INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

COGNITIVE BIAS MITIGATION: HOW TO MAKE DECISION-MAKING RATIONAL?

Tomas Kucera

IES Working Paper 1/2020

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Kucera T.(2020): "Cognitive Bias Mitigation: How to Make Decision-Making Rational?" IES Working Papers 1/2020. IES FSV. Charles University.

This paper can be downloaded at: http://ies.fsv.cuni.cz

Cognitive Bias Mitigation: How to Make Decision-Making Rational?

Tomas Kucera[a]

[a]Institute of Economic Studies, Faculty of Social Sciences, Charles University
Opletalova 26, 110 00, Prague, Czech Republic
Email (corresponding author): tomas.kucera009@gmail.com

January 2020

Abstract:
Cognitive biases distort judgement and adversely impact decision-making, which results in economic inefficiencies. Initial attempts to mitigate these biases met with little success. However, recent studies which used computer games and educational videos to train people to avoid biases (Clegg et al., 2014; Morewedge et al., 2015) showed that this form of training reduced selected cognitive biases by 30 %. In this work I report results of an experiment which investigated the debiasing effects of training on confirmation bias. The debiasing training took the form of a short video which contained information about confirmation bias, its impact on judgement, and mitigation strategies. The results show that participants exhibited confirmation bias both in the selection and processing of information, and that debiasing training effectively decreased the level of confirmation bias by 33 % at the 5% significance level.

JEL: D03, D81, Y80
Keywords: Behavioural economics, cognitive bias, confirmation bias, cognitive bias mitigation, confirmation bias mitigation, debiasing

Acknowledgements: This working paper is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 681228.

Introduction

Empirical research has documented a panoply of cognitive biases which impair human judgement and make people depart systematically from models of rational behaviour (Gilovich et al., 2002; Kahneman, 2011; Kahneman & Tversky, 1979; Pohl, 2004).

Besides distorted decision-making and judgement in the areas of medicine, law, and military (Nickerson, 1998), cognitive biases can also lead to economic inefficiencies. Slovic et al. (1977) point out how they distort insurance purchases, Hyman Minsky (1982) partly blames psychological factors for economic cycles. Shefrin (2010) argues that confirmation bias and some other cognitive biases were among the significant factors leading to the global financial crisis which broke out in 2008. Lunn (2013) concludes that confirmation bias[1] specifically had contributed to the severity of the banking crisis in Ireland.

To tackle these issues, debiasing efforts have emerged. Larrick (2008) describes three basic approaches. Firstly, incentivizing people to perform better (Stapel, Martin, & Schwarz, 1998). Secondly, designing decision environment in a way that prevents or offsets predictable biases (Klayman & Brown, 1993; Arnott, 2006; Kahneman, Lovallo, & Sibony, 2011). And thirdly, edifying and training people (Clegg et al., 2014). The effectiveness of the various approaches is, however, a matter of concern.

Increasing incentives can be effective in situations when people use suboptimal strategies while the appropriate ones are available to them or they have the capacity to acquire them in the long term (Camerer et al., 1999). However, one of the main characteristics of the vast amount of cognitive biases is that they are deeply ingrained in our mental processes and therefore the effect of increased incentives is limited (Arkes, 1991).

Introducing decision devices can be very effective in mitigating the effects of biased decisions and judgements as it transforms the decision environment. It can be done either by adopting group decision making where the members of the group first form their stands independently and then consult the final decision, which leads to, at least partial, elimination of individual biases, or by implementing computational technology in order to achieve

[1] Confirmation bias is a tendency to seek and interpret information in a way that favours current beliefs, expectations or hypothesis. It affects what kind of information one selects as well as the way one processes them (Nickerson, 1998).

unbiased and consistent evaluation of decision options (Clemen & Reilly, 2001; Dawes et al., 1989; Larrick, 2008; Soll & Larrick, 2009). However, this approach can be inconvenient in many situations as it can be costly and inflexible.

