Mistaking the Instance for the Rule: A Critical Analysis of the Truth-table Paradigm

Walter Schroyens (Walter.Schroyens@Ugent.be)  
University of Gent.  
2, H. Dunantlaan, 9000 GENT, Belgium

Abstract

Many studies probe for interpretations of <if A then C> by having people evaluate truth-table cases (e.g., <A,C>, <A_not-C>, <not-A,C>, <not-A_not-C>) as making the rule true or false, or being irrelevant. We argue that a single case can never prove a general rule to be true, as philosophy of science has taught any researcher. Giving subjects the impossible ‘true’ option would therefore bias results away from this response. In Experiment 1 people judged instead whether cases make a rule false, do not make the rule false, or are irrelevant to the rule. The experimental group (N=46) shows a significant increase in not-false vs. true responses of the control group (N=48). In Experiments 2, 3 the experimental groups judged whether cases make a rule true, corroborate it (i.e., make the rule more plausible, but neither true nor false), make it false, or are irrelevant to a rule. There was a significant reduction of irrelevant responses as compared to the default true/false/irrelevant task for the control groups. Results corroborate our conceptual analyses of the unsuitable ‘true’ response option and put into question all arguments that hinge on the presumed likelihood by which people consider truth contingencies to make a rule ‘true’.

General Introduction

As Shakespeare put it: “Much virtue in ‘if’”. Iffy propositions are the seat of our imagination; they ground our ability to explain the past and present and to foresee the future by projecting potential consequences of antecedent events. Reasoning about conditionals has accordingly attracted the interest of many cognitive scientists. The list of directly relevant papers literally runs in the thousands.

The so-called truth table task is one of the main research paradigms used to study people's understanding of conditionals. In the truth-based truth-table evaluation task, people evaluate the four truth-contingencies <A and C>, <A and not-C>, <not-A and C>, and <not-A and not-C> in relation to <if A then C>. In these contingencies or truth-table cases the antecedent and consequent event is either true (T) or false (F). An <A_not-C> contingency is accordingly referred to as the True-antecedent-False-consequent contingency (TF<A_not-C>). Table 1 gives and example of truth-table cases for the conditional like “if the figure is a circle, then the figure is coloured red”.

Our focus is on the truth-based format of the truth-table evaluation-task, in which people evaluate contingencies as making the rule true, making the rule false, or being irrelevant to evaluate conditionals. If our analysis of this task is right, it means that for more than 25 years, many researchers have used a conceptually flawed reasoning task in which people are asked to do something that is impossible: Any science student knows the induction problem: The impossibility of establishing the truth of a universal claim on the basis of a single case. “All swans are white “or “if it is a swan, then it will be white” is not rendered a true statement after observing a white swan, or even after having observed the one-trillionth white swan. Some truth-table task studies really give the evaluation-option 'true' whereas other studies instead use 'consistent'. The consistency-based tasks avoid the induction-problem by not asking people to evaluate conditionals as rendered 'true' by a particular truth-table case..

We present three new studies, showing that results and a fortiori theorizations have been biased by giving people the theoretically impossible (or at least problematical) task of judging whether a singular case makes a universal rule true. Use of 'true' has resulted in an overestimation of 'irrelevant' evaluations, and these 'irrelevant' evaluations have been crucial in arguments for or against theories of conditional reasoning. To consider the potential theoretical implications consider the following statement from Evans and Over (2004): “Let us recap some of the experimental evidence that has been presented in this book for the suppositional nature of if'. The most direct evidence comes from two main sources: studies of the how people judge the truth value of conditionals in truth table tasks (Chapter 2) and studies which ask people to assess the probability of conditional statements (Chapter 8). If the evidence for the suppositional nature of 'if' is so strongly tied to the empirical evidence on the truth-table task, it seems warranted to establish the firmness of these empirical foundations underneath the suppositional (and/or other) theories.

Experiment 1

We suggest one cannot ask people to judge whether a test-case makes a rule, such as <if A then C> true. It is impossible for a single case (or a subset of cases/tokens of a particular type like FT) to make a universal rule true. A theoretically sound truth-table task can therefore not use the unqualified option 'true'. We thus constructed an alternative truth-table task. The falsification-based truth-table task gives people the option 'not false' instead of 'true'. This falsification-based truth-table task bypasses the induction problem and should therefore reduce the frequency of 'irrelevant' evaluations due to people's sensitivity to the induction problem.

