of new theories in keeping with the Bayesian movement that has recently developed in psychology and neurosciences. These theories that Elqayam and Over (2013) have suggested to widely name *probabilistic* are assumed by their defenders to constitute a *new paradigm* that goes far beyond the mere study of deductive processes (Chater & Oaksford, 2008; Evans, 2002, 2012; Oaksford & Chater, 2007, 2009). Accordingly, Elqayam and Over (2013, p. 259) do not hesitate to state that "studying probability judgments will tell us more about the psychology of reasoning than trying to figure out how far people conform to binary extensional logic in any deductive reasoning in which they engage".

However, though the new paradigm has certainly enriched the range of problems addressed by the psychology of reasoning, it has left almost unexplored key questions that were considered as central by the deduction paradigm. This is the case of development. Our remarkable reasoning capacities, including the ability to reason abstractly, are usually seen as a distinctive characteristic of human beings, and as such the question of their origin is central for understanding human cognition and, more generally, our human nature. Accordingly, the innate, learned, or constructed nature of our capacity to reason was one of the main questions debated by what Elqayam and Over (2013) call the "old" paradigm (Inhelder & Piaget, 1958; Overton, 1990). Notwithstanding its age, the questions that were debated by this "old" paradigm are not necessarily obsolete. If probability judgments are the basis of human reasoning as the new paradigm claims, how these judgments evolve with age becomes a major issue for psychology. The aim of the present study was to address this question in the domain of conditional reasoning.

1.1. The new paradigm and the question of development

One of the main innovations of the new paradigm is undoubtedly the renewal it has introduced in the study of conditional reasoning (Oaksford & Chater, 2010). Conditional reasoning is a key process of human mind. Permitted by propositions containing the connector "*If*", it allows human beings to think about hypotheses and suppositions. Accordingly, it underpins scientific reasoning (Kuhn, 2011), but also our capacity to think about causal relations (Kushsnir & Gopnik, 2007), to comply with social rules (Harris & Nunez, 1996; Light, Blaye, Gilly, & Girotto, 1989), to understand inducements such as promises and threats (Newstead, Ellis, Evans, & Dennis, 1997), and even to think in a counterfactual way ("*If only I had* ...", Beck & Riggs, 2014). Thus, it is not surprising that the origins and development of conditional reasoning were a matter of debate for the "old" paradigm. Whereas some authors assumed that the basic logical rules governing conditional reasoning such as Modus Ponens were based on mapping the *If* of the natural language to some innate concept of contingency (Braine, 1990), others assumed that these rules could be acquired through learning processes (Falmagne, 1990). Subsequent views of the development of conditional reasoning have assumed that it is provoked by an age-related increase in world knowledge and working memory capacities permitting the construction of richer mental model representations with age (Barrouillet & Lecas, 1998, 1999; Markovits & Barrouillet, 2002). More recently, the role of different levels of divergent thinking has been emphasized in the evocation by children and adolescents of the different possibilities compatible with conditionals involving different contents (Markovits, 2014).

Departing from the formerly prevailing extensional approach, the new paradigm assumes that the meaning of a natural language conditional *If p then q* is probabilistic in nature. More precisely, it is assumed that the probability of the conditional is the conditional probability, $P(q|p)$, in such a way that $P(\textit{if p then q}) = P(q|p)$. This formal position is so important for the new paradigm theoreticians that, following Edington (1995), they call it *the Equation*. Empirical evidence supporting this proposal is based on a probabilistic truth table task (hereafter, the *probability task*) in which participants are asked to assess the probability of an *If p then q* conditional from the probabilities of the four truth-table possibilities *p* & *q*, *p* & *not-q*, *not-p* & *q*, *not-p* & *not-q*. Several studies observed that a majority of adults judge the probability of the conditional as the conditional probability $P(q|p)$, that is the probability of *p* & *q* divided by the summed probabilities of *p* & *q* and *p* & *not-q* (Evans, Handley, & Over, 2003; Fugard, Pfeifer, Mayerfofer, & Kleiter, 2011; Oberauer & Wilhelm, 2003; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007).

However, contrary to the traditional psychology of reasoning, new paradigm theoreticians seem unconcerned by development. For example, there is no mention of developmental questions in the recent special issue of the journal *Thinking & Reasoning* devoted to basic and applied perspectives for the new paradigm psychology of reasoning (Elqayam, Bonnefon, & Over, 2013). In the same way, although dual-process theories of reasoning have been successfully used to account for the development of reasoning (Barrouillet, 2011; Brainerd & Reyna, 2001; Gauffroy & Barrouillet, 2009; Klaczynski & Cottrell, 2004; Klaczynski & Felmban, 2014; Markovits, 2014; Reyna & Brainerd, 2011), Evans and Stanovich (2013) do not even evoke developmental issues in their last review on dual-process theories. Thus, the new paradigm remains silent about the question of the origins of *the Equation* and its possible evolution with age. In the following, we will try to derive developmental predictions from one of the most coherent theoretical frameworks pertaining to the new paradigm, namely Evans' (2007) suppositional theory of conditional along with the heuristic-analytic approach (Evans, 2006) as its algorithmic counterpart. These predictions will be compared to those that can be derived from our own revised mental model theory of conditional concerning its probability (Barrouillet, Gauffroy, & Lecas, 2008; Gauffroy & Barrouillet, 2009, 2014b).

1.2. The suppositional conditional and its putative development

According to Evans (2007), the conditional is suppositional in nature and based on a mental procedure that philosophers call the *Ramsey test*, by which when assessing an *If p then q* conditional, reasoners hypothetically add *p* to their stock of knowledge and argue on that basis about *q* (Evans, 2007; Evans & Over, 2004; Evans, Over, & Handley, 2005). This Ramsey test acts as an *if-heuristic* as it was described by Evans (1989). Directing automatically attention toward cases where the antecedent is true (i.e., *p* cases), it leads to disregard *not-p* cases as irrelevant for assessing the truth of the conditional. For example, for evaluating the conditional "If Queen Elisabeth dies, then Prince Charles will become King", people would perform some thought experiment and imagine a world in which Queen Elisabeth is dead, evaluating the likelihood of Prince Charles becoming King. This would result in a model representing the conditional relation between the antecedent and the consequent including the degree of belief in this relation:

Queen Elisabeth dies → (.95) Prince Charles becomes King

The value of .95 in the diagram indicates a high degree of belief, the maximum value of 1 corresponding to certainty. This would explain why, when asked to evaluate the probability of the conditional, people produce the conditional probability. Because the Ramsey test focuses on *p* cases, the probability of the conditional to be true is the proportion of cases that make it true (i.e., *p* & *q* cases) among those that are relevant for truth evaluation (i.e., *p* & *q* and *p* & *not-q* cases).

It is worth to note that the Ramsey test, conceived as an *if-heuristic*, pertains to the Type 1 processes described by the dual-process theories of reasoning (Evans & Frankish, 2009). These processes that are unconscious, automatic, rapid, pragmatically cued and not related to individual differences in general intelligence, are contrasted with Type 2 processes that are analytic, controlled and slow, a distinction that maps the contrast described by Evans (2006) between heuristic and analytic processes. If it is assumed, as Evans (2006) does, that the default model described above is produced by heuristic processes, it must be admitted that the processes estimating the ratio between *p* & *q* and *p* & *not-q* cases are also part of the heuristic or Type 1 processes. Indeed, as we have seen, the default model already captures the degree of belief on which the suppositional theory and more generally the probabilistic approach is based. As Barrouillet (2011) noted, this renders developmental predictions uncertain. Type 1 processes are considered as not requiring controlled attention (Evans & Stanovich, 2013), and as such they would remain largely unaffected by age-related changes in working memory capacity. Of course, because they are also based on associative learning, their output should evolve with mundane experience and therefore age, but it is difficult to predict in which way developmental changes in the stock of automatically triggered associations would affect probability judgments. At best, it could be assumed with Evans (2007) that some failure can occur in the Ramsey test. Evans et al. (2003) observed that a substantial minority of participants did not give the conditional but the conjunctive probability, P(*p* & *q*), when assessing the probability of the conditional. They explained this response by supposing an incomplete Ramsey test in shallow processors who would cut short the test and stop at the *p* & *q* cases. This explanation was corroborated by the fact that conjunctive responders are lower in general intelligence than conditional responders (Evans, Handley, Neilens, & Over, 2007). Working memory capacities are known to be highly related with Gf (Conway, Kane, & Engle, 2003). Because these capacities strongly evolve with age, the suppositional approach could at best predict higher rates of conjunctive responses to the probability task in younger participants.

1.3. The probability of the conditional in the revised mental model approach

We have recently suggested to account for the development of conditional by reinterpreting the developmental mental model theories of conditional (Barrouillet & Lecas, 1998; Markovits & Barrouillet, 2002) within a dual-process framework (Gauffroy & Barrouillet, 2009, 2011, 2014a; Vergauwe, Gauffroy, Morsanyi, Dagry, & Barrouillet, 2013). We assume that, when understanding a conditional sentence of the form *If p then q*, Type 1 heuristic processes deliver a default model representing the relation between the antecedent and the consequent (i.e., a mental model of the form *p q*). Because Type 1 processes are not accessible to introspection, their output would be delivered early in the reasoning process and constitute what is called initial model in the mental model theory (Johnson-Laird & Byrne, 1991, 2002). Coming spontaneously to mind, this initial representation would be considered by reasoners as the core meaning of the conditional, representing those cases that make it true when they occur. Accordingly, several developmental studies have shown that the most basic interpretation is conjunctive in nature, young children representing the conditional by the mere co-occurrence of the antecedent and the consequent (Barrouillet, Grosset, & Lecas, 2000; Barrouillet & Lecas, 1998, 2002; Gauffroy & Barrouillet, 2011). This representation is assumed to be the default model for several types of conditionals, like indicative or causal conditionals, but not necessarily for all conditionals. Heuristic processes could also deliver pragmatic implicatures susceptible to enrich the output of Type 1 processes (Evans, 2006; Evans & Over, 2004). This should be the case for promises and threats that strongly cue invited inferences that are inherent to their meaning (Gauffroy & Barrouillet, 2009; Newstead et al., 1997). More will be said about promises and threats later in Experiment 3.

Within this framework, Type 2 processes could intervene to enrich the initial representation with alternative possibilities. This process, named fleshing out in the mental model theory (Johnson-Laird & Byrne, 1991), involves working memory and strongly evolves with age (Barrouillet & Lecas, 1999). Developmental studies have shown that the first alternative model constructed is of the form $\neg p \neg q$, resulting in a biconditional interpretation, whereas the construction of the model $\neg p\ q$ leading to a complete conditional understanding requires additional processes and occurs later in development. This developmental trend from a conjunctive to a biconditional and then a conditional interpretation has been found in tasks involving the identification of cases that are either compatible or compatible with the conditional and the production of inferences (Barrouillet & Lecas, 1998, 1999, 2002; Barrouillet et al., 2000; Lecas & Barrouillet, 1999).

