A psychological study of bipolarity in the possibilistic framework

Rui Da Silva Neves, Eric Raufaste

A psychological study of bipolarity in the possibilistic framework

Rui da Silva Neves
Université de Toulouse Le Mirail
D.S.V.P., Maison de la Recherche
5, Allées A. Machado,
31058 Toulouse Cedex 9
neves@univ-tlse2.fr

Raufaste Eric
Université de Toulouse Le Mirail
L.T.C (CNRS-UMR 5551),
Maison de la recherche
5, Allées A. Machado,
31058 Toulouse Cedex 9
raufaste@univ-tlse2.fr

Abstract

The expression "bipolar information" denotes the fact that positive and negative information are not processed in the same way. This study aimed at testing the psychological plausibility of the possibilistic treatment of bipolar information. Rather than testing empirically some predictions derived from possibility theory, we tested whether the possibilistic representation of bipolarity allows understanding previously unexplained psychological data. Psychological works related to Human hypothesis testing furnished the task for this study, namely, the "2-4-6 rule discovery task". Based on an analysis of the status ("impossible", "guaranteed possible", "non impossible") of the states of the world ruled out or allowed by positive or negative pieces of information, we found that the so-called "positive test strategy" (previously seen as irrational) produces on the whole information less uncertain than the concurrent so-called "negative test strategy" (previously seen as the normative one). Further works on the same line of research are pointed out.

Keywords: Bipolarity, possibility theory, hypothesis testing, positive and negative information.

1 Introduction

Bipolar information is pervasive in human mental life and behavior. Many times a day, in order to make a choice, to solve a problem or to understand texts or discourses, we must combine positive and negative preferences; pieces of information that rule out some states of affair with pieces of information that emphasize the plausibility of other states of affair or hypotheses; aspirations (desires) and rejections (aversions)… As such, human mind seems to be expert in the processing of bipolar information, and it seems very likely that a large part of this expertise would not be acquired, but would be the result of the cognitive evolution of the specie. It would be a direct consequence of the mind and affect system(s) architecture(s) [3], [13].

Curiously, very few psychological studies have been devoted to or have integrated considerations on bipolar information. To our knowledge, until recently, the same conclusion holds in artificial intelligence studies. However, Ughetto, Dubois and Prade [17], Dubois, Hajeck and Prade [6], and Van der Torre and Weidert [18] have advocated the idea that there exist situations where the information could be bipolar. This idea has been explored in particular in possibility theory (see also [1], [8]).

The possibilistic treatment of bipolarity proposes that two kinds of information corresponding to different notions of possibility are involved in problem solving tasks. The first one is close to the classical view in logical modeling according to which (i) each piece of information declares some worlds as impossible and (ii) what is possible is what is consistent with what is known [7]. Following this view, every piece of information conjunctively added to a data base acts as negative information, and states of the world that are not ruled out by negative information – not being not impossible - are possible. The second kind of information allows representing what is possible for sure, that is, what is guaranteed possible, since it has been observed, for example. This kind of information is said positive, and pieces of positive information are combined disjunctively. Hence, the possibilistic framework allows to distinguish between impossible (I), guaranteed possible (Δ) and non-impossible (NI) states of the world given the positive or the negative nature of some piece of information. In terms of uncertainty, I is more certain than Δ and Δ is more certain than NI.

In line with previous studies of the psychological plausibility of possibility theory and possibilistic logic [2], [5], [16], this paper addresses the question of the psychological plausibility of the possibilistic representation of bipolarity. More precisely, do human recognize and make use of negative and positive pieces of information in a manner compatible with the possibilistic framework?

A classical way of investigation would be to produce experimental data in order to test the fit between human productions with the possibilistic machinery(ies) ones. The formal outcomes would be collected first given a relevant set of axioms or properties. Next, an experimental device would be constructed so as to immerse participants into conditions able to provoke the kind of mental activity (e.g. judgment, reasoning, decision-making tasks…) under consideration. Human productions (e.g. score values on numerical or qualitative scales; choices; response times)) should be recorded and compared to the formal conclusions previously derived according to some statistical criteria.

