Psihologijske teme, 30 (2021), 3, 447-466

Izvorni znanstveni rad UDK: 159.955.6

https://doi.org/10.31820/pt.30.3.3 519.2

Is Automation of Statistical Reasoning a Suitable Mindware in a Base-Rate Neglect Task?

Klara Rapan[1] and Pavle Valerjev[2]

[1] University of Zadar, Department of Teachers and Preschool Teachers Education, Zadar, Croatia

[2] University of Zadar, Department of Psychology, Zadar, Croatia

Abstract

Until recently, studies within the dual-process approach were mainly focused on group differences in processing, and individual differences were neglected. However, individual differences have proven to be a significant factor in conflict detection efficiency and the overall success in base-rate neglect and similar tasks. This should be taken into consideration within the framework of the Hybrid Model of Dual Processing. New tendencies in the development of this model have focused attention on the degree of mindware instantiation as a predictor of base-rate neglect task efficiency. This study aimed to examine the relationship between mindware and base-rate neglect task efficiency and to test and explore the relationship between base-rate response frequency and conflict detection efficiency and the degree of mindware instantiation. All participants solved base-rate neglect tasks, made judgments of confidence in their responses, and solved the Statistical Reasoning Test, Cognitive Reflection Test and Numeracy Scale. We used the Statistical Reasoning Test as a measure of mindware instantiation. The degree of mindware instantiation was found to be the only significant predictor of base-rate neglect task efficiency and the results showed that participants with a higher degree of mindware instantiation generally made more base-rate responses. No correlation was found between the degree of mindware instantiation and conflict detection efficiency. These findings support the hypothesis that the power of logical intuition depends on the individual's degree of mindware instantiation. Therefore, the results of this research indicate the importance of further research into the role of statistical reasoning in base-rate neglect task efficiency. However, we discuss that there are some methodological limitations in this research which might explain why the degree of mindware instantiation had no relationship with conflict efficiency.

Keywords: mindware, statistical reasoning, base-rate neglect, conflict detection, Hybrid Model of Dual Processing

✉ Klara Rapan, Department of Teachers and Preschool Teachers Education, University of Zadar, Ulica dr. Franje Tuđmana 24 I, 23000 Zadar, Croatia. E-mail: *krapan@unizd.hr*

Introduction

The traditional dual process approach refers to complementary Type 1 (T1) and Type 2 (T2) thought processes (Evans, 2008, 2010; Evans & Stanovich, 2013; Kahneman, 2011). T1 processes are typically described as fast, emotionally charged, heuristic-based, implicit, and error-prone, while T2 processes are typically described as slower, analytical, explicit, and more accurate than T1 processes (Bago & De Neys, 2020; Kahneman, 2011; Pennycook et al., 2015; Thompson, 2009). The interaction between these two processes has traditionally been described as *default-interventionist* (Bago & De Neys, 2020; Pennycook, 2018), primarily because it was assumed that T2 processes are triggered shortly after T1 processes provide an incorrect response or output. Therefore, T2 processes were considered to be corrective in nature. However, traditional dual process theories have proven to be problematic. Some of the most important and researched problems are the following: 1.) even individuals who have provided the wrong answers detected conflict (e.g. De Neys et al., 2013; De Neys & Glumicic, 2008; Gangemi et al., 2015; Pennycook et al., 2014; Stupple & Ball, 2008); 2.) T1 processes sometimes provide correct responses (e.g. Bago & De Neys, 2017; Newman et al., 2017); 3.) it is unclear which processes are in charge of detecting conflict between T1 and T2 processes (Pennycook et al., 2015). Therefore, a new dual processing model was recently introduced, based on somewhat different settings - *the Hybrid model of dual processing*.

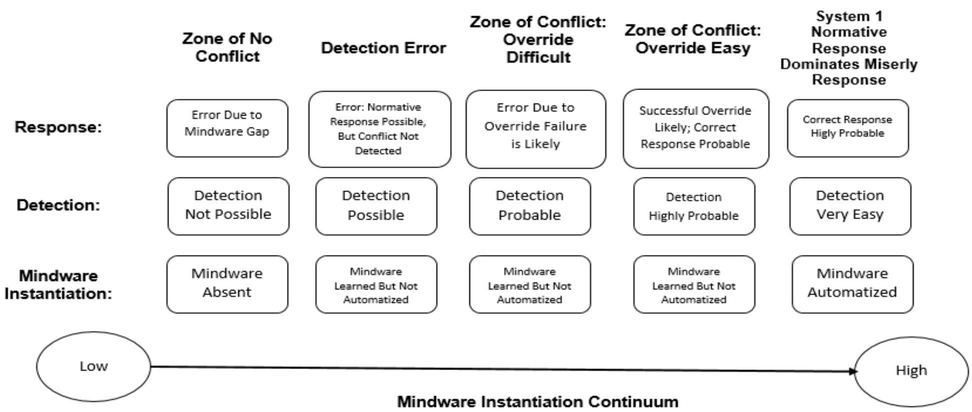
The Hybrid model of dual processing was presented by Pennycook et al. (2015) as an extension of the *Logical intuition model* presented by De Neys (2012). One of the hybrid models is a *Three-stage dual-process model of analytic engagement* (Pennycook et al., 2015), according to which the existence of two thought processes is not in doubt, but the cause of the activation of T2 processes should be considered. For clarity, the authors divide the process of reasoning into stages. In the first stage of the model, the presented problem causes the generation of several different initial responses in one's mind, which can either be heuristic intuition or logical intuition. Heuristic intuition and logical intuition are both part of T1 processes and are distinguished by the fact that logical intuition, unlike heuristic, is based on the automatically triggered implicit knowledge of a problem. On the other hand, heuristic intuition is what is traditionally known as System 1 (or S1) processes, that is, intuition-based on heuristics and biases, which, therefore, often results in poor decision making (De Neys, 2014; Pennycook et al., 2015). These intuitive answers differ in the fluency with which they come to one's mind. Therefore, in the second stage of this model, a conflict might be detected if there are several responses with a similar or same fluency (Pennycook et al., 2015), due to the arousal which normally occurs as a result of conflicting responses (De Neys, 2014). If the conflict is successfully detected, there are two possible outcomes in stage three - either to choose a different answer or to confirm the initial one. If no conflict detection occurs, an initial response with the highest answer fluency is given (Pennycook et al., 2015).