Importantly, our theory departs from the traditional mental model approach by assuming different epistemological statuses for the different types of models constructed (Barrouillet et al., 2008). Whereas the conditional is deemed true for those cases that match the initial model that comes spontaneously to mind, the alternative models constructed through fleshing out have a different status. Being optional in nature, they are not part of the core meaning of the conditional sentence, but just represent possibilities that are compatible with it. As a consequence, when people are asked to list the possibilities compatible with the conditional, they produce responses that match the initial and the fleshed out models, but when asked to assess the truth value of the conditional, those cases that match the models constructed through fleshing out are considered as leaving this truth value indeterminate (i.e., *not-p* cases for a majority of conditionals). Finally, the cases that do not correspond to any of the models constructed are considered as falsifying the conditional.

Interestingly, the hypothesis of different epistemological statuses for initial and additional models leads to unique developmental predictions when considering the truth-table task in which participants are asked to judge the truth-value of a conditional from different cases. Young children who endorse a conjunctive interpretation and construct the sole $p\ q$ model should judge the conditional true for p & q cases and false for all the other cases. Adolescents who add the $\neg p \neg q$ model through fleshing out should produce a defective biconditional response in which the conditional is true for p & q, false for p & *not-q* and *not-p* & q cases, its truth value remaining indeterminate for the sole *not-p* & *not-q* cases. Finally, older adolescents and adults who construct a three-model representation by adding two models (i.e., $\neg p \neg q$, and $\neg p\ q$) should deem the conditional true for p & q cases, false for p & *not-q* cases, *not-p* cases being judged as irrelevant, a response pattern corresponding to the well-known defective[1] or De Finetti truth-table of the conditional. These predictions have been verified in several studies (Barrouillet et al., 2008; Gauffroy & Barrouillet, 2009, 2011).

In the same way, our theory makes straightforward predictions concerning the probability task. We assume that people assess the probability that a conditional is true (or false) for a case drawn at random from a given set by estimating the ratio between those cases that make the conditional true (or false) and those that they consider as relevant for its truth-value (i.e., those that they consider as making the conditional either true or false), leaving aside those cases they consider as irrelevant (i.e., those that match models constructed through fleshing out). Children who endorse a conjunctive interpretation should judge the probability of the conditional to be true as the probability of the conjunction:

$$P_{\text{Conj. True}} = P(p \ \& \ q) \tag{1}$$

Because $p\ q$ is the sole model they construct (i.e., there is no fleshing out), all the other possibilities make the conditional false. Consequently, for conjunctive responders, the probability of the conditional to be false should be:

$$P_{\text{Conj. False}} = 1 - P(p \ \& \ q) \tag{2}$$

Things are more complex for adolescents who endorse a biconditional reading and add a $\neg p \neg q$ model to their representation. Cases that match this model constructed through fleshing out should be considered as irrelevant for the truth value of the conditional, while p & *not-q* as well as *not-p* & q cases should be considered as making it false because they do not match any of the models constructed. As a consequence, the probability for the conditional to be true should correspond to a defective biconditional probability:

$$P_{\text{Def. Bicond. True}} = P(p \ \& \ q)/[P(p \ \& \ q) + P(p \ \& \ \textit{not-q}) + P(\textit{not-p} \ \& \ q)] \tag{3}$$

whereas the probability for the conditional to be false should be:

$$P_{\text{Def. Bicond. False}} = 1 - P(p \ \& \ q)/[P(p \ \& \ q) + P(p \ \& \ \textit{not-q}) + P(\textit{not-p} \ \& \ q)] \tag{4}$$

Finally, for those individuals who are able to construct a complete conditional representation by adding the $\neg p \neg q$ and the $\neg p\ q$ models through fleshing out, and for whom *not-p* & *not-q* and *not-p* & q cases are consequently irrelevant, the probability for a true conditional should be the conditional probability:

$$P_{\text{Def. Cond. True}} = P(p \ \& \ q)/[P(p \ \& \ q) + P(p \ \& \ \textit{not-q})] \tag{5}$$

whereas the probability for a false conditional should be:

$$P_{\text{Def. Cond. False}} = P(p \ \& \ \textit{not-q})/[P(p \ \& \ q) + P(p \ \& \ \textit{not-q})] \tag{6}$$

These latter responses correspond to *the Equation*, reflecting a defective conditional reading.

1.4. The present study

The aim of the present study was to investigate the developmental origins of *the Equation*, the tendency in adults to assess the probability of the conditional as the conditional probability. Can *the Equation* be considered as the core meaning of the conditional understood as conveying a basic and inherent probabilistic meaning, in which case it should exhibit some pervasiveness through development, types of conditional and contents? By now, our

[1] This truth-table is called defective because it leaves the truth-value of the conditional indeterminate for two logical cases ($\neg p \neg q$ and $\neg p\ q$), whereas the conditional is deemed true for these two cases in the truth-table of formal logic, resulting in an interpretation known as *material implication*.

first and sole attempt in this direction suggests that young adolescents, adolescents and adults conform to the conjunctive, defective biconditional and defective conditional responses described above when assessing the probability of the conditional (Gauffroy & Barrouillet, 2009). However, in this previous study, we only investigated basic conditionals[2] without any variation in the type of conditionals or their content. Here, apart from basic conditionals, we will also consider causal conditionals and inducements. Moreover, these conditionals will be manipulated by introducing contents that are known to affect their interpretation in determinate ways.

2. Experiment 1: Probability of basic conditionals

As we mentioned above, Gauffroy and Barrouillet (2009) already investigated the developmental changes in the evaluation of the probability of basic conditionals using the task introduced by Evans et al. (2003) and Oberauer and Wilhelm (2003) in which participants are given frequency information about the cases *p* & *q*, *p* & *not-q*, *not-p* & *q*, and *not-p* & *not-q* from which they are asked to evaluate the probability of an *If p then q* statement. For example, the four types of cases are represented by cards that are either yellow or red with either a circle or a diamond printed on them (say 1 yellow circle, 4 yellow diamonds, 16 red circles, and 16 red diamonds), the participants being asked to evaluate the probability that a claim like "If the card is yellow then it has a circle printed on it" is true for a card drawn at random from the pack. We adapted this task by presenting smaller sets of cards, their total number varying from 6 to 9 and the number of cards of each type varying from 1 to 3. To make the task more understandable by adolescents and to avoid complex calculations, participants were given instructions about what the term probability means and were invited to fill sentences of the following form with two numbers "the probability that the statement is true (or false) is ... out of ...". This experiment revealed a developmental trend from a majority of conjunctive responses in sixth graders to defective conditional responses (i.e., the conditional probability) in adults with defective biconditional responses as an intermediate level in ninth graders.

The following experiment aimed at replicating and extending these findings. For this purpose, we used the evaluation probability task described above with basic conditionals the content of which was manipulated. We took advantage of a phenomenon studied by Barrouillet and Lecas (1998, 2002) who observed that conditionals with binary terms in both the antecedent and the consequent strongly elicit biconditional readings. Binary terms are terms that allow for a single alternative such as in the conditional "If the bird is male, then it has a dark plumage" in which "male" and "dark" admit a single alternative (female and light, respectively). These conditionals with binary terms (referred to as BB conditionals) are contrasted with NN conditionals (N for non-binary) like "If the car is red than it is a Ford". We have already observed that the biconditional reading induced by BB conditionals affects the production of cases compatible with the conditional (Barrouillet & Lecas, 1998), the production of inferences (Barrouillet & Lecas, 2002), but also the responses in truth table tasks (Gauffroy & Barrouillet, 2009, 2014a). In these latter tasks, whereas NN conditionals elicit the conjunctive – defective biconditional – defective conditional developmental trend, BB conditionals elicit a majority of defective biconditional responses in adolescents but also in adults. Following our example, a male black bird is judged as making the conditional true, a male white bird and a female black bird as making it false, but a female white bird leaves its truth value indeterminate. We suggested that this interpretation results from an incomplete fleshing out limited to the $\neg p \neg q$ model. Because both the antecedent and the consequent only admit a single alternative, the construction of the $\neg p \neg q$ model leads to what Barrouillet and Lecas (1998) called a *complete* representation that establishes a one to one correspondence between the possible values of the antecedent and the consequent, blocking the fleshing out process at this stage and impeding the production of the $\neg p\ q$ model. The hypothesis that the biconditional interpretation associated with BB conditionals results from an incomplete fleshing out is corroborated by the fact that young children who do not flesh out the initial model remain unaffected by the BB nature of the conditionals and still produce conjunctive patterns in a truth-table task (Gauffroy & Barrouillet, 2009).

If we are correct in assuming that people evaluate the probability of the conditional to be true by calculating the ratio between those cases that make the conditional true (i.e., that match the initial model produced by Type 1 processes) and those that are considered as relevant for its truth value (thus excluding the cases that match fleshed out models), then BB and NN conditionals should elicit different responses in the evaluation of probability task. Because truth-table tasks have revealed a strong tendency in adolescents and adults to produce defective biconditional patterns for BB conditionals (Gauffroy & Barrouillet, 2009, 2014a), both adolescents and adults should produce probability evaluations that conform to this truth-table and not *the Equation*, this latter response being limited to NN conditionals in adults. For sake of comparison with previous studies by Gauffroy and Barrouillet (2009, 2011), this and the other experiments of our study involved the same age groups (i.e., 3rd, 6th, 9th graders, and adults), except that the probability task proved too difficult for 3rd graders.

2.1. Method

2.1.1. Participants

Thirty sixth graders (mean age = 11.9 years, SD = 0.4, 18 females), 28 ninth graders (mean age = 15.8 years, SD = 0.7, 16 females), and 30 students from the University of Geneva (mean age = 26.2, SD = 1.55, 17 females) performed a probability task on both BB and NN conditionals. Sixth and ninth graders participated as volunteers and students for partial fulfilment of a course requirement.

[2] According to Johnson-Laird and Byrne (2002), basic conditionals are conditionals in which the antecedent and the consequent have no semantic or referential relations, or relations based on knowledge.

2.1.2. Material and procedure

The conditionals used in this experiment were the same as those used by Gauffroy and Barrouillet (2009) with four NN and four BB conditionals (see Appendix A). The probability task was administered using a video projector. In this and the following experiments, the task was administered by class groups for adolescents and groups of about 15 participants for adults. A short scenario was displayed at the top of the screen and introduced the conditional statement as a judgment produced by a mother. For example, in one of these scenarios, after having met pupils in a classroom, a mother claims: "If the pupil is a boy, then he wears glasses". Different drawings representing the four logical cases were presented under the conditional statement. In our example, a set of six drawings was presented: one boy with glasses (i.e., *p* & *q*), one boy without glasses (i.e., *p* & *not-q*), three girls with glasses (i.e., *not-p* & *q*), and one girl without glasses (i.e., *not-p* & *not-q*). The experimenter described aloud each set of drawings. Across the conditionals, the total number of drawings varied from 6 to 9, and the number of drawings of each type varied from 1 to 3 in such a way that two different interpretations could not give rise to the same response. Below the drawings was displayed a question of the form: "how likely is it that what the mother claims is true (false) for a drawing taken at random from the set?" (Fig. 1). For the four BB conditionals and the four NN conditionals, half of the problems were presented with the "true" question and the other half with the "false" question. We created two versions of the task with "true" problems in one version becoming "false" problems in the other. The eight problems were presented in random order as well as the four types of drawings within each problem. Participants were asked to evaluate the probability that what the mother claims is true or false by filling sentences of the following form with two numbers: "the probability that what the mother claims is true (or false) is ... out of ...". The experimenter controlled when the next trial was initiated. This procedure ensured that all the participants had enough time to answer.