In previous works devoted to the psychological study of possibility theory (e.g. [2], [5], [16]) this methodology has allowed to confirm that possibility theory and possibilistic logic have a strong descriptive power with regard to human judgment and reasoning under uncertainty. Such a work could and should be done with tasks involving bipolar information. A less classical way of investigation is to search if possibility theory can illuminate any unexplained psychological phenomena involving bipolar information. This is the way followed in this paper.

Section 2 introduces previous psychological results related to human hypothesis testing. Hypothesis testing tasks are particularly relevant to our objective because they are known as involving positive and negative pieces of uncertain information, and because it has been recognized that participants in experiments exhibits what has been seen initially as a cognitive bias. Section 3 presents an analysis of testers' strategies based on the possibilistic treatment of bipolarity. It is argued that the tests preferred by participants are those that exhibit the best epistemic utility in order to reduce the uncertainty attached to the issues of the tests. Section 4 presents some critics and future extensions to this work.

2 Empirical results in human hypothesis testing

In cognitive psychology, Wason [19] explicitly studied human hypothesis testing (HHT) the first time. Wason's analysis of HHT was based on the falsificationist theory of the philosopher Karl Popper [15] who proposed that theories in science should be exposed to empirical falsification. He considered that it should also apply to to common sense knowledge and human reasoning. Clearly, the falsificationist principle focuses on negative information, that is, pieces of evidence that rule out some worlds as impossible. In the field of philosophy of science, Popper's theory has been opposed to logical positivism [4]. Following logical positivism, statements only have meaning if they are capable of being verified. As such, this theory focuses on confirmation rather falsification, is inductive rather than deductive, and focuses on positive information rather on negative one. Wason and most psychologists have adopted the principle of falsification rather than that of confirmation as a standard for HHT[1].

Wason designed two tasks for the study of human rationality in hypothesis testing: the "selection task" and the "rule discovery task". Here, we will focus on the latter one, but both tasks continue to be replicated and analyzed in recent research. In the rule discovery task, also called "2-4-6 task", participants have the job of discovering a rule governing a combination of three numbers. In Wason's classical experiments, the rule to discover was "three

[1] See however Poletiek [15] for an unified view.

numbers in increasing order of sequence", and the example given was "2-4-6". The task consisted of testing hypotheses until the rule had been discovered. Classically, participants enunciate one by one triples of three numbers, and for each triple, the experimenter responds "yes" or "no". "Yes" means "the triple conforms to the rule", and "no" means "the triple does not conform to the rule". When the tester considers that he has discovered the correct rule, he proposes this rule to the experimenter. The task is over if the rule is the correct one or continues until the correct rule be discovered.

Wason distinguished between a confirming and a falsifying strategy. The confirming strategy consists in giving triples that conform to the participant hypothesis concerning the rule (e.g. "2-6-8" for the rule "three even numbers in increasing order of sequence"). The falsifying strategy consists in giving triples that do not conform to the rule (e.g. "3-5-7"). For Wason, a participant is rational only when he applies a falsifying strategy. Finally, he founded that very few intelligent adults spontaneously seek to falsify their beliefs and concluded that common sense hypothesis testing exhibited a "confirmation bias".

This conclusion has been however discussed by several authors [9], [10], [11], [12], [14], [20] who argued that given the common hypothesis initially formulated by most participants ("three even numbers in increasing order of sequence"), participants cannot refute the induced hypothesis. Indeed, all triples that fulfill this hypothesis also fulfill the target rule ("3 numbers in increasing order of sequence"). As such, the efficiencies of the confirming and falsifying strategies depend on the kind of relation between the participant's hypothesis and the target rule the experimenter has in mind. In addition, it has been pointed out that when participants present a triple that does not conform to their hypothesis and which they do not expect to conform to the rule, they are not seeking a falsification but rather a confirmation. Conversely, apparent confirmation seeking can correspond to falsification seeking (see [12], [14], [20]). Therefore, confirmative and eliminative test strategies can both produce positive and negative information.