However, what has remained unexplored is the source of logical intuition. Where does logical intuition come from? The concept of *mindware* described by Keith Stanovich (2018) provides various potential answers to this question. According to Stanovich and West (2008), mindware refers to the degree of the automation of knowledge relevant to a given situation. For example, the relevant mindware for riding a bicycle would be the knowledge of the procedure itself, and success in riding a bicycle would be the result of the degree of the automation of that knowledge, which Stanovich (2018) simply refers to as *mindware instantiation*. Since mindware is not reserved exclusively for expert skills (chess players are common example), it could be applied to any ability or knowledge (De Neys & Pennycook, 2019).

According to Stanovich (2018), the degree of mindware instantiation is strongly positively correlated with the likelihood of successful conflict detection and rejection of an incorrect response (Figure 1). In theory, participants will show less confidence in their incorrect responses (products of the T1 process) if they have strong normative mindware, that is, response confidence should increase in a situation when the intuitive response is in line with normative mindware. In contrast to this, response confidence should be lower when the intuitive incorrect response is not in line with normative mindware. Furthermore, the incorrect response is most likely to occur if there is no corresponding mindware to solve the problem, if conflict is not detected or if override failure occurs (Stanovich, 2018).

Figure 1

Processing States on the Mindware Continuum



Note. Adapted from "Miserliness in human cognition: The interaction of detection, override and mindware" by K. E. Stanovich (2018), *Thinking & Reasoning*, *24*(4), p. 435.

Mindware (Stanovich, 2018) and logical intuition (Pennycook et al., 2015) share many similarities - they both operate on an unconscious (automated) level, both are based on rules derived from logic, both, if sufficiently automated and fast, lead to accurate outputs. This is clear from the fact that the tasks commonly used to test a Hybrid model of dual processing consist of fairly simple logical principles and rules that can be automated through education, repetition in daily life or practice (De Neys, 2012). Consider the following base-rate neglect problem:

> Person X is smart.
>
> Person X was selected at random from a group that consists of 100 scientists and 700 butchers.
>
> It is more probable that Person X is:
>
> 1. A scientist
> 2. A butcher

Once one is confronted with the problem, intuitive heuristic and logical responses are automatically generated in one's mind. The answer generated by logical intuition is answer number 2 (because it takes into account statistical principles and probability), while the answer generated by heuristic intuition is answer number 1 (because it is based on the mere similarity between the person and their stereotypical description). Therefore, one can assume that the relevant mindware in the context of the base-rate neglect task is statistical reasoning, because the correct response is based on an automated understanding of statistical principles. If the degree of mindware instantiation (or logical intuition) is high, there will be no conflict between heuristic and logical responses and the person will provide the correct (logical) response, and response confidence will be higher as well.

In base-rate neglect tasks, there are usually a few variables worth noticing. First, attribute/ratio congruency – a base-rate neglect task can either be congruent or non-congruent. Congruent base-rate neglect tasks are instances when a person's attribute is congruent with the sample ratio, for example: "Person X is tall. Person X was selected at random from a group which consists of 995 basketball players and 5 teachers." Non-congruent base-rate neglect tasks are instances when a person's attribute is not congruent with the sample ratio, for example: "Person Y is funny. Person Y was selected at random from a group which consists of 995 students and 5 comedians.". Second, base-rate neglect tasks can differ in sample ratios. The ratios commonly implemented in research are either extreme (e.g. 5/995) or moderate (e.g. 300/700), yet according to Dujmović and Valerjev (2018) these ratios have been extended to include high (e.g. 200/800) and low (e.g. 450/550) ratio. It is widely considered that higher ratios make conflict detection more pronounced (e.g. Bago & De Neys, 2020; De Neys, 2014; Pennycook et al., 2012).

In a recent study, Šrol and De Neys (2021) explored the role of mindware instantiation, cognitive reflection, numeracy, thinking dispositions and cognitive ability as potential predictors of the overall accuracy on reasoning problems (syllogistic reasoning tasks, base-rate neglect problems, conjunction fallacy items and the bat-and-ball problem). The mindware instantiation index was computed using neutral versions of all reasoning problems. Mindware has been found to be the strongest independent predictor of reasoning problems efficiency (Šrol & De Neys, 2021). Also, this research tested the correlation between the aforementioned predictors and conflict detection efficiency. Conflict detection efficiency was described as the difference in the latency between correct and incorrect responses and also, as a difference in confidence in correct response and incorrect response. The degree of mindware instantiation has not been found to be a significant predictor of response latency, but is found to be a significant positive predictor of response confidence (Šrol & De Neys, 2021).