2.2. Results

2.2.1. Analysis of responses

We categorized the responses of each participant according to the expected interpretations described by Eqs. (1)–(6). We collapsed the data from "true" and "false" problems because there was no reliable difference and performed a 3 (age group: sixth graders, ninth graders and adults) $\times$ 2 (conditionals: BB and NN) analysis of variance on the number of conjunctive, defective biconditional and defective conditional responses.

As we expected concerning conjunctive responses, their rate was lower for BB than for NN conditionals, $F(1,85) = 34.66$, $p < .001$, $\eta_p^2 = .29$. In line with our hypotheses, conjunctive responses on NN conditionals were predominant in younger and rarer in older participants (80%, 35%, and 37% in 6th, 9th graders and adults, respectively, Fig. 2), $F(2,85) = 8.86$, $p < .001$, $\eta_p^2 = .17$, with a levelling between the two older groups. Although the rate of conjunctive responses in older adolescents and adults is higher than in our previous research, it is in line with most of the studies that used the probability task (see Manktelow, 2012, for a review). By contrast, the rate of conjunctive responses did not vary significantly with age for BB conditionals (33%, 21% and, 17% respectively), $F(2,85) = 1.20$,

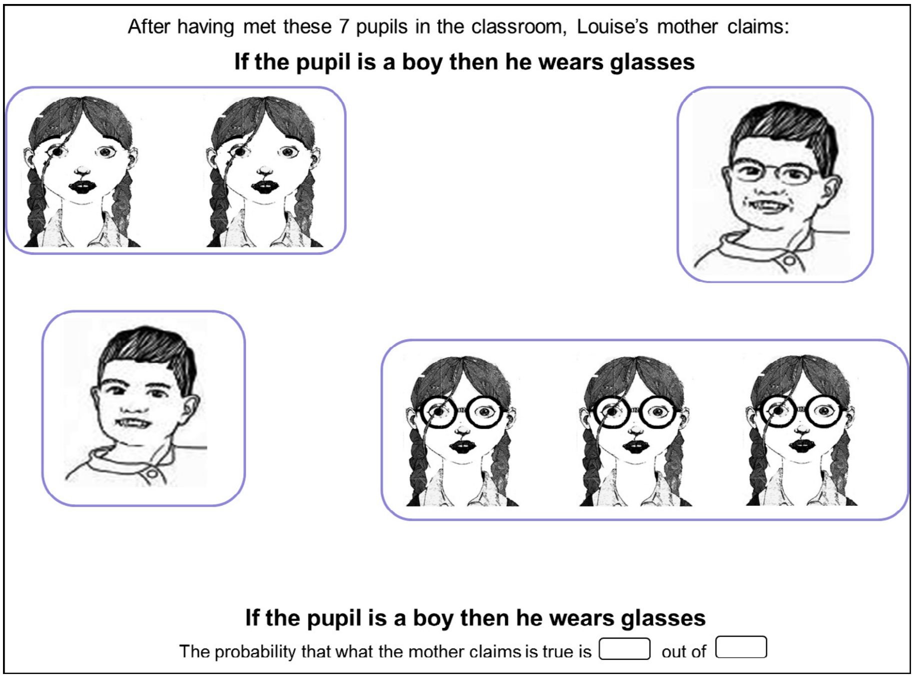

Fig. 1. Example of material used in the probability task involving basic conditionals.

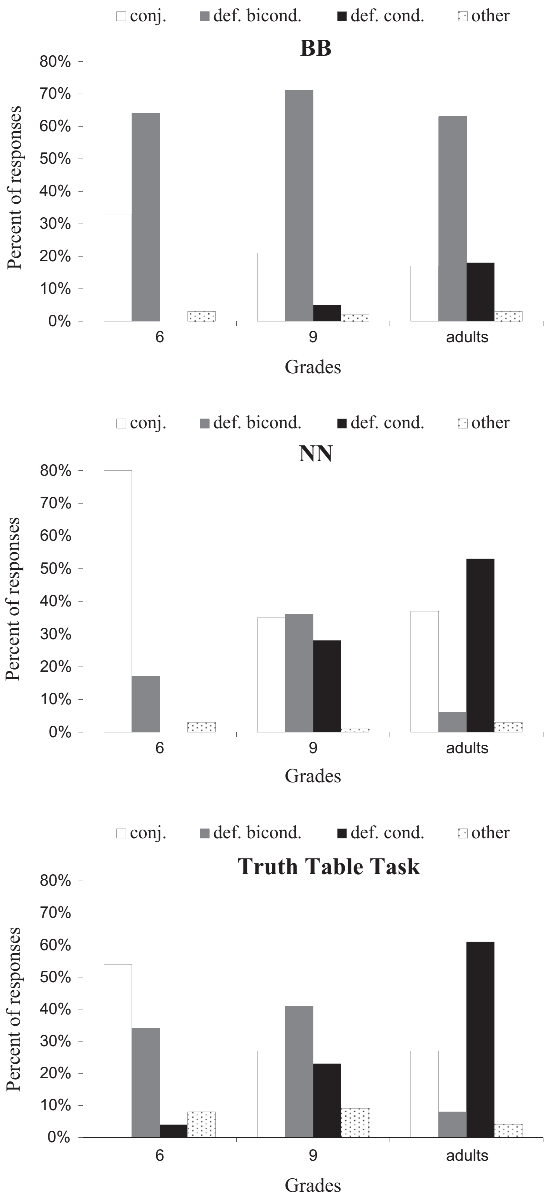

Fig. 2. Percent of responses categorized as conjunctive (Conj.), defective biconditional (Def. Bicond.), and defective conditional (Def. Cond.) in the probability task involving BB and NN basic conditionals. The truth table task panel reports the percent of responses categorized as conjunctive, biconditional and conditional in the truth-table task studied by Gauffroy and Barrouillet (2011).

$p = .30$, $\eta_p^2 = .03$, resulting in a significant interaction between age and type of conditional, $F(2,85) = 4.76$, $p < .05$, $\eta_p^2 = .10$.

Our main prediction was that binary terms should affect the construction through fleshing out of the $\neg\, p\ q$ model leading to predominant defective biconditional interpretations. Accordingly, the rate of defective biconditional responses was higher for BB than for NN conditionals (66% and 20% respectively), $F(1,85) = 75.50$, $p < .001$, $\eta_p^2 = .47$, with no effect of grades, $F(2,85) = 2.10$, $p = .13$, $\eta_p^2 = .05$, and no interaction, $F(2,85) = 1.27$, $p < .29$, $\eta_p^2 = .03$. The defective biconditional response was the predominant response for 6th graders, 9th graders and adults for BB conditionals (64%, 71%, 63%, respectively), whereas these responses represented an intermediate level for NN conditionals (17%, 36%, and 6%, respectively), $F(1,85) = 7.53$, $p < .01$, $\eta_p^2 = .08$, for the quadratic trend.

Finally, while there were no defective conditional responses in 6th graders, their rate increased to 28% in ninth graders for NN conditionals and they became predominant in adults (53%). In accordance with the idea that binary terms block the fleshing out process, this interpretation was rarer with BB conditionals (0%, 5% and 18% in 6th graders, 9th graders and adults, respectively), $F(1,85) = 24.83$, $p < .001$, $\eta_p^2 = .23$. Thus, the main effect of age, $F(2,85) = 12.60$, $p < .001$, $\eta_p^2 = .23$, significantly interacted with the type of conditional, $F(2,85) = 7.48$, $p < .01$, $\eta_p^2 = .15$.

2.2.2. Response patterns analysis

Participants were considered as being consistent on a given interpretation when at least 3 out of their 4 responses corresponded to this interpretation. This classification revealed very clear results. Concerning NN conditionals, 80% of the 6th graders conformed to a conjunctive interpretation and 15% to a biconditional interpretation. All the 9th graders were consistent in their responses, 36% giving conjunctive responses, 36% defective biconditional responses and 28% defective conditional responses. Finally, the main response in adults was the defective conditional, 53% of participants favouring it and 7% exhibiting a defective biconditional interpretation. Note that, as in 9th graders, 37% of adults still gave conjunctive responses.

As BB conditionals were concerned, for all grades, the results revealed that the main interpretation was the defective biconditional. In 6th graders, 29 out of 30 participants were consistent, 34% giving conjunctive responses and 66% defective biconditional responses. In 9th graders, 27 out of 28 participants were consistent, with 22% of the participants producing conjunctive responses, 74% producing defective biconditional responses and 4% giving the conditional probability. Finally, 29 out of the 30 adults were consistent, 17% giving the conjunctive probability, 66% the defective biconditional probability and 17% the conditional probability.

2.3. Discussion

Two main findings arose from this experiment. First, as Gauffroy and Barrouillet (2009) already observed, the evaluation of the probability of the conditional as the conditional probability is a late developmental achievement characteristic of adulthood that is preceded by two developmental levels. Because the task of evaluating the probability of the conditional requires assessing the truth-value of the conditional, it does not come as a surprise that the responses in the probability task reflect the responses usually observed in the traditional truth-table task. We hypothesized that people use a strategy by which they evaluate the probability of the conditional to be true (or false) as the ratio between the number of cases that make the conditional true (or false) and the number of cases for which the conditional has a truth-value, either true or false. Our results confirm this hypothesis, responses in the probability task evolving with the judgments of irrelevance in the truth-table task. Whereas the number of cases considered as making the conditional true does not evolve with age, the number of cases considered as relevant progressively decreases (Gauffroy & Barrouillet, 2011). Whereas all the cases are relevant for the younger participants, *not-p* & *not-q* cases and then all the *not-p* cases are progressively disregarded as participants get older. The hypothesis of probability evaluations reflecting truth-table patterns is confirmed by the results concerning BB conditionals. In the same way as they elicit defective biconditional patterns in the truth-table task (Gauffroy & Barrouillet, 2009), BB conditionals gave rise to defective biconditional responses in the probability task in both adolescents and adults. These results suggest that it is not *the Equation* that drives the judgments of irrelevance of the *not-p* cases and the resulting defective truth table so often observed in truth-table tasks, but the reverse. The evaluation of the probability of the conditional is determined by truth-value judgments. Accordingly, this probability evaluation evolves developmentally in close connection with the judgments of irrelevance in the truth-table tasks, from conjunctive to defective biconditional and defective conditionals, except when the content of the conditional constraints the fleshing out process that leads to these judgments of irrelevance. In other words, it seems that it is the developmental evolution in the irrelevance judgments that finally underpins *the Equation*, rather than *the Equation* and the associated Ramsey test that would cause irrelevance judgments.