A second aim was to provide a more direct measure of people's sensitivity to the induction problem, which is presumed by the induction-problem hypothesis. Scholars and researchers recognize -- or should recognize -- one cannot prove a rule (hypothesis) true on the basis of a single case (experiment): the next test-case can always bring a potential falsification. But how do philosophically-scientifically untutored people, i.e., ordinary people think
about this? To see whether naïve reasoners understand the induction problem, we asked them about it in a meta-theoretical verification task. Imagine a set of 100 coloured figures and a description of 90 of these figures: 10 figures remain unidentified. When there are TF<\neg A \_\neg C> cases in the identified subset, we can decide that the rule is false without knowing the exact nature of the 10 remaining figures. When there are no TF cases in the identified subset, we are dealing with an indeterminate problem: one cannot, as yet decide whether the rule is true or false.

**Method**

**Design.** Participants served as their own control for evaluating the four truth-contingencies. Task format formed was a between-group variable. A first group (N = 48) received the truth-based task; a second group (N = 46) received the falsity-based task.

**Materials and procedure.** Instructions told participants had to "indicate for each combination (i.e., a type of colored figure) whether it makes the rule false, does not make it false or is irrelevant to the rule." For instance:

- The figure is a circle and is not colored red
- This combination:
  - O shows that the rule is false
  - O does not show the rule is false
  - O is irrelevant.

In the truth-based task, the ‘true’ option read: “shows that the rule is true”.

The meta-theoretical verification task asked people to "imagine a collection of 100 colored geometrical figures. You have identified the shape and color of 90 figures. Ten figures remain hidden and unidentified". The indeterminate problem-set described 65, 0, 20 and 5 TT, TF, FF and FT cases and explicated 10 figures were unidentified. The indeterminate problem-set described 50/15/20/5 different such cases. Participants judged whether it is "possible to evaluate whether the rule is true or false, taking into consideration that there remain unidentified figures?" and selected on of three options:

- O No, for the time being, one cannot know whether the rule is irrefutably true or false.
- O Yes, the rule is irrefutably true.
- O Yes, the rule is irrefutably false.

**Participants.** Participants were 17-19 final-year Quebecois high-school students (N = 94) who participated on a voluntary basis and received 25 for their time.

**Results**

**Truth-Table task.** Figure 1 presents selection rates for all response options in either the truth-based or falsity-based truth-table evaluation task. To analyze the task effect, we performed non-parametric tests on the frequency of True vs. Not-False responses. Figure 1 conveys that the overall task effect (True vs. Not-False; Mann-Whitney U = 454, Z = 4.916, p < .005) is only significant in size for the two false-antecedent cases (FT and FF: respectively, U = 717, Z = 2.927, p < .005 and U = 521, Z = 4.409, p < .0001). It also shows that the reduction in affirmative selections (True vs. Not-False) is mostly mirrored by an increase in irrelevancy judgements of both FT and FF (respectively, .261 vs. .542: U = 794, Z = 2.345, p < .05 and .239 vs. 583: U = 724, Z = 2.874, p < .005). There was a comparable tendency for increased false judgements of FT<\neg A \_C> and FF<\neg A \_\neg C>, but these numerical differences did not reach statistical significance.

**Meta-theoretical verification task.** When we have incomplete information about the world, we can never judge a rule to be true, that is, strictly and irrefutably true. The as yet unidentified cases could still falsify the rule. However, when the available information includes a falsification of the rule, it can be evaluated as such. When the problem was indeterminate (TF=0) most people (76.6%) correctly recognized that one cannot determinately establish the truth or falsity of the rule. When the problem was determinate (i.e., there were TF cases), fewer participants gave the indeterminate response (43.6%; Wilcoxon-T = 260.0, Z = 3.777, p < .001). The reduction in indeterminate responses (76.6 vs. 43.6) was matched by an increase in falsification responses (13.8 vs. 39.4: T = 225.0, Z = 3.151, p < .005). This is the correct response for determinate problem.