Taking these considerations into account, the aim of this study was to examine the relationship between mindware instantiation and base-rate neglect task efficiency and conflict efficiency within the Hybrid Model of Dual Processing. Given that the relevant knowledge for success in a base-rate problem is assumed to be statistical reasoning, the mindware measure used in this study is a newly constructed *Statistical Reasoning Test* (SRT) (Rapan & Valerjev, 2020). In accordance with the above, we present the following hypotheses:

1. We predict that there will be more correct base-rate responses in congruent and extreme base-rate tasks.
2. We predict that judgments of confidence will be the highest for base-rate responses in congruent tasks, and that judgments of confidence will be lower for base-rate responses in congruent tasks and stereotypical responses in non-congruent tasks.
3. We predict that mindware instantiation will be a significant predictor of base-rate task efficiency and conflict detection efficiency, independent of education, cognitive reflection and numeracy.

Method

Participants

A total of 258 participants began responding, but only 43.80% of them completed the survey. The final sample consisted of 113 participants (77.88% women), in the age range from 16 to 40 (M = 23.54, SD = 3.94). Participants were recruited online, via social media. The sample was predominantly represented by students (66.37%), followed by participants with secondary education (14.16%), upper-secondary education (12.39%), university degree (5.31%) and finally, high-school students (1.78%).

Materials

Base-Rate Neglect Tasks

The base-rate neglect tasks used in this study were originally developed by Dujmović and Valerjev (2017, 2018). All tasks differed in sample ratios and attribute/ratio congruence in such a manner that 4 different conditions were created, as shown in Table 1. The size of the ratio varied as either extreme (997/3, 996/4, 995/5, 994/6, 993/7) or moderate (700/300, 710/290, 720/280, 730/270, 740/260). The means of varying the ratio were taken from Pennycook et al. (2015), but instead of three, five different ratios were used at each level in order to avoid repetition. The attribute was extremely typical for one of the groups in the sample, but not exclusive.

Table 1

Examples of Four Different Base-Rate Neglect Problems

Condition	Attribute	Ratio
Extreme ratio/congruence	Person A is educated.	Group consists of 997 professors and 3 cashiers.
Moderate ratio/congruence	Person B is romantic.	Group consists of 720 poets and 280 surgeons.
Extreme ratio/incongruence	Person C is well-read.	Group consists of 994 welders and 6 professors.
Moderate ratio/incongruence	Person D is famous.	Group consists of 700 physical therapists and 300 musicians.

As can be seen from Table 1, the first condition represents the highest level of congruence because both the sample ratio and the attribute point to the same response. The situation is similar in the second condition, as both the person attribute and the ratio (though less extreme) suggest the same response. In the third and fourth condition the ratio and the attribute are incongruent - they indicate different answers and therefore, these two conditions represent a low-level congruence, that is, high levels of conflict.

Statistical Reasoning Test

The Statistical Reasoning Test (*SRT*) (Rapan, 2019; Rapan & Valerjev, 2020) is designed to measure the automation of statistical reasoning, that is, the speed and accuracy of the comprehension of statistical information and ideas such as probabilities and ratios. We used SRT as a measure of mindware instantiation. The correlations between SRT and other measures of similar constructs such as numeracy and cognitive reflection, that represent the criterion validity of the test, are presented in Table 4. SRT is moderately positively correlated with Numeracy Scale and

Cognitive Reflection Test, and weakly correlated with base-rate neglect efficiency. Despite the relatively weak correlation with base-rate neglect efficiency, the correlation coefficient of SRT and base-rate response frequency is higher than the correlation of Cognitive Reflection Test and base-rate response frequency, and Numeracy Scale did not show significant correlation with base-rate task efficiency.

The test consists of 11 multiple-choice items in Croatian, which vary in difficulty (e.g. “A box contains 4 white, 6 blue and 8 black balls. A single ball is drawn. What is the probability that the ball is blue?”; “Two fair coins are tossed simultaneously. What is the probability of obtaining two ’tails’?”; “Water to air ratio in vessel A is 4:3. Water to air ratio in vessel B is 5:4. Which vessel contains more water, if the two vessels are the same size?”). There are four possible answers presented after each question and only one answer is correct. The time allowed to respond is limited independently for each item, depending on the average solving time for each item in the initial application of the test. Therefore, each item has a different time limit, and the time limit for the whole test is 4 minutes and 50 seconds in total. Each correct answer has a value of one point. An overall test result is formed as the sum of all correct answers. Efficiency in the SRT test was used as a measure of mindware in this study.

Cognitive Reflection Test

As a measure of cognitive reflection, a version of the three-item *Cognitive Reflection Test* (*CRT*) was used. This test is designed to measure the tendency to override an incorrect response and to engage in further reflection that leads to the correct response (Toplak et al., 2011). The original version of the CRT test consists of three items (Frederick, 2005), four more items were added later (Toplak et al., 2014). Due to the participant’s potential familiarity with the first three items of the test, the other version of the test without the first four items was translated and used in this study. One of the items was multiple-choice, while the remaining two items required a typed answer. Correct answers had the value of one point, and the overall test result was formed as a sum of all correct answers.

Numeracy Scale

As a measure of numeracy, the Numeracy Scale (Schwartz et al., 1997) was used. The *Numeracy Scale* is designed to measure the ability to perform mathematical tasks and to understand and deal with numbers (Fagerlin et al., 2007). The scale consists of three items that need to be answered by typing the correct answer. As was the case with previous measures, each correct answer had the value of one point, and the total score was formed as the sum of all correct answers.