It could be argued that there is at least one exception to the close relationship between truth-table judgments and probability evaluations. A substantial minority of adults produce conjunctive probabilities on NN conditionals whereas conjunctive patterns in the truth table task almost never occur at these ages (see for example Barrouillet et al., 2008, or Gauffroy & Barrouillet, 2009). However, the occurrence of conjunctive response patterns in truth table tasks depends on the number of alternative responses available in these tasks. These patterns almost never occur in adults when they are given the three-value choice task with "true", "false", and "indeterminate" as alternative responses. However, we observed that when restricted to a forced choice between "true" and "false", a substantial minority of adults produce conjunctive response patterns in the truth-table task, deeming the conditional false for *not-p* cases (Gauffroy & Barrouillet, 2011). For sake of comparison, we have reported in Fig. 2 the distribution of response patterns observed by Gauffroy and Barrouillet (2011) with such a two-valued truth-table task. As it can be seen, there is a striking parallel with the distribution of responses in the probability task. We will defer the discussion of this phenomenon to the general discussion and provide another test of our hypothesis that truth-value judgments drive the evaluations of probability. Gauffroy and Barrouillet (2009) investigated the developmental trend in truth tables for causal conditionals with either few or many alternative antecedents. The next experiment used the same causal conditionals in the probability evaluation task.

3. Experiment 2: Probability of causal conditionals

Causal conditionals are conditionals in which the antecedent refers to a cause while the consequent refers to its effect as in "If a rock is thrown at a window, the window will break". Several studies have established that the main factor that affects causal conditional interpretation is the number of alternative causes (of a broken window in our example) that people can evoke. Causal conditionals for which people do not have access to alternative causes tend to be interpreted in a biconditional way, whereas several alternatives easily accessed usually induce a conditional interpretation in adults (Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; Markovits, 1984; Quinn & Markovits, 1998). Developmental studies have shown that the tendency to interpret conditionals with few or no alternatives as biconditionals is already observable in early adolescence (Barrouillet, Markovits, & Quinn, 2001; Markovits, Fleury, Quinn, & Venet, 1998). Despite this difference between causal conditionals with few or several alternatives, for both types of conditionals the mental model theory assumes the same initial mental model in which the cause and its effect would co-occur (Goldvarg & Johnson-Laird, 2001). Within this framework, conditional and biconditional interpretations would result from different levels of fleshing out: retrieved alternative causes would ease the construction of a $\neg p\ q$ model leading to the conditional interpretation whereas the fleshing out process would be limited to the $\neg p \neg q$ model when no alternative cause is available, resulting in the biconditional reading. Thus, Gauffroy and Barrouillet (2009) predicted that, in a truth-table task, conditionals with few and many alternatives should differ by the number of responses of irrelevance they elicit in the same way as BB and NN basic conditionals do. Whereas causal conditionals with few alternatives should mainly elicit defective biconditional patterns in both adolescents and adults, causal conditionals with many alternatives should reveal the developmental trend from conjunctive, defective biconditional and defective conditional patterns we described above. This is exactly what Gauffroy and Barrouillet (2009) observed. In the present study, the same causal conditionals as those studied by Gauffroy and Barrouillet (2009) were used in a probability evaluation task with the hypothesis that the response in this task should mirror those observed in the truth-table task. Although the number of alternatives must be distinguished from the strength of the causal relation, the two dimensions being orthogonal, for sake of simplicity and to keep with the denominations used by Gauffroy and Barrouillet (2009), we will refer to causal relations with either few or many alternatives as *strong* and *weak* causal relations respectively.

3.1. Method

3.1.1. Participants

Twenty-seven sixth graders (mean age = 11.4 years, SD = 0.6, 15 females), 27 ninth graders (mean age = 15.2 years, SD = 0.5, 14 females), and 28 students from the University of Geneva (mean age = 24.5, SD = 1.42, 18 females) performed a probability task with strong and weak causal relations. Sixth and ninth graders participated as volunteers and students for partial fulfilment of a course requirement. None of them took part in the previous experiment.

3.1.2. Material and procedure

The conditionals were the same as those used by Gauffroy and Barrouillet (2009, see Appendix A). The procedure and the instructions given to participants were the same as in the previous experiment. For example, for the strong causal conditional "If the lever 2 is down, then the rabbit's cage is open" a set of 7 drawings were presented: one with the lever 2 down and the cage open (i.e., *p* & *q*), one with lever 2 down and the cage close (i.e., *p* & *not-q*), two with the lever 2 up and the cage close (i.e., *not-p* & *not-q*) and three with the lever 2 up and the cage open (i.e., *not-p* & *q*). Participants were asked to evaluate how likely it is that what the mother claims is true (or false) for a drawing taken at random from the set of drawings. Four weak causal relations and four strong causal relations were presented with half "true" problems and half "false" problems. Participants estimated the probability by filling a sentence with two numbers as in Experiment 1. "True" problems in the first version of the task became "false" problems in the second version and conversely.

3.2. Results

3.2.1. Analysis of responses

The percentages of conjunctive, defective biconditional, and defective conditional responses were the same for both "true" and "false" problems. So, the analyses combined responses from both conditions. As we predicted, and as for NN conditionals, the rate of conjunctive responses decreased with age for weak causal relations (67%, 26%, and 29% for 6th graders, 9th graders and adults, respectively), $F(2,79) = 6.59$, $p < .01$, $\eta_p^2 = .14$, with difference reaching significance only between grades 6 and 9, $t(52) = 3.23$, $p < .01$. The rate of conjunctive responses was higher for weak than strong causal relations (41% vs. 18%), $F(1,79) = 23.97$, $p < .001$, $\eta_p^2 = .23$ (Fig. 3) and was not affected by age for strong relations (30%, 11%, and 14%), $F < 1$. Nonetheless, the significant main effect of grade, $F(2,79) = 5.18$, $p < .01$, $\eta_p^2 = .12$, did not significantly interact with the effect of type of conditional, $F(2,79) = 2.76$, $p = .06$, $\eta_p^2 = .07$.

As we predicted, strong causal relations induced a predominance of defective biconditional responses from 6th graders to adults (63%, 78% and, 54% in 6th graders, 9th graders and adults, respectively). On the contrary, this response constituted an intermediate level of interpretation for weak relations (26%, 43%, and 11%), $F(1,79) = 5.92$, $p < .05$, $\eta_p^2 = .07$, for the quadratic trend. Thus, in accordance with the hypothesis that strong causal relations, like binary terms, should affect the fleshing out process, the rate of defective biconditional responses was higher for strong than weak causal relations, $F(1,79) = 47.50$, $p < .001$, $\eta_p^2 = .37$.

Finally, the rate of defective conditional responses was lower for strong than weak causal relations $F(1,79) = 17.88$, $p < .001$, $\eta_p^2 = .18$. For weak relations, this response appeared on grade 9 and became predominant in adults (30% and 61%, respectively), $F(2,79) = 16.27$, $p < .001$, $\eta_p^2 = .29$, for the main effect of grade, whereas adults gave 32% of defective conditional responses for strong relations, the main effect of grade, $F(2,79) = 15.77$, $p < .001$, $\eta_p^2 = .18$, significantly interacting with the type of conditional, $F(2,79) = 4.66$, $p < .05$, $\eta_p^2 = .11$.

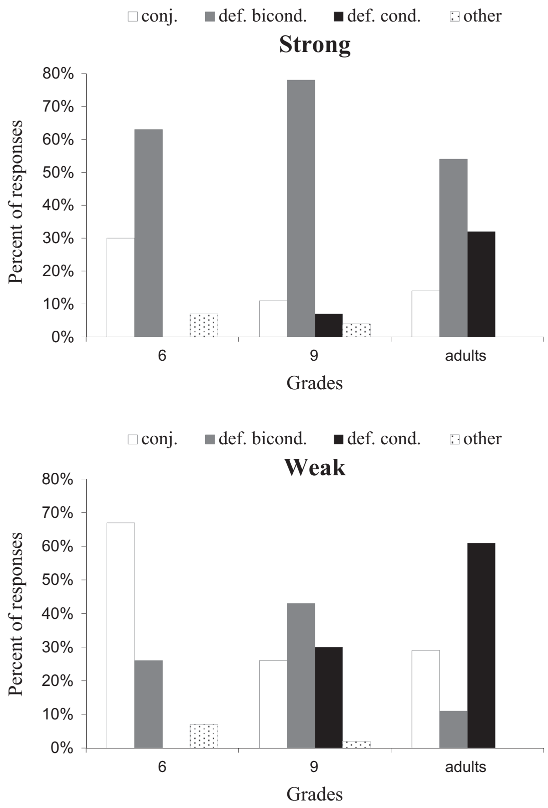

Fig. 3. Percent of responses categorized as conjunctive (Conj.), defective biconditional (Def. Bicond.), and defective conditional (Def. Cond.) in the probability task involving strong and weak causal conditionals.

3.2.2. Response patterns analysis

The same criterion for consistency as in the previous experiment was used (3 out of 4 responses of the same type). Only three participants (2 sixth graders and 1 ninth grader) were not consistent in their responses to both weak and strong causal conditionals. The results with weak causal relations revealed the standard developmental trend. In 6th grade, 67% of participants were conjunctive responders while 26% adopted a defective biconditional interpretation. This latter response represented an intermediate level with 41% of 9th graders who adopted it, 26% giving a conjunctive response, and 30% being consistent on defective conditional responses. The rate of defective conditional responders increased up to 61% in adults whereas a substantial minority of 29% produced conjunctive responses, defective biconditional responders being rarer (11%).

In line with our hypotheses, strong causal relations abolished developmental differences. In all age groups, a majority of participants were defective biconditional responders (63%, 78%, and 54% in 6th graders, 9th graders, and adults, respectively). Conjunctive responders were rarer (30%, 11%, and 14%, respectively), while defective conditional responders only appeared in 9th grade (7%), but constituted a substantial minority in adults (32%).

3.3. Discussion

The results of this experiment revealed that the evaluation of causal conditionals varies with the availability of alternative causes in the same way as the evaluation of basic conditionals varies with the availability of alternative values on the antecedent and the consequent. The probability of causal conditionals with few alternative causes is evaluated in the same way as BB conditionals, whereas the probability of causal conditionals with many alternatives evolves with age as NN basic conditionals. As it was observed for basic conditionals, the evaluation of the probability of causal conditionals mirrors the judgments of their truth-value reported by Gauffroy and Barrouillet (2009) when using a truth-table task. There is a developmental trend from conjunctive, to defective biconditional and then defective conditional responses for causal conditionals with many alternative causes, and a predominant defective biconditional response for those that admit few alternative causes.

The present results extend to another category of conditionals our hypothesis that participants evaluate the probability of conditionals as the ratio between those cases that make the conditional true and those that are considered as relevant for its truth value. However, a further test is needed. Indeed, the different responses we observed in the two previous experiments only vary by the range of cases considered as relevant for this evaluation, whereas the cases seen as making the conditional true remain unchanged through ages and conditionals, always corresponding to *p* & *q* cases. It could be then concluded that evaluating the probability of conditionals is a matter of identifying irrelevant cases. However, Gauffroy and Barrouillet (2009) observed that there are conditionals for which the cases making them true go beyond the sole *p* & *q* contingencies. This is the case for inducements like promises and threats that elicit an equivalence reading in both adolescents and adults. When presented with a promise like "If you mow the lawn, I will give you 5€", adolescents and adults judged the promise to be true for a child receiving 5€ after having mown the lawn (i.e., a *p* & *q* case), but also for a child receiving nothing when it did not mow the lawn (i.e., a *not-p* & *not-q* case), the two other cases being judged as falsifying the conditional. If our hypothesis concerning the evaluation of the probability of conditionals is correct, the probability of promises and threats should reflect this equivalence reading and correspond in both adolescents and adults to:

$$P_{\text{Equiv. True}} = P(p \,\&\, q) + P(not\text{-}p \,\&\, not\text{-}q) \qquad (7)$$

The next experiment tested this hypothesis.