The last approach focuses on the possibility of training people in corrective strategies in order to make them able to recognize the bias or the situation in which the bias is likely to appear and then to apply a strategy so they avoid or mitigate the bias. Although the early attempts to mitigate cognitive biases turned mostly unsuccessful (Kahneman, 2003b), some recent studies present this approach as a viable option (Clegg et al., 2014; Morewedge et al., 2015). With the use of a computer game, both Clegg et al. and Morewedge et al. managed to decrease the level of fundamental attribution error, bias blind spot, and confirmation bias by approximately 30 %. Beside the computer game, they also used a training in the form of a 30-minute-long video but their results diverge on its effects. While Clegg et al. do not report a significant effect in the confirmation bias mitigation for the video, Morewedge et al. present an effect comparable to the one of the computer game, i.e. a reduction of confirmation bias by 30 %.

This work focuses on the third approach in order to help address the question whether teaching people about their cognitive biases and possible mitigation strategies makes any sense. Specifically, I carried out an experiment which investigates the effectiveness of training in mitigation of confirmation bias.

In contrast to the Morewedge et al.'s experiment, I deployed much shorter training intervention in a form of a video, which lasted only 5 minutes compared to 30 minutes in Morewedge et al.'s study, while maintaining the same structure and comparable content. The results show a significant effect of 30% decrease, which is consistent with Morewedge et al.'s (2015) findings.

Method

Participants

The experiment was conducted with 138 participants. All of them were high school students with the age ranging from 16 to 19. The pool of participants consisted of six groups – each group was a separate class. There were two 4th grade classes, i.e. the final grade (37 participants), three 3rd grade classes (78 participants), and one 1st grade class (23 participants). The experiment was carried out in an IT classroom at the high school. The experiment took place as a replacement for a normal teaching lesson and it was compulsory for all the students in the class. Statistics of participants' age are provided in the table below.

Table 1 – Age of Participants

Number of participants	Mean age	Standard deviation
138	17,93	0,97

Note: All participants were high school students within the age range 16-19.
Source: *Own data*

Experimental Procedure

The experiment was carried out via computers and the interface was adopted from Protivínský (2013). A single blind between-group design was deployed. There were two kinds of groups – a treatment group (TG) and a control group (CG). Participants were randomly assigned to either group at the beginning of the experiment without being told which group they were assigned to. The experiment consisted of 12 rounds in which a level of confirmation bias was measured. In the middle of the experiment, between 6th and 7th round, participants watched a video. The video was different for either group (for detailed description of this intervention see chapter *Training*). The effect of the intervention on the level of confirmation bias was then measured in the final 6 rounds.

Confirmation Bias Measurement Method

An adjusted form of the Wason's selection task (Wason, 1968) was adopted for measuring the level of confirmation bias. Participants were given a task to verify validity of a logical statement in a form of $\mathbf{p} \Rightarrow \mathbf{q}$[2] for a given sample of 4 cards. There were following types of cards in the sample: **p**, ¬ **p**[3], **q**, and ¬ **q**. Participants could turn the cards over to test the statement. The types of the cards that people selected for verifying the statement were the major matter of interest.

[2] „⇒“ stands for implication.
[3] „¬“ means logical negation.

The task was assigned repeatedly in 12 rounds. There were three topical modifications of the task background (framings) and were assigned in a repeated sequence so every participant solved the task in each framing twice in the first 6 rounds and then again twice in the final 6 rounds.

Manifestation of Confirmation Bias

For measurement of confirmation bias, I adopted the approach of Jones and Sugden (2001). It is based on the idea that in the absence of any bias, mistakes (responses not complying with normative model) are distributed randomly. It means that if there was no bias, there would be no systematic pattern in selection of card combinations except the selection of normative choice of **p** and ¬ **q** cards. All other combinations would be then evenly distributed.

On the other hand, more frequent selection of potentially confirming cards **p** and **q** would suggest the presence of confirmation bias. In order to examine this, Jones and Sugden (2001) recommend comparing sets of cards which fulfil the condition that after eliminating cards common to both sets, one set contains only potentially confirming cards while the other then contains only cards which are not potentially confirming. Let A_i denote a set of cards and $f(A_i)$ its isomorphic counterpart. Consider the following sets: A_1 = (**p**), A_2 = (**q**), A_3 = (**p, q**), A_4 = (**p, q, ¬ p**), and A_5 = (**p, q, ¬ q**). It isomorphic opposites are then $f(A_1)$ = (**¬ q**), $f(A_2)$ = (**¬ p**), $f(A_3)$ = (**¬ p, ¬ q**), $f(A_4)$ = (**q, ¬ p, ¬ q**), and $f(A_5)$ = (**p, ¬ p, ¬ q**). Significantly higher selection frequency of A_i sets compared to the selection frequency of $f(A_i)$ can be then interpreted as confirmation bias. The magnitude of the eventual difference in selection frequencies of these respective sets can be referred to as the level of confirmation bias.