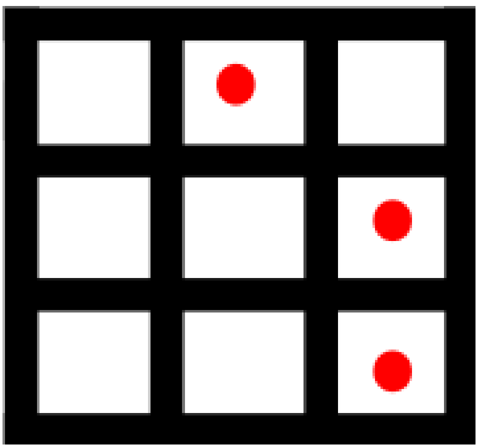
Procedure

The study was conducted online. All participants were clearly informed that their participation in the study would be completely anonymous and that they were free to opt-out of the study at any time. In order to recruit a large sample of participants and to keep participants motivated, they were informed that one random participant will be given a 300 HRK cash prize after the study was finished. The study consisted of two parts. In order to match the results from both parts and to enable the announcement of the prize game winner, participants were instructed to write the same code name at the beginning of each part of the study. Then, the study began. All participants took part in both parts of the study. The first part of the study was designed in PsychoPy v3.1.5. (Peirce et al., 2019). It consisted of 20 trials, preceded by three exercise trials. Each trial consisted of the display of a dot-pattern, a base-rate problem, judgment of confidence and a dot pattern memory task. Correct base-rate response and stereotypical response frequency and judgment of confidence were measured.

Furthermore, we will describe the order in which the stimuli were displayed on the screen. As a way of disabling T2 process activation, it was required that the participant's cognitive load be increased. Here we followed the procedure used by Bago and De Neys (2020). Previous to each base-rate problem, participants were presented with a different dot-pattern (Figure 2). It was required that the pattern be remembered and, after providing a response to a base-rate problem, participants were presented once more with a dot-pattern and their task was to determine whether the pattern was the same or different from the one originally presented, by pressing the appropriate button (left - same; right - different). The patterns were presented to participants for 3000 ms before each base-rate problem.

Figure 2

An Example of Dot-Pattern Used as Cognitive Load



Next, we will describe the presentation of base-rate neglect stimuli on screen. First, information about the person drawn from the sample was displayed for 3000 ms, then the sample and ratio for 4000 ms, while the previous information about the person's attribute remained on the screen. Finally, the question about which group the person was more likely to belong was answered by pressing the left or right button. The interval to answer the question was also limited to 3000 ms, as shown in the example below:

Person A is educated.	3000 ms
Person A is educated. Group consists of 720 dancers and 280 surgeons.	4000 ms
Person A is educated. Group consists of 720 dancers and 280 surgeons. It is more probable that Person A is: 1. A dancer 2. A surgeon	3000 ms

The limited interval for showing the problem and answering the question served as another way of disabling potential T2 process activation, and this was to ensure that the answer was the result of intuitive T1 processes. Also, conflict detection efficiency was measured. The difference between judgment of confidence in correct base-rate response in congruent tasks and incorrect stereotypical response in incongruent tasks was used as a measure of conflict detection (*conflict detection index*). Confidence assessment is widely used in dual process approach (e.g. Ackerman & Thompson, 2017; Bago & De Neys, 2020; De Neys & Glumicic, 2008; Stupple & Ball, 2008; Thompson et al., 2018; Valerjev & Dujmovic, 2017). Participants were asked to assess their confidence in a given response immediately after making the decision on a base-rate problem, on a six-point scale from 50 to 100%. Base-rate neglect tasks were presented in random order to each participant.
At the end of the first part of the study, participants were redirected to the second part of the study which was designed in PsyToolkit v2.6.1 (Stoet, 2010, 2017) in which they solved the Statistical Reasoning Test, Numeracy Scale and the Cognitive Reflection Test. Finally, all participants completed a short questionnaire on their demographic characteristics (age, gender, and education).

Results

The descriptive parameters for all variables used in the study are presented in Table 2 and Table 3.

Table 2

Mean Values, Standard Deviations and Cronbach's Alpha of Overall Scores on The Statistical Reasoning Test, Cognitive Reflection Test and Numeracy Scale (N = 113)

	M	*SD*	α
Statistical Reasoning Test	5.84	2.23	.58
Cognitive Reflection Test	1.15	1.10	.64
Numeracy Scale	1.67	1.03	.50

As seen in Table 2, Cognitive Reflection Test and Numeracy Scale mean scores are similar to those in previous research (e.g. Frederick, 2005; Haigh, 2016; Lipkus et al., 2001; Pennycook et al., 2015). The Cronbach alpha internal consistency coefficient of the Statistical Reasoning Test detected in previous research is .73 (Rapan & Valerjev, 2020), and in this study a slightly lower internal consistency was detected: the Cronbach alpha was .58. Cognitive Reflection Test and Numeracy Scale also displayed slightly lower internal consistencies than in previous research (Frederick, 2005; Lipkus et al., 2001).

Table 3

Mean Values and Standard Deviations of Base-Rate Response Frequency, Judgments of Confidence of Stereotypical Response and Judgments of Confidence of Base-Rate Response (N = 113)

Condition	Base-rate response frequency		Judgment of confidence of stereotypical response (%)		Judgment of confidence of base-rate response (%)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Moderate ratio/congruence	4.55 (91%)	0.78	-	-	89.43	9.55
Extreme ratio/congruence	4.68 (93.6%)	0.67	-	-	92.34	8.97
Moderate ratio/incongruence	1.15 (23%)	1.40	87.51	10.93	83.11	14.71
Extreme ratio/incongruence	1.82 (36.4%)	1.74	84.91	13.01	86.21	12.65

Note. Mean proportions of correct responses are in brackets.