4. Experiment 3: Probability of inducements

As we explained above, Gauffroy and Barrouillet (2009) surmised that promises and threats should elicit implicatures susceptible to enrich the output of Type 1 processes because these implicatures are part of the core meaning of inducements. Understanding a promise like "if you mow the lawn, then I'll give you 5€" requires understanding that the aim of the locutor is not to give money, but to have her lawn mown and that, consequently, nothing will be given if this service is not done. Thus, the $\neg\, p \neg\, q$ model should be part of the core meaning of the promise. In the same way, the aim of a mother saying "If you break the vase, then I'll take your ball away" is to have her vase not broken, the $\neg\, p \neg\, q$ model being an integral part of the core meaning of the sentence. These two models would block any fleshing out, the $p\ q$ model excluding the $p \neg\, q$ possibilities whereas the $\neg\, p \neg\, q$ model excludes $\neg\, p\ q$ possibilities. Accordingly, Gauffroy and Barrouillet (2009) observed that *p* & *q*, but also *not-p* & *not-q* cases were considered as making promises and threats true by children, adolescents, and adults, whereas *not-p* & *q* and *p* & *not-q* cases were considered as making these inducements false, reflecting an equivalence interpretation. We expected the evaluation of the probability of these inducements to conform to this equivalence interpretation in all the age groups tested.

4.1. Method

Twenty-four sixth graders (mean age = 12.2 years, SD = 0.5, 12 females), 26 ninth graders (mean age = 14.9 years, SD = 0.6, 13 females), and 30 students from the University of Geneva (mean age = 22.7, SD = 2.10, 21 females) performed a probability task on the same four promises and four threats already used by Gauffroy and Barrouillet (2009, see Appendix). As in the two previous experiments, participants had to estimate in a ratio form the probability that the promise or the threat uttered by the mother (e.g., "If you exercise the dog then I cook you a cake for dinner") is true or false for a drawing taken at random from a set of drawings representing the four possible cases.

4.2. Results

We expected that the heuristic process would produce a two-model initial representation ($p\ q$ and $\neg\, p \neg$ q) that would block in turn any fleshing out, resulting in an equivalence interpretation. So, we added the equivalence response to the three other responses previously observed in the probability task. As in Experiments 1 and 2, the percentages of conjunctive, defective biconditional, defective conditional and equivalence responses did not vary between "true" and "false" problems. So, we collapsed the responses from the two conditions.

In accordance with our hypothesis, equivalence was the main response for all grades for both promises (59%, 62%, and 63%) and threats (59%, 62%, and 63%, Fig. 4). There was no effect of grade, no effect of type of conditional and no interaction, *Fs* < 1. Apart from equivalence

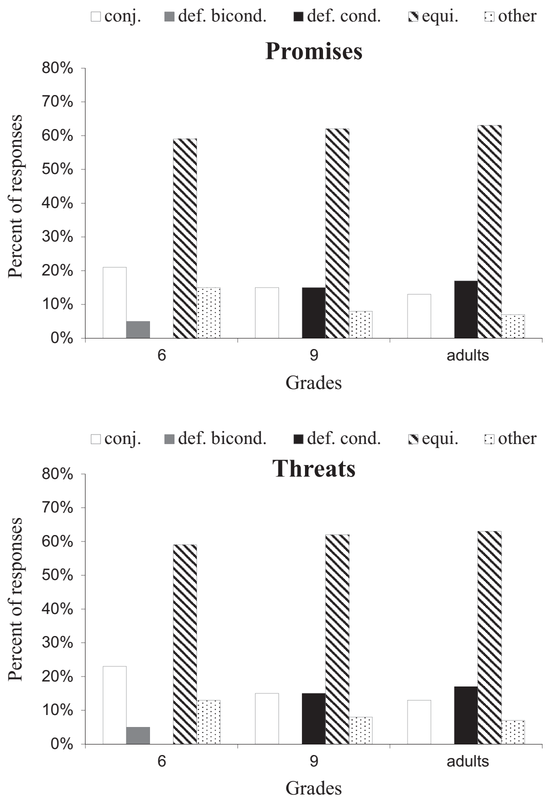

Fig. 4. Percent of responses categorized as conjunctive (Conj.), defective biconditional (Def. Bicond.), defective conditional (Def. Cond.), and equivalence (Equi.) in the probability task involving promises and threats.

responses, there were some conjunctive responses the rate of which was not affected by the type of inducement and that slightly decreased in older groups, a decrease that did not reach significance, $F < 1$, whereas defective conditional responses appeared in ninth grade for both types of inducement. Concerning the response patterns analysis, the results were exactly the same for promises and threats. Moreover, this analysis confirmed our first results because a majority of participants in all grades were consistent in giving equivalence responses (59%, 62%, and 63% of the 6th graders, 9th graders, and adults, respectively). There were some conjunctive responders (21%, 15%, and 13%, respectively), whereas rarer participants adopted defective biconditional (a single participant in 6th grade) or conditional responses (15% in 9th graders and 17% in adults).

4.3. Discussion

The results of this experiment confirmed that the evaluation of the probability of conditionals mirrors the judgments of truth-value revealed by the truth-table task. Depending on the nature of the conditionals, these probability evaluations can sometimes strongly differ from *the Equation*, even in educated adults, as it is the case for inducements. It could of course be supposed that the addition of the inverse implicature "*If not-p then not-q*" to the "*If p then q*" conditional under study would lead reasoners to evaluate the probability of inducements as a composite of the probabilities associated with each conditional. For example, an intuitive probability calculus integrating the two conditional probabilities P(q|p) and P(*not-q*|*not-p*), could lead to respond P(p & q) + P(*not-p* & *not-q*), which corresponds to the probability associated with the equivalence response. For example, presented with 2 *p* & *q* cases, 1 *p* & *not-q* case, 3 *not-p* & *not-q* cases, and 2 *not-p* & *q* cases, such a reasoner could evaluate the probability of the *If p then q* conditional as 2 out of 3, of the *If not-p then not-q* conditional as 3 out of 5, and give a response of the form 5 out of 8. The problem with such an explanation is that the previous experiments have established that the way the different age groups evaluate conditionals varies and, as a consequence, such a strategy of calculation should lead to different responses in the different groups, whereas the three age groups gave the same predominant response corresponding to the equivalence reading. A simpler explanation is to assume that our participants based their responses on the ratio between the cases that make, in their view, inducements true (or false), and those cases that are relevant for their truth-value. Our previous results established that these two sets of cases do not evolve with age from childhood to adulthood (Gauffroy & Barrouillet, 2009) and determine the responses observed in the probability task.

5. General discussion

This study explored the way in which the evaluation of the probability of a variety of conditionals evolves with age from early adolescence to adulthood. Our aim was to assess the pervasiveness of *the Equation*, the evaluation of the probability of the conditional as the conditional probability, through ages, types of conditionals and contents. Three main findings arose from our study. First, evaluating the probability of the conditional as the conditional probability is a late developmental achievement. Totally absent at the beginning of adolescence, the response corresponding to *the Equation* does not become predominant before adulthood. Second, even in a highly educated adult population selected on stringent academic criteria such as Swiss university students, a substantial minority of about one third of this population is consistent in responses that differ from *the Equation*. Third, variations in the content of basic and causal conditionals affect the basis on which the probability of the conditional is evaluated, often abolishing developmental differences in the range of ages considered here. In the following, we discuss these three points in turn.

5.1. A late developmental achievement

The present study has fully confirmed the results of our earlier investigations concerning the evaluation of the probability of basic conditionals (Gauffroy & Barrouillet, 2009). For conditionals characterized by a strong developmental trend (e.g., basic NN or weak causal conditionals), the probability corresponding to *the Equation* does not become predominant before adulthood, being preceded by at least two developmental levels that mirror the conjunctive and defective biconditional patterns observed in the development of the truth-table task (see for example Gauffroy & Barrouillet, 2009, 2011). In light of these findings, we must consider whether the *Equation* should continue to be considered the foundation of understanding in humans of the conditional, as the new-paradigm theoreticians argue?

For the theoretical approaches pertaining to the "old" paradigm, such as the mental logic or the mental model theories, a strong argument supporting the psychological validity of the hypothesized fundaments of reasoning was their availability in the youngest reasoners. For example, the mental logic theories assumed that human reasoning is underpinned by logical inference schemas triggered by the syntactic structure of the premises (Braine & O'Brien, 1991, 1998; Rips, 1994). In one of the most popular of these mental logic theories, Braine (1990; Braine & O'Brien, 1991) assumed that these schemas constitute a natural logic that makes use of what he defined as primary skills. The inferences pertaining to these primary skills were assumed to be universal, often made more or less automatically and as a consequence essentially without errors by adults (Braine, Reiser, & Rumain, 1984). Concerning *If*, it was assumed that the primary skill corresponds to Modus Ponens: when two propositions of the form *if p then q* and *p* are available, the rule derives a conclusion of the form *q*. Importantly, among several kinds of evidence for his theory, the first evidence advocated by Braine (1990) was the primitiveness of the hypothesized inferential forms and the fact that they are available early in children. He noted that children make inferences of the Modus Ponens form as early as it seems practicable to test them, at least by five years of age (Ennis, 1976; Hawkins, Pea, Glick, & Scribner, 1984), and he added "intuitively, if one is unable to make inferences of this form, one does not understand the meaning of *If*" (Braine, 1990, p. 149). Those reasoning forms that appeared more difficult and of later advent in development were excluded from the primary skills and the basic meaning of the connective. This was the case for schemas supporting suppositional reasoning such as the schema of *If* introduction. Naturally, the developmental origin of this mental logic was a key question that received various responses, from innateness (Macnamara, 1986) to more nuanced conceptions. Braine (1990) suggested that this would depend on the type of schema and proposed that the schema for *If* was learned by mapping the conditional on to a pre-linguistic *concept* of contingency relation. In other words, the meaning of the conditional was conceived as rooted in the most basic and elementary cognitive functions. Although the mental model theory rejects the idea of a mental logic made of schemas of inference, the adequacy of its most basic processes with children's reasoning performance is frequently stressed and taken as evidence for the theory (Johnson-Laird, 2006; Johnson-Laird & Byrne, 2002). According to this theory, the first representation coming to mind when understanding an *If p then q* sentence is a model representing the co-occurrence of *p* and *q* (Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992). A straightforward developmental prediction is that young children should be limited to this first processing step and interpret conditionals as akin to conjunctions, a prediction confirmed by several studies (Barrouillet & Lecas, 1998; Delval & Riviere, 1975; Gauffroy & Barrouillet, 2011; Paris, 1973; Politzer, 1986).