**Discussion**

The true/not-false effect corroborates our analyses regarding the conceptually flawed nature of the truth-based truth-table task. When the option is to choose between ‘true’ or ‘irrelevant’; the latter is the best alternative available: FF<\neg A \_\neg C> does not make <if A then C> true, and neither does it make the rule false. Our study confirmed that
using the unsuitable ‘true’ (vs. not-false) option increases the proportion of ‘irrelevancy’ responses or -- phrased the other way round -- when we used a suitable ‘not false’ response we see a decrease of ‘irrelevancy’ responses. That is, some people would state FF is irrelevant in relation to the rule, not because it is really irrelevant, but because it is as close to a correct response as one can get when neither one of the other two response alternatives are acceptable.

The induction-problem hypothesis, when taken on its own, suggests TT cases also do not make universal conditionals true. In Experiment 1 we nonetheless did not observe an increase of ‘correct’ not-false’ evaluations of TT cases. When we consider the induction-problem hypothesis not on its own, but within mental-models theory, we are given a suggestive explanation for the overall high TT[true] ratings. The so-called implicit-model or initial-model principle of mental-models theory (Johnson-Laird & Byrne, 2002) reflects the idea that people are cognitive miser and will not consider all possibilities all at once (even provided they could). "If it rains, then the streets get wet" is about it raining and the streets getting wet. This is accordingly the possibility people consider ab initio. Context and content might trigger consideration of other possibilities, i.e., construction of mental models representing these possibilities. Since TT is represented within the initial representation of <if A then C> it is difficult to image it would be ‘irrelevant’ regarding the rule: it is what the rule is about. False-antecedent cases are not relevant at first because they are not represented ab initio. TT cases are relevant at first because they are represented ab initio. The induction-problem reinforces the initial-model based irrelevancy judgment of false-antecedent cases. TT cases are relevant and since they clearly do not falsify the <if A then C> rule, ‘True’ is the best ‘non-irrelevant’ choice that is available in the ‘true/false/irrelevant’ task.

Experiment 2
The meaning of ‘not-false’ is clear per se, but is large in scope. The contrast-class of the negative ‘not-false’ subsumes both incorrect ‘true’ responses and the logico-philosophically correct response indicating that a confirmatory case increases the verisimilitude of a rule. Verisimilitude (truthfulness as opposed to categorical truth, i.e., verity) was popularized by Sir Karl Popper (e.g., Popper, 1963) and reflects an increased degree of corroboration, that is, resistance to falsification attempts. We suggested false-antecedent cases in the truth-based task are evaluated as ‘irrelevant’ because it is the best alternative available. That is, the ‘irrelevant’ option subsumed both cases where people really judge a case being irrelevant and cases judged ‘irrelevant’ because neither the true nor the false option adequately reflect the impact of these cases regarding the truth of a rule. We therefore gave reasoners a fourth response alternative, i.e., the corroboratory response which disambiguates ‘irrelevant’ and provides the logico-philosophically correct evaluation of non-falsifying truth contingencies. As compared to the classic ‘true/false/irrelevant’ task, we thus expect a decrease of ‘irrelevant’ evaluations in favour of a correct evaluation of the non-falsifying truth-contingencies. People should now also tend to judge TTnot as making the rule true, but as giving support to the rule without making the rule irrefutably true.

Method

Participants and design. Participants -- 17-23 year old Quebecois students -- served as their own control for evaluating the four truth-contingencies (cf. Table 1). Task format formed was a between-group variable. A first group (N = 61) received the truth-based task; a second group (N = 67) received the corroboratory-based task.

Materials and procedure. The truth-based task was identical to Experiment 1. The corroboratory-based task gave four responses options, translated from French:

- O Shows the rule is false,
- O Neither shows the rule is false, nor shows the rule is irrefutably true: The combination supports the rule, without proving it true or false.
- O Shows the rule is true.
- O Is irrelevant.

Results and Discussion

Figure 2 presents the selection rates of the different response options in both the truth-based (T/F/I) and corroboratory-based truth-table tasks (C/T/F/I). As expected, there was a significant reduction of irrelevant responses to the false-antecedent cases in the corroboratory-based task (FF: .705 vs. .328, U = 1274.0, Z = 3.671, p < .001; FT: .705 vs. .448, U = 1518.0, Z = 2.507, p < .005). This reduction of irrelevant responses corroborates our thesis that at least part of the ‘irrelevant’ evaluations are due to not being given a correct response option in the classic truth-based task. The false-antecedent cases do not make the conditional false, and neither do they make the conditional true. Irrelevant is the best option available, except when the correct indeterminate response is given.