Mean proportions of correct base-rate responses in conflict and no-conflict tasks with moderate and extreme ratios, and incorrect stereotypical responses in conflict tasks are similar to those in previous research (e.g. Pennycook et al., 2012). Also, judgments of confidence that can be seen in Table 3 reflect a typical difference in confidence in responses in conflict and no-conflict tasks (Bago & De Neys, 2020; Šrol & De Neys, 2021).

In order to examine the impact of congruence and ratio on frequency of base-rate responses, a 2x2 ANOVA was conducted. The results indicate the significant main effects of congruence ($F = 1105.86$, $p < .01$, $\eta_p^2 = .81$) and ratio ($F = 18.32$, $p < .01$, $\eta_p^2 = .23$), as well as the interaction of congruence and ratio ($F = 8.23$, $p < .01$, $\eta_p^2 = .11$). Post-hoc analysis showed that more base-rate responses were given to tasks with a congruent attribute and ratio, as well as to tasks with an extreme ratio. But there was no difference in base-rate response frequency depending on ratio when the task consisted of congruent attribute and ratio.

Next, we tested the difference in confidence judgments between different responses (1. base-rate response in congruent task, 2. base rate response in incongruent task, 3. stereotypical response in incongruent task) ($F = 730.5$, $p < .01$, $\eta_p^2 = .19$). Post-hoc analysis showed that base-rate responses in congruent tasks had higher confidence judgments than both stereotypical and base-rate responses in incongruent tasks. There was no significant difference between confidence judgments for base-rate and stereotypical responses in incongruent tasks.

The main purpose of this research was to determine the effect of mindware on predicting success in base-rate neglect tasks, as well as the efficiency of conflict detection. In line with previous research (Frey et al., 2018; Šrol & De Neys, 2021), we used the difference between judgment of confidence in correct base-rate response in congruent tasks and incorrect stereotypical response in incongruent tasks as a measure of conflict detection (*conflict detection index*). Only participants who got at least one item correct (base-rate response in congruent tasks) and one item incorrect (stereotypical response in incongruent tasks) were taken into consideration. Also, only participants who had higher judgments of confidence for base-rate response in congruent tasks than for stereotypical response in incongruent tasks were considered for further analyses. Among those participants, the difference between confidence in correct base-rate response in congruent tasks and incorrect stereotypical response in incongruent tasks was statistically significant ($t = 8.36$, $p < .01$, $df = 58$). The average difference between confidence in correct base-rate response in congruent tasks and incorrect stereotypical response in incongruent tasks was 7.93 ($M = 7.93$, $SD = 7.29$). Before conducting multiple regression analyses, the intercorrelations among all used variables were calculated (Table 4).

Table 4 displays the intercorrelations among the 6 variables used in this study. Statistical reasoning is significantly positively correlated with cognitive reflection, numeracy, frequency of total base-rate responses and frequency of base-rate responses in incongruent tasks with moderate ratios. Cognitive reflection is significantly positively correlated with statistical reasoning, numeracy and frequency of total base-rate responses. Due to lower internal consistencies of Statistical Reasoning Test, Cognitive Reflection Test and Numeracy Scale in this study, disattenuating coefficients among the tests were calculated. As seen in Table 4, disattenuating coefficients are significantly higher and indicate a moderate correlation

Table 4

Correlations Between Statistical Reasoning, Cognitive Reflection, Numeracy and Frequency of Base-Rate Responses (N = 113)

	SRT	CRT	NS	Total BR frequency	BR frequency for noncongruent tasks with moderate ratios	Conflict detection index
SRT	-					
CRT	$.32^{**}$ ($.52^{**}$)	-				
NS	$.26^{**}$ ($.48^{**}$)	$.39^{**}$ ($.68^{**}$)	-			
BR frequency	$.21^{*}$	$.19^{*}$	.18	-		
BR frequency for noncongruent tasks with moderate ratios	$.23^{**}$	.16	.12	$.63^{**}$	-	
Conflict detection index	.20	.11	.23	$.36^{**}$	.16	-

Note. SRT = Statistical Reasoning Test; CRT = Cognitive Reflection Test; NS = Numeracy Scale. Disattenuating correlation coefficients among the tests are in brackets. Conflict detection index shows correlations only on a subgroup of participants who successfully detected conflict (N = 59); other correlations are calculated on all participants (N = 113). $^{**}p < .01$; $^{*}p < .05$.

between measures, which suggests that correlations among measures are influenced by salient measurement error. Next, conflict detection index is significantly positively correlated only with the frequency of total base-rate responses. The frequency of correct base-rate responses in incongruent tasks with moderate ratios is significantly positively correlated with statistical reasoning and with total frequency of base-rate responses in all tasks.

For the final analysis, we conducted a hierarchical multiple regression as shown in Table 5. We tested the unique predictive contribution of statistical reasoning in explaining the variance of base-rate response choices frequency in noncongruent tasks with moderate ratios. It is known that extreme base rates present a stronger cue for logical intuition, as already reported in this research, and in some previous research (e.g. Bago & De Neys, 2020; Pennycook et al., 2015). Therefore, we only used frequencies of base-rate responses in noncongruent tasks with moderate ratios in order to test the predictive contribution of mindware, because we predict that mindware instantiation should be more important for base-rate tasks that present a weaker cue (moderate ratios) than those that present a stronger cue (extreme ratios). This is also evident from the fact that a significant effect of ratio was determined. Base-rate response frequency was higher for extreme ratios than for moderate ones

(Table 3). Other predictors used in the analysis were education, cognitive reflection and numeracy.