These analyses have led to the introduction of two kinds of strategies called "positive test strategy" and "negative test strategy". However, the question remains: why do participants preferentially use positive test strategy rather than negative test strategy?

The main idea is that the human tendency to prefer positive tests against negative ones is based on a maximization of the "epistemic utility" of these two strategies. The informal notion of epistemic utility in this work represents the gain of information, belief or knowledge that each strategy is able to provide in relation to the truth or the falsity of some hypothesis. More precisely, participants in the 2-4-6 task should prefer positive test strategy because, when the nature of the relation between their hypothesis and the experimenter's rule is unknown, its repeated application (in particular at the beginning and at the end of the hypothesis testing process) should overall permit to draw less uncertain conclusions than the negative test strategy.

In the next section, we present an analysis of the epistemic utility of the use by participants of the positive and negative test strategies based on the bipolar possibilistic treatment of positive and negative information. These analyses are conducted given the four possible kinds of relations between the participant's hypothesis and the target rule (this relation being unknown).

3 How bipolarity in possibility theory can explain "the confirmation bias" in the 2-4-6 task?

Consider the general case where the participant (P) has in mind a hypothesis h belonging to the set Hc of the rules that can generate the triple 2-4-6. P can produce two kinds of triples (t) in relation to h: incompatibles (ti) or compatibles (tc). Before any feedback from the experimenter (E), any rule h that leads to at least one triple of the ti kind cannot be said impossible in relation to the target rule, nor it can be said guaranteed possible: it is only non-impossible (NI). After the feedback of E, if t is a ti instance of the target rule R, h is falsified but any hypothetic rule able to generate t is guaranteed possible because one instance of the rule has been observed. If t is not an instance of R, h is not falsified but every rule able to generate t is impossible.

By comparison, before any feedback from E, any rule h belonging to Hc is guaranteed possible - although still uncertain and possibly false - because at least one positive instance of the rule is known, namely the 2-4-6 triple. After a feedback from the experimenter, any confirmation does not change the status of h since this rule is already guaranteed possible. Yet, h receives more inductive support. If h is falsified, every rule able to generate t is also impossible.

Let us consider now the consequences of the "yes" or "no" feedback given the kind of relation between the participant's hypothesis and the target rule.

Let U be the set of all possible triples of 3 numbers, Tr be the set of triples compatible with the target rule (e.g. "3 numbers in increasing order of sequence"), and Th be the set of triples compatible with the hypothesis rule under consideration by the participant (e.g. "three even numbers in increasing order of sequence").

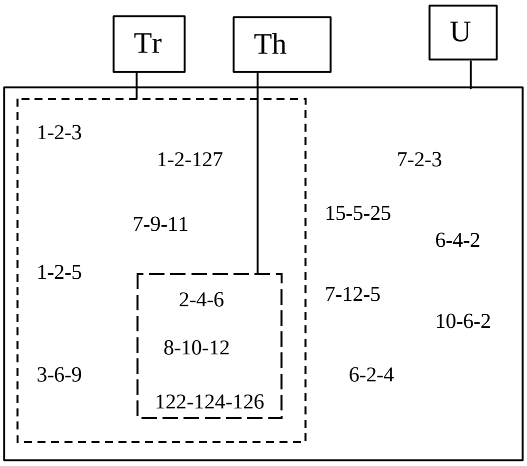

Figure 1: Case where the participant's hypothesis (h) is included in the target rule (R).

First, consider the case represented in figure 1, where Th ⊂ Tr. Let us also consider the case where the participant applies the positive test strategy and proposes for example the triple 8-10-12 (which belongs to both Th and Tr). In such a case, the "no" feedback from the experimenter will never occur, because every triple t in Th is also in Tr. The only possible response being "yes", every hypothesis that generates t is guaranteed possible (Δ). Consider now the case where P applies the negative test strategy. If the triple t belongs to Tr, it will receive a positive feedback, which falsifies the participant's hypothesis h and establishes the impossibility of every hypothesis compatible with t (note that this response carries also positive information related to Tr). If t does not belong to Tr, the feedback is negative and it can be concluded that the hypotheses belonging to the set Ht of the rules able to generate t are NI. For convenience, in the sequel of the paper, we will say "Ht is NI", "Ht is I" and "Ht is Δ". It is to be noted that we cannot concluded that Ht is guaranteed possible, because there is no positive information per se in this case.