The reduction in ‘irrelevant’ evaluations is largely due to the selection of the indeterminate ‘neither true nor false’ option in the corroboratory-based task. Figure 2 shows that the FF[false] and FT[false] rates are not affected by the logically correct response alternative, and that the effect of task format on their irrelevant ratings (d(FF)= -.477 and d(FT) = -.357) is closely matched by the ‘neither true nor false’ ratings (respectively .448 and .338). Figure 3 also shows that about 45% of people select the logico-philosophically correct evaluation of TT<A_C>, when given the opportunity to do so.

Figure 2 also shows that 17.9% judged that TF<A_not-C> makes the rule neither true nor false. This might be taken to suggest that ‘neither true nor false’ is a cautious response selected to express a general degree of uncertainty. This general-caution hypothesis clearly fails as a viable alternative explanation since it does not explain the variation in the frequency by which the ‘neither true nor false’ option is selected. To test the general-caution hypothesis we could exclude overly ‘cautious’ participants who judge TF<A_not-C> as making the rule ‘true’. There is not much of a point to do such, since it is a mathematical impossibility that eliminating these 17.9% participants
We also added a fourth response alternative to the truth-based task. The standard truth-based task gives only three response options to which we added the corroboratory response in the corroboration-based task. None of the three standard response options is a correct response for the corroboration truth-table cases. To increase the comparability between the truth-based and corroboration-based task we therefore introduced a fourth option "none of the above" in the former. In the corroboration-based task, this fourth response option and alternative to the default "true", "false", "irrelevant" option, expicates that a confirmatory case makes the rule neither true nor false, but makes the rule less likely to be false.

**Experiment 3**

We argued that 'irrelevant' evaluations in the standard truth-based paradigm are partly due to 'irrelevant' being the best option available for false-antecedent cases. FF cases neither proof the rule true nor prove the rule false. They are consistent but do not, cannot proof a rule to be true. It is only when one misinterprets truth as 'truthfulness' (verity versus verisimilitude) that FF can be taken to make a rule more truthful (though not true as such). Hence, if we block or hinder this misinterpretation fewer people should evaluate FF as making the rule true. When blocking a fuzzy interpretation of truth more people should judge FF is corroborating (or irrelevant when that is the best alternative available, as in the truth-based task).

A pre-study provided suggestive evidence for the hypothesis that stressing strict truth makes people more sensitive to the induction problem. It is only upon a strict versus fuzzy notion of truth that a single case can never prove a general rule to be true. To strengthen the experimental manipulation of stressing strict truth (and, thus, blocking a fuzzy notion of truth) we decided to introduce the manipulation directly into the responses options for all four truth-table cases. We stressed the contrast between truth (verity) and truthfulness (verisimilitude) by rephrasing the standard response option "makes the rule true" as "makes the rule irrefutably true". The standard false response "makes the rule false" was similarly extended as "makes the rule strictly speaking false". Stressing that 'true' means 'irrefutably true' increases the contrast between the normatively incorrect true response and the correct corroboratory response. That is, we should observe that the corroborating truth-table cases are more likely to be judged as such, when the meaning of the corroborating response option is accentuated and clarified in virtue of its contrast with 'irrefutable truth'.

**Method**

**Participants and Design.** All 152 participants were 2nd year bachelor. Seventy-nine students received the Standard Truth-based task; 42 with and 37 without stressing strict, irrefutable truth. Seventy-three students received the corroboration-based; 36 with and 37 without stressing irrefutable truth. Within each of the four experimental groups, each participant evaluated the four truth-contingencies.

**Materials and procedure.** The corroboration-based truth-table task gave the following response options:

- O Makes the rule [strictly speaking] false
- O Makes the rule [irrefutably] true
- O Is irrelevant regarding the rule
- O Is relevant but neither makes the rule false, nor does it make the rule [irrefutably] true: The figure is not excluded by the rule and it thus makes it less likely the rule is not true.

In the truth-based task, the fourth response alternative translates "none of the above".