Table 5

Hierarchical Regression Analysis with Education, Cognitive Reflection, Numeracy and Statistical Reasoning as Predictors of Frequency of Base-Rate Responses for Incongruent Tasks with Moderate Ratios (N = 113)

	I.	II.	III.
	β	*β*	*β*
Education	.10	.13	.14
CRT		.15	.10
NS		.07	.03
SRT			.20*
R	.10	.21	.28
R^2	.01	.04	.08
Adjusted R^2	.002	.02	.05
F (*df*)	1.20 (1, 111)	1.70 (3, 109)	2.34 (4, 108)
ΔR^2		.03	.04
F_{Δ} (*df*)		1.93 (2, 109)	4.12 (1, 108)*

Note. SRT = Statistical Reasoning Test; CRT = Cognitive Reflection Test; NS = Numeracy Scale. * $p < .05$.

A three-step multiple hierarchical regression analysis was conducted. In the first step, education was introduced. Education did not prove predictive in explaining the variability of frequency of base-rate response choices ($p > .05$). In the second step, cognitive reflection and numeracy were introduced, and there was also no significant increase in the explained variance of the frequency of base-rate response choices. In the third step, statistical reasoning was introduced and it accounted for the increase of explained variance of 4% ($\Delta R^2 = .04$, $p < .05$). Statistical reasoning showed the only significant positive contribution ($\beta = .20$, $p < .05$) in explaining the variance of frequency of base-rate response choices in noncongruent tasks with moderate ratios even though multiple regression coefficient was not statistically significant.

Although one of the initial intents of this research was also to test the predictive contribution of mindware instantiation in explaining the variance of conflict detection efficiency, multiple regression was not conducted because the conflict detection index showed no significant correlation with mindware instantiation (statistical reasoning test).

Discussion

This study aimed to examine the relationship between mindware instantiation and base-rate neglect task efficiency by operationalizing mindware as a degree of the automation of statistical reasoning. Within the dual process approach, there have been only a few studies on mindware which incorporated base-rate neglect tasks as evidence of successful mindware instantiation, and such studies have not included the automation of statistical reasoning as a measure of relevant mindware instantiation. However, we hypothesize that the automation of statistical reasoning is the best operationalization of the degree of mindware instantiation due to the same principles that underlie both statistical reasoning and base-rate task efficiency – they are both based on a knowledge of statistical principles such as ratio and probability.

One of the goals of this study was to determine the difference in base-rate response frequency depending on the task ratio and congruency. There is numerous research the results of which suggest that there should be a higher frequency of base-rate responses when base-rate tasks consist of extreme ratios and ratio/attribute congruence (e.g. Bago & De Neys, 2020; De Neys & Glumicic, 2008; Dujmović & Valerjev, 2018; Pennycook et al., 2015). The findings from this study were similar. Namely, more base-rate responses were made when tasks were attribute/ratio congruent rather than incongruent. These results provide more evidence of base-rate task ability to provoke incorrect (stereotypical) responses when the task consists of conflicting information. Every base-rate neglect task consists of two kinds of information, which can be considered as cues, and these are attribute and ratio. They can either be congruent (e. g. "Person A is fast. Person A was selected at random from a group which consists of 995 runners and 5 singers.") or conflicting (e. g. "Person B is brave. Person B was selected at random from a group which consists of 995 painters and 5 firefighters."). When cues are conflicting, it is more likely that an incorrect response will be triggered. Also, the results of this study suggest that more base-rate responses were made when tasks consisted of more extreme ratios rather than moderate ones, but this only applied to incongruent tasks. There was no difference in base-rate response frequency when the tasks were congruent. A possible explanation of these results is that attribute/ratio congruency was a strong enough cue for logical intuition to reach its maximum potency. As a result, there was no difference in correct response frequency for extreme ratios when attribute and ratio were congruent. On the other hand, when there were conflicting cues (when attribute and ratio are incongruent), the ratio extremity cue was needed in order to provide a correct response. This explanation is in accordance with the Hybrid Model of Dual Processing. When there is a cue which triggers heuristic intuition, logical intuition also needs to be triggered in order to increase the chances of the override of the incorrect response in order to occur (Pennycook et al., 2015). As was the case in some of the examples of prior research, it was established that the variation of ratios in base-rate neglect tasks affected the strength of logical intuition (e. g. Bago & De Neys, 2020; Newman et al., 2017; Pennycook et al., 2015).

The following goal of this study was to determine whether judgments of confidence are different for correct and incorrect responses given in conflict and no-conflict tasks. In accordance with the results of previous research (Bago & De Neys, 2020; Šrol & De Neys 2021), judgments of confidence in this study were highest for correct (base-rate) responses in congruent tasks, as there was no cue which triggered heuristic intuition. Judgments of confidence were lower for incorrect (stereotypical) responses in incongruent tasks and for correct (base-rate) responses in incongruent tasks. Although differences between confidence in different responses were detected, conflict detection index showed no correlation with the degree of mindware instantiation, and predictive contribution of mindware instantiation in explaining the variance of conflict detection efficiency was not tested further. A possible reason why no correlation occurred is low statistical power due to the relatively small sample size. Namely, as mentioned previously, conflict detection index was calculated as the difference between judgment of confidence in correct base-rate response in congruent tasks and incorrect stereotypical response in incongruent tasks. Only participants who got at least one item correct and one item incorrect were considered for further analyses. Also, only participants who had higher judgments of confidence for correct response in no-conflict tasks than for incorrect response in conflict tasks were included in the final sample. The final sample consisted of only 59 participants. Further methodological explanations as to why this correlation was not determined are presented later in the discussion.