In summary, it was of crucial importance for the theories of the "old" paradigm to have their main tenets corroborated by developmental evidence, with the hypothesis that the most basic processes underpinning adult performance should characterize the earliest developmental levels. Our results showed that this is far from being the case for *the Equation*, which characterizes the most elaborated developmental level. Of course, it could be argued that children become Bayesian at the end of adolescence as they become logical in Piaget's theory. However, Piaget provided a rationale for this evolution with children's thinking going through different levels from mere intuitions to concrete logic and culminating with a formal logic. Our modified mental model theory provides such a rationale for the age-related evolution of the evaluation of the probability of conditionals. If we assume that individuals assess this probability by calculating the ratio between those cases that make the conditional true and those that are relevant for its truth value, setting aside the cases matching fleshed out models, it becomes easy to understand the developmental trend leading from conjunctive to defective biconditional responses and then defective conditional responses (i.e., those corresponding to *the Equation*). The conjunctive response comes from the fact that the initial *p q* model is not fleshed out, all the other cases making the conditional false. The defective biconditional response comes from the construction of the $\neg p \neg q$ model through fleshing out, making *not-p* & *not-q* cases irrelevant whereas *not-p* & *q* cases still make the conditional false. Finally, the defective conditional response comes from a complete representation in which the $\neg p \neg q$ and $\neg p\ q$ models are constructed through fleshing out, the corresponding cases being irrelevant for the truth value of the conditional. However, such an explanation makes *the Equation* a mere by-product of an interpretation that is not probabilistic in nature. When confronted with the probability task, people produce a response corresponding to *the Equation*, but the defective truth table underpinning this response would result from the processes by which mental models are constructed and their resulting epistemic status, and not from a probabilistic understanding of the conditional.

It should be noted that an alternative explanation of this developmental trend has been recently proposed by Fugard et al. (2011) who suggest an inhibitory account. The trend from a conjunctive to a defective biconditional and then a defective conditional response would reflect a narrowing of the hypothetical scope. At the beginning, the whole event space is seen as relevant, resulting in a conjunctive interpretation and response. The defective biconditional response (i.e., *biconditional event* in Fugard et al.'s terms) reduces the relevant events to those cases where either *p* or *q* is true (i.e., excluding *not-p* & *not-q* cases). Finally, the defective conditional response (the *conditional event*) exclusively focuses on *p* cases, excluding *not-p* & *not-q* and *not-p* & *q* cases. This reduction in scope with age would be due to an increasing ability to inhibit irrelevant cases from the event space. There is no doubt that the capacities of inhibition increase with age (e.g., Davidson, Amso, Anderson, & Diamond, 2006). However, it is doubtful that the conjunctive and defective biconditional responses are due to the inefficiency of some inhibitory process. The conjunctive and biconditional developmental levels have been observed in virtually all the types of tasks used to study conditional reasoning (Barrouillet & Lecas, 1998, 2002; Barrouillet et al., 2000, 2008; Gauffroy & Barrouillet, 2009, 2011). Of course, when focusing on the probability task, it can be thought that conjunctive and defective biconditional responders fail to inhibit *not-p* & *not-q* and *not-p* & *q* cases, taking them into account in their computations. The problem with the inhibition account is that in other tasks such as the production of cases compatible with the conditional, conjunctive and biconditional responses are also observed in which children and adolescents *fail* to produce the same *not-p* & *not-q* and *not-p* & *q* cases (the conjunctive response corresponds to the production of the sole *p* & *q* cases, whereas the biconditional response consists of producing only *p* & *q* and *not-p* & *not-q* cases, Barrouillet & Lecas, 1998). It is difficult to imagine that adolescents are, in the probability task, unable to inhibit those cases that they fail to evoke in other tasks in which their production is required. It is far simpler to assume that the conjunctive and biconditional developmental levels are due to the inability to construct a complete three-model representation. Those cases that do not correspond to any of the models constructed are considered as making the conditional false. Consequently, they are not constructed in production tasks, but they are taken into account in the probability task because they are relevant for the truth-value of the conditional, being considered as making it false.

Another attempt to accommodate the probabilistic approach with our results could be to adopt a dual-process conception first introduced by Verschueren, Schaeken, and d'Ydewalle (2005) and extended to development by Markovits (2014). In this conception, two types of reasoning would coexist. The first, based on Type 1 processes, would rely on the automatic retrieval of knowledge from long-term memory and underpin an assessment of the probability of the conclusion under study. The second, more analytical and involving Type 2 processes, would underpin deduction through the active search for counterexamples. This approach has recently been used by Markovits (2014) to account for the development of conditional reasoning from childhood to adulthood. The later developmental stage of this search of counterexamples consists of the capacity to reason from abstract conditionals. The conditionals we used involved unfamiliar relations between their antecedent and their consequent for which participants cannot retrieve knowledge from long-term memory. Thus, it could be argued that our task mainly solicited counterfactual-based reasoning, hence the strong development we observed, instead of the probabilistic reasoning theorized by the new paradigm psychology of reasoning. However, this same task is usually considered as providing one of the strongest evidence supporting the probabilistic view of reasoning (Evans et al., 2005; Fugard et al., 2011). Moreover, the dual-process view advocated by Verschueren et al. (2005) and Markovits (2014) establishes a clear demarcation between deductive and probabilistic forms of reasoning, whereas one of the ultimate claims of the new paradigm is that deductive processes are probabilistic in nature. Thus, this dual-process approach does not seem to be sufficient to re-establish *the Equation* as the kernel of the conditional.

Finally, a further possibility would be to assume that *the Equation* does not result from the development of analytical processes but is instead a sophisticated intuition of late development, as described by the fuzzy trace theory (Brainerd & Reyna, 2001). According to this theory, processing based on fuzzy representations that capture the gist of concepts and situations progressively supersedes over development more analytical processing based on verbatim representations. In other words, the *If-heuristic* described by Evans (1989) would represent the gist of the conditional, only available in late adolescence and adulthood. Because semantic and pragmatic knowledge combine to support extraction of gist (Brainerd & Reyna, 2001), this approach could also help to understand the variations we observed through types of conditionals and contents. Though being plausible, such an account would remain ad hoc and would leave unexplained the sequence of developmental stages leading to the Equation. As we have seen, our modified mental model theory provides such an account without endorsing the probabilistic nature of human reasoning.

In summary, the late developmental advent of *the Equation* does not fit very well with the idea that it would reveal something fundamental in the way humans understand conditionals. It is true that *the Equation* can be related with the Ramsey test and that this procedure could account for the defective truth table, but this truth table is itself a quite late developmental achievement. A modified mental model theory can account for the different developmental levels that lead to its emergence in late adolescence and its predominance in adults (Barrouillet et al., 2008). The conjunctive, defective biconditional, and then defective conditional response patterns are predicted by the modified mental model theory and occur in the probability task as they occur in the truth-table task previously studied by Gauffroy and Barrouillet (2009, 2011). These three levels reflect the progressive increase in the number of mental models that children and adolescents can construct in understanding conditionals and that we have observed in virtually all the tasks used to investigate conditional reasoning (Barrouillet & Gauffroy, 2013). In the same way, our theory predicts the effects of the pragmatic and semantic modulations introduced by conditionals of different type and content that affect either the fleshing out process (e.g., BB conditionals or causal conditionals with alternative causes) or the construction of the initial model (inducements).

5.2. The lack of universality and the conjunctive response

Although evaluating the probability of the conditional as the conditional probability was the predominant response in adults for NN basic and weak causal conditionals, these responses were far from being universal, with only 53% and 61% of adults who were consistent in this response, respectively. Though these percentages could be judged as especially low, they do not differ from other studies using the probability task. Evans et al. (2003) reported about 50% of conditional probability patterns and about 43% of conjunctive patterns, whereas Oberauer and Wilhelm (2003) in their Experiment 3 identified only 44% of their participants as clearly conforming to the conditional probability response, whereas 34% privileged conjunctive responses. Oberauer, Geiger, Fischer, and Weidenfeld (2007) reported 57% of their participants exhibiting a conditional probability response in their Experiment 1, and 70% in their Experiment 3. Interestingly, Oberauer and Wilhelm (2003) asked their participants to assess both the probability of the conditional and the conditional probability. They noted, "only a subset of participants understood the two as equivalent. In Experiment 1B, 52% of participants gave the same value for P(q|p) and P(if p then q); only 23% of participants in Experiment 1A gave the same values consistently over all four conditions" (p. 685). Overall, the proportion of adults evaluating the probability of the conditional as the conditional probability seems to be only slightly higher than a half, with a substantial minority who produce conjunctive responses. When considering the proportion of adults conforming with the expected response in the probability task, the key question becomes: Are these relatively low percentages compatible with the claim that the core meaning of the conditional is probabilistic in nature and reflected by *the Equation*? Once more, the comparison with the traditional psychology of reasoning can be enlightening.

Consider for example the case of the Modus Tollens inference (from *If p then q* and *not-q*, concluding *not-p*). Reviews of the literature such as Evans, Newstead, and Byrne (1993) reported a frequency of endorsement of this inference in adults varying from about 50% to 80%. Although Modus Tollens is a valid inference for the formal logic, these frequencies of endorsement were usually considered as sufficiently low to set this inference aside from the core meaning of the conditional. The difficulty of Modus Tollens for adults led Braine and O'Brien (1991) to suggest that it was not part of the lexical entry for *If*, which comprises in this model Modus Ponens and a schema for Conditional Proof. In the same way, Modus Tollens is not part of the routines of the PSYCOP model proposed by Rips (1994). It can also be noted that, for the same reason, the $\neg p \neg q$ model that supports Modus Tollens in the mental model theory is not part of the initial model of *If* (Johnson-Laird & Byrne, 1991, 2002). This example makes clear that, for the "old" paradigm, the core meaning of a connective was expected to be almost universally shared by adult reasoners. At the very least, it is difficult to consider that a response endorsed by a narrow majority of adults reflects the deep meaning of a word as frequent as *If*.

To preserve the idea that *the Equation* reflects something fundamental, it can be argued that both *the Equation* and the conjunctive response often observed in adults reflect basically the same process. This was what Evans et al. (2003) seemed to assume by suggesting that both *the Equation* and conjunctive responses result from a Ramsey test, but incomplete in the case of conjunctive responses. Instead of comparing the probability of the p & q cases with that of the p & *not-q* cases, shallow processors would cut short the Ramsey test and stop at the p & q cases. This interpretation is corroborated by the fact that the participants giving conjunctive responses to the probability task have lower cognitive capacities than the others (Evans et al., 2007). However, we have already argued elsewhere that this explanation is not entirely satisfactory because it does not account for the conjunctive response observed when evaluating the probability of the conditional to be false. If conjunctive responses were produced by shallow processors performing an incomplete Ramsey test, these reasoners should cut short the Ramsey test at the p & *not-q* cases when evaluating the probability of a false conditional and respond with this probability. However, this response is practically never observed, conjunctive responders evaluating the probability of a false conditional as $1 - \text{P}(p \ \& \ q)$.