Training

As already stated above, participants watched a video after the 6th round. There were two different videos, video A and video B, and the video that a participant watched depended on the group he or she was assigned to at the beginning of the experiment. Participants assigned to the treatment group were presented with the video A, and participants in the control group watched the video B. The video A served as a training tool. The video B, though very similar in its form, does not contain any training aspects and served only as a control instrument for an effect of a video per se. The form of the videos is very simple - a narrator, kindly embodied by professor Potužák,[4] stands in front of a white board and speaks to the camera. Both videos are, as well as all instructions of the experiment, in Czech language. The detailed description of the videos' content follow.[5]

Video A

The video A was deployed as a debiasing tool. It first introduces the concept of cognitive biases by providing an abstract definition and then demonstrating the principle on a practical example as recommended by Fong et al. (1986). The example reveals the presence of confirmation bias. The narrator explains at which point in the example the confirmation bias has appeared and how it has influenced judgement. A theoretical explanation of confirmation bias is also provided. Then follows other examples of confirmation bias in various contexts. Finally, in the last part of the video, mitigation strategy, namely "Consider the opposite" strategy, is described. The video lasts for 5 minutes, which is relatively short compared to 30-minute-video in Morewedge et al.'s experiment (2015).

Video B

In order to create comparable conditions for both treatment and control group, I created the video B (again with the kind assistance of professor Potužák) with neutral content in respect to confirmation bias mitigation. The video B has the same opening as the video A, introducing the concept of cognitive biases. What follows, however, is a list of some cognitive biases, namely hindsight bias, the law of small numbers fallacy, and representativeness heuristic. All these biases are briefly explained and described on attractive experiments. The video B lasts for 5 minutes as well as the video A.

After watching a video, participants continued with solving the final 6 rounds of the adopted version of Wason's selection task.

[4] Ing. Pavel Potužák, Ph.D., Assistant Professor at University of Economics, Prague

[5] Videos are also available at the following links: https://youtu.be/m2Ro9wHTCPs (Video A), https://youtu.be/-1LXKo9SHXI (Video B).

Table 2 - Distribution of participants between treatment and control group

Group	Number of participants	Relative share
Treatment	68	49,3 %
Control	70	50,7 %

Note: Participants were distributed to the groups randomly at the beginning of the experiment.

Source: *Own data*

Incentives

It is common in economic experiments that financial incentives are deployed and that a part of the remuneration is performance dependent (Hertwig & Ortmann, 2001). An awarding scheme was designed to motivate participants to deploy the appropriate strategy.

Every participant started with 120 points. There was a reward of 30 points for every correct answer, but no points were neither awarded nor subtracted for an incorrect judgement. When turning cards over, participants were charged according to the following scheme:

- 1 card turned over = - 3 points,
- 2 cards turned over = - 7 points,
- 3 cards turned over = - 12 points,
- 4 cards turned over[6] = - 18 points.