The main goal of this study was to examine the relationship between mindware instantiation and base-rate neglect task efficiency by operationalizing mindware as a degree of the automation of statistical reasoning. In both previous studies conducted to measure the degree of mindware instantiation (Frey et al., 2018; Šrol & De Neys, 2021), a neutral success in all tasks (*base-rate neglect task*, *conjunction fallacy, belief bias*, *bat-and-ball problem*) was used as a measure of mindware instantiation. Such a test in the study conducted by Šrol & De Neys (2021) showed low inter-reliability ($\alpha = .28$), but the results indicate that mindware was a significant predictor of success in the test. This study provided similar, but not the same results. Interestingly, in the regression model both in this study and the study conducted by Šrol & De Neys (2021), mindware increased the percentage of explained variance by a similar amount – 4% in this study and 5% in the latter. Specifically, with the control of education, cognitive reflection and numeracy, mindware instantiation (statistical reasoning) proved to be the only significant positive predictor of success in base-rate neglect tasks in this study, in contrast to the study conducted by Šrol and De Neys (2021), where cognitive ability, numeracy, faith in intuition and cognitive reflection were all significant predictors. In other words, a higher score on the Statistical Reasoning Test was found to indicate more normative responses in base-rate tasks, and the findings of our study indicate that it is the only significant predictor. The probable reason that different research findings are obtained is that in previous studies (Frey et al., 2018; Šrol & De Neys, 2021), a composite score was used as a criterion variable on all previously mentioned tasks, and in this study only base-rate response choice in noncongruent tasks was used, because mindware is known to be task-specific (Stanovich, 2018). Moreover, it was determined that there were more base-rate responses in conflict tasks with extreme ratios than moderate ratios and correlation with base-rate response frequency with extreme ratios and with merged extreme and moderate ratios was not significant. We consider this as evidence of test sensitivity, because mindware instantiation should be more important for base-rate tasks that present a weaker cue (moderate ratios) than those that present a stronger cue (extreme ratios). Also, the Statistical Reasoning Test used in this study measures both speed and accuracy, and is time-limited, which indicates that the emphasis was on the automation of knowledge itself rather than the mere possession of the knowledge of the general rules required for solving given tasks.

Since conflict detection efficiency is not confirmed to be correlated with mindware instantiation, we will attempt to draw attention to the possible methodological flaws in this study that may have interfered with the obtained results. First, it is important to emphasize that judgments of confidence were the only measure used for conflict detection. However, there are some other indirect measures of conflict detection, such as response latency (e. g. De Neys et al., 2013; De Neys & Glumicic, 2008; Pennycook et al., 2014; Stupple & Ball, 2008), judgments of confidence latency (Frey et al., 2018), and neural activity measures (De Neys et al., 2008). We assume that response latency might be a better measure of conflict detection than judgments of confidence which were used in this study. Next, the findings of this study may not be concordant with the findings of the study conducted by Šrol and De Neys (2021) because both studies included different measures of mindware instantiation, as stated earlier. Finally, we suggest including more measures of individual differences in the regression model, such as cognitive ability, belief in intuition and thinking disposition, in order to identify the most significant predictors of success in base-rate neglect tasks (e.g. Klaczynski, 2014; Teovanović et al., 2015; Toplak et al., 2011).

Conclusion

The results obtained in this study are mostly consistent with Stanovich's (2018) theoretical assumptions that mindware is a significant predictor of success in cognitive tasks, which suggests the importance of including mindware in the research of individual differences in performance in base-rate neglect tasks. Automation of statistical reasoning has been found to be a suitable measure of base-rate neglect task efficiency. However, judgment of confidence which was used as a measure of conflict detection in this study, did not prove valid, and other conflict detection measures are recommended in further research.

References

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607–617. https://doi.org/10.1016/j.tics.2017.05.004

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109. https://doi.org/10.1016/j.cognition.2016.10.014

Bago, B., & De Neys, W. (2020). Advancing the specification of dual process models of higher cognition: A critical test of the hybrid model view. *Thinking & Reasoning*, *26*(1), 1–30. https://doi.org/10.1080/13546783.2018.1552194

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177%2F1745691611429354

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*(2), 169–187.

De Neys, W. D., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science*, *19*(5), 483–489. https://doi.org/10.1111%2Fj.1467-9280.2008.02113.x

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248–1299. https://doi.org/10.1016/j.cognition.2007.06.002

De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, *28*(5), 503–509. https://doi.org/10.1177%2F0963721419855658

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, *20*(2), 269–273. https://doi.org/10.3758/s13423-013-0384-5

Dujmović, M., & Valerjev, P. (2017). An image is worth a thousand words, but what of numbers? The impact of multi-modal processing on response times and judgments of confidence in base-rate tasks. In O. Tošković, K. Damnjanović, & Lj. Lazarević (Eds.), *XXIII Science Conference Empirical Studies in Psychology* (pp. 30–36). Faculty of Philosophy, University of Belgrade.

Dujmović, M., & Valerjev, P. (2018). The influence of conflict monitoring on meta-reasoning and response times in a base rate task. *Quarterly Journal of Experimental Psychology*, *71*(12), 2548–2561. https://doi.org/10.1177%2F1747021817746924

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry*, *21*(4), 313–326. https://doi.org/10.1080/1047840X.2010.521057

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177%2F1745691612460685

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177%2F0272989X07304449

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frey, D., Johnson, E. D., & De Neys, W. (2018). Individual differences in conflict detection during reasoning. *The Quarterly Journal of Experimental Psychology*, *71*(5), 1188–1208. https://doi.org/10.1080/17470218.2017.1313283

Gangemi, A., Bourgeois-Gironde, S., & Mancini, F. (2015). Feelings of error in reasoning – in search of a phenomenon. *Thinking & Reasoning*, *21*(4), 383–396. https://doi.org/10.1080/13546783.2014.980755