Second, consider the case represented in figure 2, where Th ∩ Tr ≠ ∅ and Th ⊄ Tr.

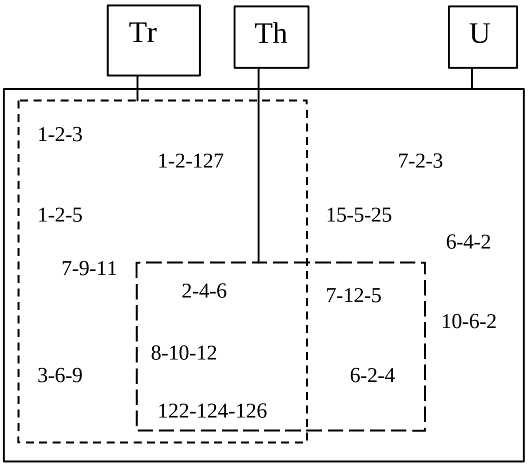

Figure 2: Case where the target rule (R) partially overlaps the participant's hypothesis (h).

When applying the positive test strategy, P receives a positive or a negative feedback depending on whether t belongs to Tr or not. If t belongs to Tr, the feedback is positive and Ht is Δ. If t does not belong to Tr, the feedback is negative, and Ht is impossible. When P applies the negative test strategy, if t belongs to Tr, the feedback is "yes", and Ht is impossible. If t does not belong to Tr, Ht is NI.

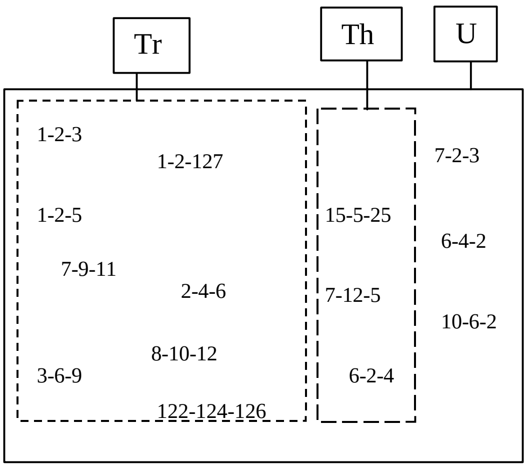

Figure 3: Case where the target rule (R) does not overlap the participant's hypothesis (h).

Third, consider the case represented in figure 3, where $Th \cap Tr = \emptyset$. When P applies the positive test strategy, it can only lead to a negative feedback. Then, participant's hypothesis is falsified and Ht is I. When P applies the negative test strategy, if t belongs to Tr, the feedback is "yes", which carries a positive information about Tr but falsifies h. If t does not belong to Tr, h is falsified and Ht is NI.

Finally, consider the case represented in figure 4 where $Tr \subset Th$ and $Tr \neq Th$. When P applies the positive test strategy, if t belongs to Tr, the feedback is positive, h is confirmed, and Ht is Δ. If t does not belong to Tr, the feedback is negative, and Ht is I. When P applies the negative test strategy, the feedback can only be negative, and Ht is NI.

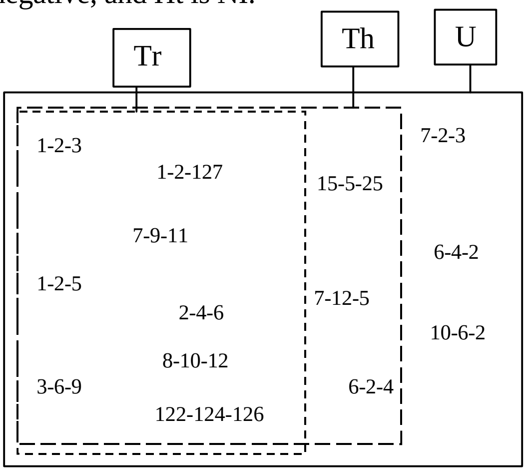

Figure 4: Case where the target rule (Tr) is included in the participant's hypothesis (h).