What is surprising with the conjunctive responses produced by adults in the probability task is that conjunctive response patterns are practically never observed at these ages in truth-table tasks (e.g., less than 5% in Barrouillet et al., 2008, and 0% in Gauffroy & Barrouillet, 2011). However, as we noted above, it is interesting to note that about 30% of adults adopt a conjunctive response in truth-table tasks when the alternative responses are reduced to a true/false choice instead of allowing for the "irrelevant" response (Gauffroy & Barrouillet, 2011). Even more interestingly, Oberauer et al. (2007) observed that the adoption of a conjunctive response pattern in a two-valued truth-table task correlates with the tendency for conjunctive responses in the probability task. By contrast, the adoption of a material implication reading of the conditional in the two-valued truth-table task correlated with the conditional probability response in the probability task. These findings offer an explanation of the conjunctive response in the probability task that is simpler and more convincing than an incomplete Ramsey test. It should not be forgotten that the probability task is a kind of truth-table task consisting of evaluating when a conditional is true or false. However, the two types of tasks differ in the range of the possible alternatives because the irrelevant option that is often available in truth-table tasks is not explicitly available in the probability task. It is fairly possible that a substantial minority of adults interpret the probability task as a two-valued truth-table task. After all, they are asked to evaluate the probability of the conditional to be true ("how likely are the following statements to be true of a card drawn at random from the pack", Evans et al., 2003, p. 324). Thus, when they have to decide whether *not-p* cases make the conditional true or false, they decide for "false", hence the conjunctive response. This would explain the correlations observed by Oberauer et al. (2007) between responses in the two-valued truth table task and the probability task. This would also explain why conjunctive responders in the probability task have lower cognitive capacities, because Gauffroy and Barrouillet (2011) observed that conjunctive responses in the two-valued truth-table task are probably mainly produced by reasoners who adopt a defective biconditional reading when the *irrelevant* response is available, a reading that is less developmentally advanced than the defective conditional. Once more, within this account, the responses to the probability task reflect truth-value judgments rather than some probabilistic meaning of the conditional, and it is fairly possible that the Ramsey test has nothing to do with the observed responses.

5.3. Variations with types of conditionals and contents

Our hypothesis was that when evaluating the probability of a conditional, people calculate the ratio between those cases that make the conditional true and those cases that are relevant for its truth-value. Previous studies had shown that truth tables strongly vary from one type of conditional to another, and even with contents within the same type of conditional. For example, basic conditionals involving binary terms often induce a biconditional reading that can be observed in the production of cases compatible with the conditional (Barrouillet & Lecas, 1998), in inference production (Barrouillet & Lecas, 2002) and in truth-table tasks in which they elicit defective biconditional patterns (Gauffroy & Barrouillet, 2009). Accordingly, the same BB conditionals elicited defective biconditional responses in the probability task in adolescents, but also in adults. The same phenomenon was observed for causal conditionals with few alternatives. Inducements like promises and threats have been shown to mainly induce equivalence readings in truth-table tasks (Gauffroy & Barrouillet, 2009; Newstead et al., 1997). In line with our hypothesis, these same inducements elicited equivalence responses in the probability task. In other words, the responses in the probability task mirror those observed in the truth-table tasks.

Of course, it could be argued that these variations in responses are not at odds with the probabilistic approach if it is assumed that people add pragmatic implicatures to the meaning of the conditional. Evans and Over (2004) noted that adding the converse implicature (adding *If q then p* to *If p then q*) conjoins two conditionals that have a defective table leading to the defective biconditional reading, as we observed for NN and strong causal conditionals. By contrast, adding the inverse implicature (adding *If not p then not q*) leads to the equivalence interpretation we observed with promises and threats. As we suggested above, the evaluation of the probability of the resulting conditionals would reflect some combination of two *Equations* by naïve probabilistic reasoners. For example, *the Equation* corresponding to the converse implicature (P(q & p)/[P(q & p) + P(q & *not-p*)]) could be combined with that of the initial conditional (P(p & q)/[P(p & q) + P(p & *not-q*)]) following some loose additional rule leading to the defective biconditional response (P(p & q)/[P(p & q) + P(p & *not-q*) + P(*not-p* & q)]). However, this type of explanation encounters a main difficulty. We have seen that the evaluation of the probability varies with age. Consequently, combining the probabilities of the conditional and its converse implicature would lead to different outcomes in the different age groups. For example, combining two conjunctive responses should lead to a conjunctive response, leading to the same rate of conjunctive responses in BB and NN conditionals or in weak and strong conditionals. This is not what we observed, our results revealing the predominance of the same response in each age group studied for both BB and strong causal conditionals, namely the defective biconditional response. Another difficulty for the hypothesis of added implicatures is to explain why variations in content (BB conditionals or strong causal relations) call for the addition of the converse implicature whereas inducements would elicit the inverse implicature. We have argued in Gauffroy and Barrouillet (2009) that the inverse implicature is part of the meaning of inducements, hence their production by Type 1 processes and the fact that inducements are deemed true for *not-p* & *not-q* cases. By contrast, the biconditional reading of BB and strong causal conditionals would result from a limitation of the fleshing out process to the $\neg p \neg q$ model, resulting in a defective biconditional truth table (i.e., *not-p* & *not-q* cases that match a fleshed out model are considered as irrelevant whereas the conditional is deemed false for *not-p* & *q* cases that are not represented). Thus, the evaluation of the probability of the conditional as the conditional probability appears as volatile as the everyday inferences are defeasible, a defeasibility that was advocated to jettison the standard logic as a norm for human reasoning (Oaksford & Chater, 2001).

Overall, the variability in the evaluation of the probability of different types of conditionals involving different contents does not speak for *the Equation* as capturing something essential for the meaning of *If*. Instead, proposals like the concept of semantic and pragmatic modulation in the mental model theory, by which models are added or not produced, depending on content and context, seem at least as convincing (Byrne & Johnson-Laird, 2009; Johnson-Laird & Byrne, 2002). These modulations that affect the nature and the epistemic status of the models constructed have a direct impact on truth-value judgments (Gauffroy & Barrouillet, 2009). These modulations determine in turn the evaluation of the probability of these conditionals because probabilities are computed from these truth-value judgments. Indeed, as our results made clear, at each age, for each type of conditional and for every content, probabilities mirror exactly truth tables. Thus, we think that our modified mental model theory provides an account of the responses observed at different ages and for different conditionals to the probability task that is as least as convincing as the explanations offered by the new paradigm approach.

6. Conclusions

The aim of this study was to assess the pervasiveness of *the Equation*, the evaluation of the probability of the conditional as the conditional probability. Our results suggest that *the Equation* characterizes the evaluation of certain types of conditionals by a narrow majority of adult university students. When *the Equation* is the predominant response in adults, it is preceded by two developmental levels reflecting interpretations of the conditional that have been observed with the other tasks investigating conditional reasoning (production of cases compatible with a conditional, truth-table tasks, production and evaluation of inferences). Overall, *the Equation* does not exhibit the universality that was requested and taken as evidence for the role of cognitive primitive by the "old" paradigm, such as Modus Ponens for the mental logic, an inference endorsed almost systematically by adults (Evans et al., 1993) and already present in young children (Rumain, Connell, & Braine, 1983).

Although Elqayam and Over (2013) stated that studying probability judgments will tell us more about the psychology of reasoning than trying to figure out how far people conform to binary extensional logic, our results indicate that probability judgments do not tell us much more than the truth-table tasks studied by Gauffroy and Barrouillet (2009) with the same conditionals. The distribution of responses in the probability task can be predicted from the distribution of responses observed in a truth-table task. It could be argued that the truth-table task used by Gauffroy and Barrouillet (2009) goes beyond the binary logic by involving the response option "irrelevant", the responses of irrelevance resulting from the probabilistic nature of the conditional and the Ramsey test. However, it remains uncertain that this is the case. We have argued that the responses of irrelevance do not necessarily reflect a Ramsey test process, but the fleshed out nature of the corresponding mental models, an hypothesis corroborated by developmental findings (Barrouillet et al., 2008). Moreover, it is even unclear that the results of the probability task have something to do with the probabilistic conditional and are at odds with the binary logic. We have already evoked the study by Oberauer et al. (2007) who presented the same adult participants with a task of probability judgment and two variants of the truth-table task, one with and one without the response option "irrelevant" (in this latter task, people have to decide for each case whether it makes the conditional true or false). Quite amusingly, the best predictor of the tendency to evaluate the probability of the conditional as the conditional probability (i.e., *the Equation*) was the production of a truth-table task corresponding to the material implication in the two-valued truth-table task (i.e., the conditional is false for the *p* & *not-q* case and true for all the other cases). Importantly, the material implication pattern was a better predictor than the tendency to produce the De Finetti table when the "irrelevant" response was available.[3] This fact suggests that there is no empirical reason to suppose that *the Equation* is more related with De Finetti table and the Ramsey test than with the meaning of the conditional in the binary extensional logic, which is precisely the material implication. This result, judged as "unexpected" by Oberauer et al. (2007), is nonetheless easy to understand. Both the material implication pattern in the two-valued truth table task and the defective conditional pattern in the three-valued truth table task reflect the same and highest developmental level of conditional understanding (Gauffroy & Barrouillet, 2011). This highest level corresponds to the construction of a three-model representation that underpins the evaluation of the probability of the conditional as the conditional probability as we explained above. Overall, the results of the probability task can be easily accounted for by a modified mental model theory as proposed by Barrouillet et al. (2008) and Gauffroy and Barrouillet (2009) without endorsing the probabilistic approach of conditional reasoning.

To conclude, the evidence supporting *the Equation* in our developmental study can appear disappointingly scarce, and it could be wondered how it is possible to not find better empirical support for a predominant theoretical framework such as the probabilistic approach to human reasoning (Elqayam & Over, 2013; Oaksford & Chater, 2001, 2009). However, it should be noted that our disappointing results are not isolated. For example, Oberauer (2006) compared the capacity of seven formal models of reasoning to account for a very large set of data on the endorsement and rejection of the four conditional inference forms (i.e., Modus Ponens, Modus Tollens, Affirmation of the Consequent, and Denial of the Antecedent). The best fit was provided by a modified version of the mental model theory augmented by directionality of the models, as proposed in Barrouillet and Lecas (1998) and Barrouillet et al. (2000), and by the dual-process model proposed by Verschueren et al. (2005) in which heuristic processes are added to the original mental model theory. The suppositional theory proposed by Evans gave, according to Oberauer (2006, p. 267), "less than satisfactory fits to the data", whereas the probabilistic theory of Oaksford, Chater, and Larkin (2000) provided fits that were worse than all competitors. In the present study, *the Equation* appears to be a late developmental achievement only endorsed by a narrow majority of educated adults when evaluating the probability of certain conditionals. From these findings, *the Equation* and its related theory of conditional reasoning could turn out to be what Bowers and Davis (2012) have described as Bayesian "just-so" stories in psychology.

Acknowledgement

This research was supported by a grant from the Swiss National Science Foundation N° 100014_132053 to Pierre Barrouillet.