Haigh, M. (2016). Has the standard cognitive reflection test become a victim of its own success? *Advances in Cognitive Psychology*, *12*(3), 145–149. https://dx.doi.org/10.5709%2Facp-0193-5

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Klaczynski, P. A. (2014). Heuristics and biases: Interactions among numeracy, ability, and reflectiveness predict normative responding. *Frontiers in Psychology*, *5*, 665. https://doi.org/10.3389/fpsyg.2014.00665

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44. https://doi.org/10.1177%2F0272989X0102100105

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1154. https://psycnet.apa.org/doi/10.1037/xlm0000372

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pennycook, G. (2018). A perspective on the theoretical foundation of dual process models. In W. De Neys (Ed.), *Current issues in thinking and reasoning. Dual process theory 2.0* (pp. 5–27). Routledge, Taylor & Francis Group.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition*, *42*(1), 1–10. https://doi.org/10.3758/s13421-013-0340-7

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2012). Are we good at detecting conflict during reasoning? *Cognition*, *124*, 101–106.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology, 80*, 34–72. https://doi.org/10.1016/j.cogpsych.2015.05.001

Rapan, K. (2019). *Utjecaj razine statističkog rasuđivanja na uspjeh u zadacima temeljnoga omjera* [Influence of the level of statistical reasoning on success in base rate tasks]. [Unpublished master's thesis]. Department of Psychology, University of Zadar. https://urn.nsk.hr/urn:nbn:hr:162:917171

Rapan, K., & Valerjev, P. (2020). Test statističkog rasuđivanja [Statistical Reasoning Test]. In V. Ćubela Adorić, I. Burić, I. Macuka, M. Nikolić Ivanišević, & A. Slišković (Eds.), *Zbirka psihologijskih skala i upitnika, Svezak 10*. University of Zadar.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127*(11), 966–972. https://doi.org/10.7326/0003-4819-127-11-199712010-00003

Šrol, J., & De Neys, W. (2021). Predicting individual differences in conflict detection and bias susceptibility during reasoning. *Thinking & Reasoning, 27*(1), 38–68. https://doi.org/10.1080/13546783.2019.1708793

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672. https://doi.org/10.1037/0022-3514.94.4.672

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42(4)*, 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177%2F0098628316677643

Stupple, E. J., & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning*, *14*(2), 168–181. https://doi.org/10.1080/13546780701739782

Teovanović, P., Knežević, G., & Stankov, L. (2015). Individual differences in cognitive biases: Evidence against one-factor theory of rationality. *Intelligence*, *50*, 75–86. https://doi.org/10.1016/j.intell.2015.02.008

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In K. Frankish & J. St B. T. Evans (Ed.), *In two minds: Dual processes and beyond* (pp. 171–196). Oxford University Press.

Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. St. B. T. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology: General*, *147*(7), 945–961. https://psycnet.apa.org/doi/10.1037/xge0000457

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275. https://doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Valerjev, P., & Dujmović, M. (2017). Instruction type and believability influence on metareasoning in a base rate task. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 3429–3434). Cognitive Science Society.

Je li automatizacija statističkoga rasuđivanja prikladan *mindware* u zadatku temeljnoga omjera?

Sažetak

Donedavno su istraživanja unutar pristupa dvojnoga procesiranja uglavnom bila usmjerena na grupne razlike u procesiranju, a individualne su razlike bile zanemarene. Međutim, individualne su se razlike pokazale značajnim čimbenikom u objašnjenju učinkovitosti detekcije konflikta i ukupnoga uspjeha u zadatku temeljnoga omjera i sličnim zadacima. Navedeno treba uzeti u obzir u okviru *Hibridnoga modela dvojnoga procesiranja*. Nove tendencije u razvoju toga modela usmjerile su se na stupanj usvojenosti *mindwarea* kao prediktora uspješnosti u zadatku temeljnoga omjera. Slijedom navedenoga, cilj je ovoga istraživanja bio ispitati odnos *mindwarea* i uspjeha u zadatku temeljnoga omjera te ispitati i istražiti odnos između frekvencije odgovora temeljnoga omjera, učinkovitosti detekcije konflikta i stupnja usvojenosti *mindwarea*. Svi su sudionici rješavali zadatke temeljnoga omjera, donosili metakognitivne procjene sigurnosti u svoje odgovore te rješavali Test statističkoga rasuđivanja, Test kognitivne refleksije i Skalu numeričnosti. Test statističkoga rasuđivanja bio je korišten kao mjera stupnja usvojenosti *mindwarea*. Utvrđeno je da je stupanj usvojenosti *mindwarea* jedini značajan prediktor uspješnosti u zadatku temeljnoga omjera, a rezultati su ujedno pokazali i da su sudionici s višim stupnjem usvojenosti *mindwarea* općenito dali više odgovora temeljnoga omjera. Nije utvrđena povezanost između stupnja usvojenosti *mindwarea* i detekcije konflikta. Dobiveni nalazi podupiru hipotezu da moć logičke intuicije ovisi o stupnju usvojenosti *mindwarea*. Stoga rezultati ovoga istraživanja ukazuju na važnost daljnjega istraživanja uloge statističkoga rasuđivanja u objašnjenju uspješnosti u zadacima temeljnoga omjera. Nadalje, u radu se raspravlja i o određenim metodološkim ograničenjima koja bi mogla objasniti zašto nije utvrđena povezanost stupnja usvojenosti *mindwarea* i detekcije konflikta.

Ključne riječi: mindware, statističko rasuđivanje, zadatak temeljnoga omjera, detekcija konflikta, hibridni model dvojnoga procesiranja

Primljeno: 19. 9. 2020.