**Results and Discussion**

Table 3 presents the selection rates of the different response options in both the truth-based (T/F/I/N) and corroboration-based truth-table tasks (T/F/I/C). The table also includes test statistics for comparison of the conditions with or without stressing strict truth as irrefutable truth. As observed in Experiment 3, the distribution of response selections changed significantly for FF<not-A_not-C>, but now both in the truth-based and corroboration based task. In the present study also TT<A_C> evaluations were affected by stressing the irrefutability of strictly true propositions, both in the truth-based and corroboration based tasks. Though it is obvious from looking at Table 3 that these effects are caused by a shift from incorrect true' evaluations to correct corroboratory or 'none of the above' evaluations this, is not measured by the chi-square tests. (Moreover, these test may
be invalid/unreliable when the expected frequencies in are lower than 5, which is the case for almost all truth-table cases. We therefore computed the frequency of correct responses (i.e., "none of the above" and corroborating). Combined over truth-based and corroboration-based task, there was a significant increase of correct responding on both the TT and FF cases (respectively 12.2 vs. 30.8: U = 2349.0, Z = 2.772, p < .01; 13.5 vs. 28.2, U = 2462.0, Z = 2.214, p < .05). The task format did not interact with these effects of stressing the irrefutability of true assertions.

The increase in correct responding corroborates our expectations. Normative evaluations are never absolute, i.e., they are always 'correct' relative to some more or less arbitrarily chosen and always debatable norm. In the present context -- i.e., the inductive context of evaluating the truth-value of rule on the basis of observations/cases -- the normative correctness of the corroborating response is only appropriate relative to a strict notion of truth. Stressing the irrefutability of strictly true propositions accordingly highlights people's potential sensitivity to the induction problem. The effect of stressing strict truth is an example of the general idealization hypothesis proffered by Schroyens, Schaecken, & Dieussaert (2008). They argued that 'the' interpretation of conditionals does not exist and that the core meaning (vs. interpretation) of a proposition is an idealization that abstracts from the multiple processing demands and limitations. The idealization-hypothesis states that it is only when we make the context and content more 'ideal' and conducive to idealized behaviour that we will observe performance in line with the idealized core meaning of the propositions.

The present experiment allows for a direct comparison between the two task formats because they are matched on the number of response alternatives. That is, comparing performance in the two task formats shows significant shifts in the response selection rates of all three corroborating cases, but not the falsifying TF case. As before, since there are more than two response categories, the shift in the distribution of responses does not tell us which responses have shifted. Using the frequency of correct responses, we see an increase of correct responses for all three corroboration cases from the unspecified and uninformative "none of the above" in the truth-based task to the specified corroboratory response in the corroboration-based task (respectively, Mann-Whitney U = 2340.0, Z = 2.807, p < .01; U = 1881.0, Z = 5.070, p < .0001; U = 2075.0, Z = 4.220, p < .0001). Specific pair-wise comparisons show that the increase of correct responses are matched by a decrease of true evaluations for TT (84.8 vs. 64.9: U = 2294.5, Z = 2.896, p < .005), a decrease of irrelevant evaluations for FT (67.1 vs. 35.1: U =, Z = 3.867, p < .001) and at least a tendency for a decrease of both true and irrelevant evaluations for FF (36.7 vs. 25.7: U = 2575.5, Z = 1.411, p = .158 and 51.9 vs. 32.4; U = 2335, Z = 2.361, p <.05).

### General Discussion.

The unsuitable nature of the truth-based truth-table task has repercussions for contemporary theoretical debates. Our results imply the need for a critical re-analysis of all arguments grounded on the likelihood of ‘true’ judgments. If a successful or at least good argument for a position is subsequently found by any participant to be flawed in a way that raises new doubts about the merits of that position, one is obligated to reopen the issue for further consideration and resolution. (Damer, 2005, p. 7).