Appendix A

BB conditionals used in Experiment 1:

"If the pupil is a boy then he wears glasses"
"If the door is open then the light is switched on"

[3] From several problems in a probabilistic truth-table task in which was varied the probability of the four logical cases, Oberauer et al. (2007) evaluated the sensibility of participants' responses to two indices, the ratio of *p* & *q* cases to *p* & *not-q* cases, underpinning the conditional probability response, and the frequency of the *p* & *q* cases that underpins the conjunctive response. A participant conforming to *the Equation* should exhibit a sensibility to the ratio and an insensibility to the frequency in its evaluations of probability. Moreover, Oberauer et al. (2007) calculated from the truth table tasks a TFII index corresponding to the De Finetti table in the task with the response option "irrelevant" (i.e., responding True, False, Irrelevant and Irrelevant for the *p* & *q*, *p* & *not-q*, *not-p* & *q*, and *not-p* & *not-q* cases, respectively), and a TFTT index corresponding to the material implication in the truth-table task without the "irrelevant" option (i.e., responding True for the *p* & *q* case, False for the *p* & *not-q* case, and True for both *not-p* cases). Oberauer et al. (2007, Table 10) report correlations between the TFII index and the ratio and frequency indices of −.05 and .26 respectively, whereas the same correlations are of −.31 and .34 with the TFTT index, clearly indicating that the TFTT index is a better predictor of the tendency to evaluate the probability of the conditional as the conditional probability.

"If the student is a woman then she wears a shirt with long sleeves"
"If the piece is big then it is pierced"

NN conditionals used in Experiment 1:

"If the card is yellow then a triangle is printed on it"
"If there is a star on the screen then there is a circle"
"If he wears a red t-shirt then he wears a green trousers"
"If there is a rabbit in the cage then there is a cat"

Strong causal relations used in Experiment 2:

"If the button 3 is turned then the blackboard's lights are switched on"
"If the lever 2 is down, then the rabbit's cage is open"
"If the second button of the machine is green then the machine makes sweets"
"If I pour out pink liquid in the vase then stars appear on it"

Weak causal relations used in Experiment 2:

"If the touch F5 is pressed then the computer screen becomes black"
"If the boy eats *alkali* pills then his skin tans"
"If the fisherman puts flour in the water then he catches a lot of fishes"
"If the gardener pours out *buntil* in his garden then he gathers a lot of tomatoes"

Promises used in Experiment 3:

"If you gather the leafs in the garden then I give you 5 francs"
"If you score a goal then I name you captain"
"If you exercise the dog then I cook you a cake for dinner"
"If you clean your room then you watch the TV"

Threats used in Experiment 3:

"If you break the vase then I take your ball"
"If you do not buy the bread then you do not play video games"
"If you do not your homework then you do not go to the attraction park"
"If you have a bad mark then you do not go to the movie"

References

Barrouillet, P. (2011). Dual-process theories of reasoning: The test of development. *Developmental Review, 31*(2–3), 151–179.
Barrouillet, P., & Gauffroy, C. (2013). Dual processes and mental models in the development of conditional reasoning. In P. Barrouillet & C. Gauffory (Eds.), *The development of thinking and reasoning* (pp. 95–121). Hove: Psychology Press.
Barrouillet, P., Gauffroy, C., & Lecas, J. F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review, 115*(3), 760–771.
Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition, 75*, 237–266.
Barrouillet, P., & Lecas, J. F. (1998). How can mental models account for content effects in conditional reasoning? A developmental perspective. *Cognition, 67*, 209–253.
Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking and Reasoning, 5*(4), 289–302.
Barrouillet, P., & Lecas, J. F. (2002). Content and context effects in children's and adults' conditional reasoning. *Quarterly Journal of Experimental Psychology, 55*(3), 839–854.
Barrouillet, P., Markovits, H., & Quinn, S. (2001). Developmental and content effects in reasoning with causal conditionals. *Journal of Experimental Child Psychology, 81*(3), 235–248.
Beck, S. R., & Riggs, K. J. (2014). The development of counterfactual reasoning. In H. Markovits (Ed.), *The developmental psychology of reasoning and decision making* (pp. 165–181). Hove: Psychology Press.
Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin, 138*(3), 389–414.
Braine, M. D. S., & O'Brien, D. P. (1991). A theory of If: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 182–203.
Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
Braine, M. D. S. (1990). The natural approach to reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives* (pp. 135–158). Hillsdale, NJ: Lawrence Erlbaum Associates.
Braine, M. D. S., Reiser, B. J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. *Psychology of Learning and Motivation – Advances in Research and Theory, 18*, 313–371.
Brainerd, C. J., & Reyna, V. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. *Advances in Child Development and Behavior, 28*, 49–100.
Byrne, R. M. J., & Johnson-Laird, P. N. (2009). 'If' and the problem of the conditional. *Trends in Cognitive Sciences, 13*, 282–287.
Chater, N., & Oaksford, M. (Eds.). (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford: Oxford University Press.
Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science, 7*(12), 547–552.
Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory & Cognition, 23*(5), 646–658.
Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition, 19*(3), 274–282.
Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.
Delval, J. A., & Riviere, A. (1975). "Si llueve, Elisa lleva sombrero": A psychological investigation of the conditional truth table. *Revista De Psicologia General y Aplicada, 136*, 825–850.
Edington, D. (1995). On conditionals. *Minds, 104*, 235–329.
Elqayam, S., Bonnefon, J. F., & Over, D. (2013). Special issue on basic and applied perspectives for new paradigm psychology of reasoning. *Thinking and Reasoning, 19*(3–4).
Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking & Reasoning, 19*(3–4), 249–265.
Ennis, R. H. (1976). An alternative to Piaget's conceptualization of logical competence. *Child Development, 47*, 903–919.
Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge.
Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Brighton, UK: Lawrence Erlbaum Associates.
Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978–996.
Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378–395.
Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment*. Hove, UK: Psychology Press.
Evans, J. St. B. T., & Frankish, K. (2009). *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.
Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition, 35*, 1772–1784.
Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 321–355.
Evans, J. St. B. T. (2012). Dual process theories of reasoning: Facts and fallacies. In K. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 115–133). New York: Oxford University Press.
Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ: Lawrence Erlbaum.

Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press.
Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review, 112*, 1040–1052.
Evans, J. St. B. T., & Stanovich, K. E. (2013). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on Psychological Science, 8*, 263–271.
Falmagne, R. J. (1990). Language and the acquisition of logical knowledge. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives* (pp. 111–134). Hillsdale, NJ: Lawrence Erlbaum Associates.
Fugard, A. J. B., Pfeifer, N., Mayerfofer, B., & Kleiter, G. D. (2011). How people interpret conditionals: Shifts toward the conditional event. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 635–648.
Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditional: An integrative developmental theory. *Developmental Review, 29*, 249–282.
Gauffroy, C., & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. *Developmental Psychology, 47*, 1000–1011.
Gauffroy, C., & Barrouillet, P. (2014a). Conditional reasoning in context: A developmental dual processes account. *Thinking and Reasoning, 20*, 372–384.
Gauffroy, C., & Barrouillet, P. (2014b). A developmental mental model theory of conditional reasoning. In H. Markovits (Ed.), *The developmental psychology of reasoning and decision-making* (pp. 63–83). Hove: Psychology Press.
Goldvarg, E., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science, 25*(4), 565–610.
Hawkins, J., Pea, R. D., Glick, J., & Scribner, S. (1984). Merds that laugh don't like mushrooms: Evidence for deductive reasoning by preschoolers. *Developmental Psychology, 20*(4), 584–594.
Harris, P. L., & Nunez, M. (1996). Understanding of permission rules by preschool children. *Child Development, 67*, 1572–1591.
Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
Johnson-Laird, P. N. (2006). *How we reason?* Oxford: Oxford University Press.
Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 646–678.
Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418–439.
Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist, 58*, 697–720.
Kahneman, D. (2011). *Thinking fast and slow*. London: Penguin Books.
Kahneman, D., & Tversky, A. (1972). Subjective rationality: A judgment of representativeness. *Cognitive Psychology, 3*, 430–453.
Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.
Klaczynski, P. A., & Cottrell, J. M. (2004). A dual-process approach to cognitive development: The case of children's understanding of sun cost decisions. *Thinking and Reasoning, 10*(2), 147–174.
Klaczynski, P. A., & Felmban, W. S. (2014). Heuristic and biases during adolescence: Developmental reversals and individual differences. In H. Markovits (Ed.), *The developmental psychology of reasoning and decision making* (pp. 84–112). Hove: Psychology Press.
Kuhn, D. (2011). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of childhood cognitive development* (2nd ed., pp. 497–523). Oxford: Blackwell.
Kushsnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology, 43*, 186–196.
Lecas, J. F., & Barrouillet, P. (1999). Understanding conditional rules in childhood and adolescence: A mental models approach. *Current Psychology of Cognition, 18*, 363–396.
Light, P., Blaye, A., Gilly, M., & Girotto, V. (1989). Pragmatic schemas and logical reasoning in 6- to 8-year-old children. *Cognitive Development, 4*, 49–64.
Macnamara, J. (1986). *A border dispute: The place of logic in psychology*. Cambridge, MA: Bradford Books/MIT Press.
Manktelow, K. I. (2012). *Thinking and reasoning: Psychological perspectives on reason, judgment, and decision making*. Hove: Psychology Press.
Markovits, H. (1984). Awareness of the possible as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology, 75*, 367–376.
Markovits, H. (2014). How to develop a logical reasoner: A hierarchical model of the role of divergent thinking in the development of conditional reasoning. In H. Markovits (Ed.), *The developmental psychology of reasoning and decision-making* (pp. 148–164). Hove: Psychology Press.
Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22*(1), 5–36.
Markovits, H., Fleury, M. L., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development, 69*(3), 742–755.
Newstead, H. E., Ellis, M. C., Evans, J. St. B. T., & Dennis, I. (1997). Conditional reasoning with realistic material. *Thinking and Reasoning, 3*(1), 49–76.
Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608–631.
Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*(8), 349–357.
Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.
Oaksford, M., & Chater, N. (2009). The uncertain reasoned: Bayes, logic and rationality. *Behavioral and Brain Sciences, 32*, 105–120.
Oaksford, M., & Chater, N. (2010). *Cognition and conditionals: Probability and logic in human thinking*. Oxford: Oxford University Press.
Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning Memory & Cognition, 26*, 883–889.
Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology, 53*, 238–283.
Oberauer, K., Geiger, S. M., Fischer, K., & Weidenfeld, A. (2007). Two meanings of *If*? Individual differences in the interpretation of conditionals. *Quarterly Journal of Experimental Psychology, 60*, 790–819.
Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 680–693.
Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology, 54*(1), 62–97.
Overton, W. F. (1990). Competences and procedure: Constraints on the development of logical reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives* (pp. 1–32). Hillsdale, NJ: Lawrence Erlbaum Associates.
Paris, S. G. (1973). Comprehension of language connectives and propositional logic relationships. *Journal of Experimental Child Psychology, 16*, 278–291.
Piaget, J., & Inhelder, B. (1959). *La genèse des structures logiques élémentaires*. Actualités pédagogiques et psychologiques.
Politzer, G. (1986). Laws of language use and formal logic. *Journal of Psycholinguistic Research, 15*, 47–92.
Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition, 68*(3), 93–101.
Reyna, V., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review, 31*, 180–206.
Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
Rumain, B., Connell, J., & Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as in adults. *Developmental Psychology, 19*, 471–481.
Vergauwe, E., Gauffroy, C., Morsanyi, K., Dagry, I., & Barrouillet, P. (2013). Chronometric évidence for dual process mental model theory of conditional. *Journal of Cognitive Psychology, 25*(2), 174–182.
Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual process specification of causal conditional reasoning. *Thinking & Reasoning, 11*(3), 239–278.