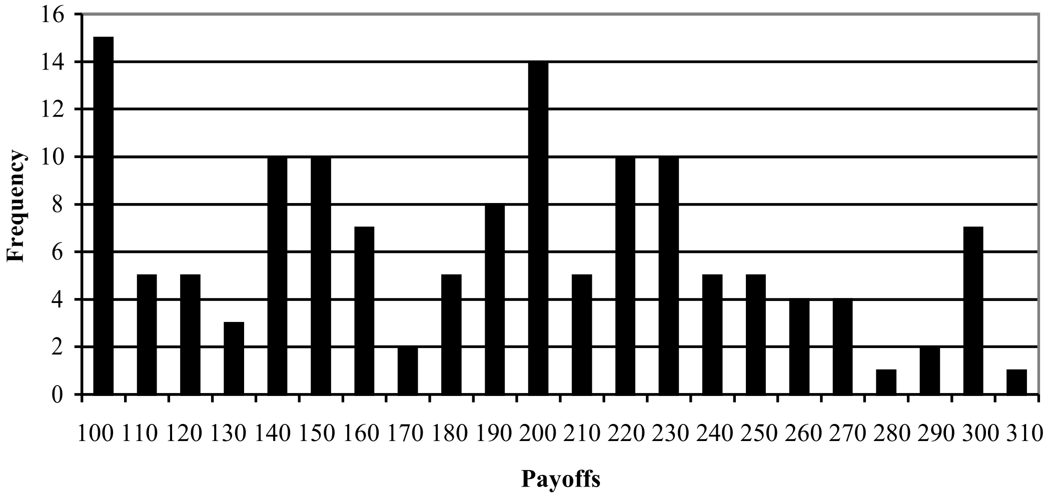
Such a scheme reflects a realistic assumption of increasing marginal costs of information acquisition. The specific values are based on the analysis of the normative model, i.e. given the distribution of probabilities and the payoffs, a rational strategy is to proceed according the normative model. At the end of the experiment, 100 points were subtracted, and the remaining number of points was the value[7] of participant's result. A minimum reward was set at 100 CZK. It meant that if a participant's result after subtracting 100 points got below 100,

[6] Number of cards turned over in a given round.
[7] Value of the result was expressed in CZK.

the participant was granted the minimum reward. A reward corresponding to the normative strategy was 296 CZK. The distribution of payoffs is depicted in the figure below.

However, as the experiment was carried out at the high school with many participants under 18, the financial rewards were eventually translated into material rewards, namely stationery, chocolates, and other sweets. In order to maintain the motivational character of the awarding scheme, participants were allowed to decide what they wanted to exchange their obtained points for. There was an experimenter's shop set up and the exchange ratio between the monetary and material reward was based on the value of the material rewards. Therefore, the risk of distortion of the rewarding scheme by the disproportion of subjective values was decreased.

Figure 1 - The payoff distribution



Note: There was a guaranteed minimum payoff set at 100 CZK. A reward attainable under the optimal strategy was 296 CZK. Higher payoff was possible if participants guessed the answer and got lucky, which was suboptimal in terms of expected value given the probabilities. Although payoffs are expressed in CZK, participants were not given the monetary prize, but chose a material prize of an equivalent value in an experimenter's shop.

Source: *Own data*

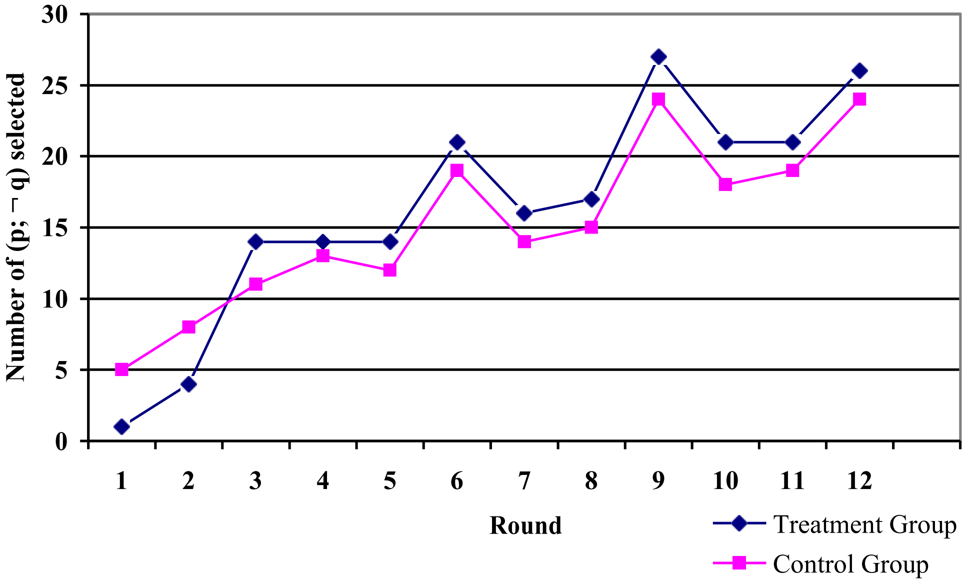
Results

Responses

The responses show that compliance with normative model of decision-making was relatively low, only 24 % (21,7 %) in the treatment (control) group, compared to 42,6 % in the results of Protivínský (2013). Participants in the Protivínský's experiment were, however, university students.