Assuming that, in one hand, the negative and positive cases under consideration have the same likelihood, and in the other hand, the four kinds of relations between Th and Tr have also the same likelihood, the examination of table 1 shows that a positive test provides either information guaranteed possible (3 cases out of 6) or impossible (3 cases out of 6), while a negative test provides either an information non-impossible (3 cases out of 7) or impossible (4 cases out of 7). Therefore, given that I is more certain than Δ, and Δ is more certain than NI, overall, it appears that the application of the positive test strategy leads to conclusions less uncertain than the negative test one. Moreover, whatever the kind of the relation between the participant's hypothesis and the target rule, a negative test can potentially produce a less certain conclusion in 3 cases out of 4. By contrast, a positive test strategy can potentially produce a conclusion at least as certain as the negative test strategy in 3 cases out of 4, including the one missing previously. In a way, participants applying a negative test strategy could be seen has more optimistic than the ones applying a positive test strategy. Indeed, negative test strategy produces more often impossible hypotheses, that is the kind of information the more certain. However, this very same strategy produces also more often non-impossible hypotheses, that is the kind of information the less certain.

Table 1: Plausibility of the set Ht of rules compatible with any triple t in function of the kind of test (positive or negative) and the kind of relation between the participant's hypothesis (H) and the experimenter's rule (R).

Relation	Positive test	Negative test
$Th \subset Tr$	Yes: Δ No: --	Yes: I No: NI
$Th \cap Tr \neq \emptyset$ and $Th \not\subset Tr$	Yes: Δ No: I	Yes: I No: NI
$Th \cap Tr = \emptyset$	Yes: -- No: I	Yes: I No: NI
$Tr \subset Th$ and $Tr \neq Th$	Yes: Δ No: I	Yes: -- No: I

"Δ" stands for "guaranteed possible", "I" for "impossible", "NI" for "non impossible", and "—"for "non-tractable".

4 Conclusion

This study has been conducted with the goal of testing the psychological plausibility of the possibilistic treatment of bipolar information. Rather than testing empirically (as in previous works) some predictions derived from possibility theory, we aimed at showing that the bipolar possibilistic representation allows to explain previously unexplained psychological data. Human hypothesis testing (involving the use of positive and negative information, and exhibiting some unexplained results), furnished the task for this study, namely, the 2-4-6 rule discovery task.

Based on an analysis of the status (impossible, guaranteed possible, non impossible) of the information given ("yes" or "no" response) as a function of the participant's testing strategy (positive or negative) and the kind of relation between the participants' hypothesis and the target rule, we have found that the positive test strategy produces on the whole less uncertain information than the concurrent negative test strategy (classically seen as the normative one). This result emphasizes the psychological plausibility of the possibilistic framework. However, we assumed that the negative and positive cases under consideration have the same likelihood. Further work should be based on the assumption that it is not the case. Moreover, it could be of particular interest to study whether participants are able to infer the kind of relation existing between the hypothesis h under consideration and the target rules R, and if they are able to adapt their testing strategies as a function of the four possible kinds of relationship between h and R.

In addition, further experimental studies should be conducted in order to verify that the certainty hierarchy we have established between the kinds of pieces of information (I, Δ and NI) is psychologically relevant for subjects in the considered contexts. Moreover, the same methodology should be applied in the context of another hypothesis testing tasks, in particular the so called selection task (see [14] for example), and further work should also be based on a more classic experimental approach in the line of previous works [2], [5], [16].

Finally, in this paper, our analyses and conclusions concerned both human hypothesis testing and the possibilistic framework. We used the possibilistic framework to analyze the hypothesis testing task thereby showing the potential interest of a well-known result in the psychology literature: people use more often the positive rather than the negative test strategy. However, we did not try to formalize human hypothesis testing in a possibilistic framework per se. Yet, our understanding of human reasoning might benefit greatly from such a work, which in turn should allow testing deeply the possibilistic treatment of polarity.