This is the 'reconsideration principle' Damer (2005) advanced in his work on "Attacking faulty reasoning: A practical guide to fallacy-free arguments".
Both the true/not-false effect (Experiment 1) and the 'true' vs. 'neither true nor false' effect (Experiment 2, 3) are consistent with Johnson-Laird and Byrne (2002). Even stronger, they are predicted by the theory given the theory's take on the so-called "paradoxes of material implication" (e.g. It is true that London is not in France; hence it is true that if London is in France, then pigs can fly). When people evaluate FF<not-A_not-C> as making if A then C true, they are endorsing the following 'paradoxical' argument, i.e., an inductive generalization:

P1: (Not-A and not-C) therefore (if A then C)

Mental-models theory adheres to the principle that people will not endorse inferences like P1 because "they throw semantic information away" (Johnson-Laird & Byrne, 2002, p. 652). Semantic informativeness refers to the number of possibilities rule out by a proposition. The assertion "not-A and not-C" rules out 3 out of 4 possibilities; only the "not-A and not-C" combination is a true possibility. In contrast, the conditional rules out only 1 out of 4 possibilities. "All theories of the conditional agree that the only state of affairs that contradicts [if the cat is happy then she purrs] 1 is a happy cat not purring (TF), and so all other cases are possible" (Evans, 2007, p. 54). That is, P1 involves a loss of semantic information (3/4 vs. 1/4). It follows within mental-models theory -- as a direct consequence of the semantic information-loss principle -- that people will tend not to endorse P1.

The core problem with induction is also that it signifies a loss of semantic information. An inductive inference is a generalization. One goes from a subset of cases to the entire class. "All swans observed up to now are white" is a subset of "all swans are white". That is, "some swans (i.e., all of those who we have observed up to now) are white" does not rule out that in the future we might observe a black swan, whereas this possibility is ruled out by the categorical claim that "all swans are white". Our meta-theoretical verification task (Experiment 1) accordingly demonstrated that people do not tend to throw away semantic information in going from "for some figures (i.e., all that are identified up to now) it applies that if it is circle, then it is red" to "for all figures it applies that if it is a circle then it is red". In short, the use of semantic-informativeness in mental-models theory shows that the induction-problem hypothesis is consistent with the model-theory (which, to apply our insights in a meta-reflective exercise, obviously does not make it a true theory, even though it the induction-problem effects show the theory is less likely to be false).

TT<A_C> cases are mostly judged 'true' even though some will judge it correctly as merely corroborating when given this option (Experiment 2-3). When we consider the induction-problem hypothesis within mental-models theory, we are given a suggestive explanation for the overall high TT[true] ratings. Since TT is explicit within the initial representation it can hardly be labeled 'irrelevant' regarding the rule. That is, while the induction-problem reinforces the initial-model based irrelevancy judgment of false-antecedent cases, it does not for TT cases, which are relevant in virtue of their status as the initial representation.

In summary, mental-models theory can explain the major pattern of results in the truth-table tasks. The initial-model principle, including the implications of mental-footnotes about what is false, explains the relatively high irrelevant ratings of false-antecedent cases, FF<not-A_not-C>and FT<not-A_C>, the high false ratings of TT<not-A_C> under both a conditional and biconditional interpretation of the conditional, as well as the non-negligible proportion of false evaluations of FT<not-A_C> under a bi-conditional interpretation. Johnson-Laird and Byrne's (2002) explicit discussion of differences between reasoning about truth versus reasoning about possibilities also shows mental-models theory is consistent with the induction-problem hypothesis and the 'induction-problem' effects we demonstrated in Experiment 1-3. Most interestingly, the mental-model theory's initial-model principle provides a suggestive explanation for a potential problem with the induction-problem hypothesis (i.e., the high true evaluations of TT, as even <A_C> cases do not make the rule true).

In conclusion, for more than 25 years many researchers have presented reasoners with a task in which they could not respond in compliance with the generally accepted thesis that a single case can never make a rule true. Our findings stress the need for a re-evaluation of arguments that make use of truth-table task results without taking into account the unsuitable nature of 'true'. In Experiment 1 we established the new phenomenon that people are more likely to evaluate false-antecedent cases as 'irrelevant' when confronted with the conceptually impossible 'true' response, as compared to a correct 'not-false' evaluation option. In Experiment 2 -- replicated and extended by Experiment 3, we similarly observed a decrease of 'irrelevant' evaluations of false-antecedent cases in favor of a correct evaluation stating that these cases neither make the rule false, nor irrefutably true but at best support the rule. About half the population also preferred this correct evaluation of TT. Any theory will need to account for these new phenomena, independent of whether it lets itself be inspired by our induction-problem hypothesis. We have shown that our new findings undermine old critiques against mental-models theory.

**References**