Acknowledgements

This research was supported by the CNRS AS119 (Action Spécifique sur la Bipolarité) and TCAN Program: "Traitement des connaissances, apprentissage et NTIC ".

References

[1] S. Benferhat, D. Dubois, S. Kaci, H. Prade (2002). Bipolar Possibilistic representation. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI 2002)*, Edmonton, Canada, Aug. 1-4, Morgan Kaufmann Publisher, San Francisco, 45-52.

[2] S. Benferhat, J.F. Bonnefon and R. Da Silva Neves (2003). An overview of possibilistic handling of default reasoning: an experimental study. *Synthese (in press).*

[3] Cacioppo, J. T., & Berntson, G. G. (1999). The affect system: Architecture and operating characteristics. *Current Directions in Psychological Science*, 8, 133-137.

[4] R. Carnap (1936-1937). Testability and meaning. *Philosophy of Science*, 3, 420-468; 4, 1-40.

[5] R. Da Silva Neves, J.F. Bonnefon and E. Raufaste (2002). An empirical test of patterns for nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence*, 34: 107-130. Kluwer Academic Publishers.

[6] D. Dubois, P. Hajeck and H. Prade (2000). Knowledge-driven versus data-driven logics. *Journal of Logic, Language and Information*, 9, 65-89.

[7] D. Dubois, L. Lang, and H. Prade (1994) Possibilistic logic. In: *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 3 (D.M. Gabbay, C.J. Hogger, J.A. Robinson, D. Nute, eds.), Oxford University Press, 439-513.

[8] D. Dubois, H. Prade and P. Smets (2001). "Not impossible" vs. "guaranteed possible" in fusion and revision. In S. Benferhat and P. Besnard (Eds.), *ECSQARU 2001*, 522-531, Springer-Verlag Berlin Heidelberg.

[9] J. McDonald (1990). Some situational determinants of hypothesis testing strategies. *Journal of Experimental Social Psychology*, 26, 255-274.

[10] Y. Kareev, N. Halberstadt and D. Shafir (1993). Improving performance and increasing the use of non-positive testing in a rule-discovery task. *The quarterly journal of experimental psychology*, 46A, 4, 729-742.

[11] J. Klayman (1995). Varieties of confirmation bias, in J. Busemeyer, D. Medin, R. Hastie (Eds.), *Decision Making from a Cognitive Perspective, The Psychology of Learning and Motivation*, vol. 32, 385-418, Academic Press.

[12] J. Klayman and Y. W. Ha (1987). Confirmation, Disconfirmation, and Information in Hypothesis testing. *Psychological Review*, Vol. 94, 2, 211-228.

[13] J. T. Larsen, P. McGraw & J. T. Cacioppo (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology,* 81, 684-696.

[14] F. Poletiek (2001). Hypothesis-testing behaviour. *Psychology Press*.

[15] K. R. Popper (1935). *Logik der forschung*, Vienne, Springer Verlag. (Edition Française: *La logique de la découverte scientifique*, Paris, Payot, 1973).

[16] E. Raufaste, R. Da Silva Neves and C. Mariné (2003). Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence*, 148, 197-218.

[17] L. Ughetto, D. Dubois and H. Prade (2004). Implicative and conjunctive fuzzy rules. A tool for reasoning from knowledge and examples. In *Proceedings of the 16th International Conference on Artificial Intelligence (AAA-99)*, Orlando, July 18-22, 1999, Morgan and Kaufman, 214-219.

[18] L. van der Torre and E. Weydert (2001). Parameters for utilitarian desires in a qualitative decision theory. *Applied Intelligence*, 14:285-301.

[19] P.C. Wason (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.

[20] N. E. Wetherick (1962). Eliminative and enumerative behavior in a conceptual task. *Quarterly Journal of Experimental Psychology*, 14, 246-249.