In order to assure that the groups exhibited comparable results, I tested the differences between the frequencies at which the cards were selected in the respective groups during the first 6 rounds (since an intervention came before the 7^{th} round). The results of t-test show that the differences in the selection frequencies of **p**, **q**, **¬ p**, and **¬ q** between the respective groups are not statistically significant at the 5% level of significance (t-test, p = 0,064; 0,302; 0,355; 0,542; in the respective order). There is also no difference between the groups in the frequency at which participants complied with the normative model during the first 6 round. In fact, the mean frequencies are identical. Therefore, the groups can be deemed as comparable.

Figure 2 - Trend of the number of participants complying with normative model



Note: The figure shows how the number of participants complying with the normative model evolved throughout the game. The normative model assumes the optimal strategy that maximizes the expected payoff. Such optimal strategy, given the experiment settings, defines the minimum number of cards that are needed for verifying the statement, namely cards of type *p* and ¬ *q*.

Source: *Own data*

There is also an obvious trend of learning in the compliance with normative model as can be clearly seen on the following graph. Such a trend is also found in other similar studies (Jones & Sugden, 2001; Protivínský, 2013).

Presence of Confirmation Bias in the Selection of Cards

Looking at Table 3, the presence of confirmation bias is clearly visible in each round of the experiment. As argued above, in the absence of confirmation bias, the values of A_i and $f(A_i)$ sets would not be significantly different. However, here we can see that it is not the case. The binomial test confirms that the values are significantly different from the values expected in the absence of confirmation bias ($p < 0{,}001$ in each round for both groups).

Table 3 – Frequency of combinations of cards turned over by treatment group

Group	Round	1	2	3	4	5	6	7	8	9	10	11	12
Treatment	A_i sets	54	53	42	48	41	38	32	31	23	32	31	27
	$f(A_i)$ sets	1	4	6	0	7	3	7	8	8	7	7	6
	level of CB	53	49	36	48	34	35	25	23	15	25	24	21
Control	A_i sets	53	51	41	44	43	35	39	38	30	35	33	30
	$f(A_i)$ sets	8	3	6	8	7	4	8	7	6	6	5	5
	level of CB	45	48	35	36	36	31	31	31	24	29	28	25

Note: The table shows the selection frequencies of card combinations belonging to set A_i or $f(A_i)$, and the level of selection bias for both treatment and control group in each round. The level of confirmation bias is calculated as the difference between the values of A_i and $f(A_i)$ sets. Set A_i consist of potentially confirming cards and set $f(A_i)$ is its isomorphic counterpart. This classification follows Jones and Sugden (2001). For detailed description of these sets, see the chapter Manifestation of Confirmation Bias above.

Source: *Own data*

Table 3 shows the level of confirmation bias for each respective round, calculated as the difference between the values of A_i and $f(A_i)$ sets. The level of confirmation bias is further elaborated below in the chapter on the effects of the training intervention. For detailed numbers on the selection frequencies of all 16 possible card combinations, please see Tables A.1 and A.2 in the Appendix.

The results were also aggregated by the framings. In order to find out whether a given framing had any impact on the level of confirmation bias, I carried out a chi-squared test. The results for both treatment (χ^2-test, p = 0,481) and control (χ^2-test, p = 0,870) show that at the 5% level of significance there was no significant difference between the results under the various framings. The data reported in Table 4 are the average frequencies of respective sets under given framings. The row called *level of CB* reports the average level of confirmation bias under the given framing.

Table 4 - Comparison of confirmation bias across framings

Group	Framing	Restaurants (neutral)	Rivers (causal)	Shops (deontological)
treatment	A_i sets	41,5	39	32,5
	$f(A_i)$ sets	3,75	6,5	5,75
	level of CB	37,75	32,5	26,75
control	A_i sets	42,75	41,25	34
	$f(A_i)$ sets	7,5	5,5	5,25
	level of CB	35,25	35,75	28,75

Note: The table shows average selection frequencies of respective sets of card combinations as defined above, and of the level of confirmation bias under given framings for both treatment and control group. A chi-squared test showed no significant difference between the framings at 5% level of significance.

Source: *Own data*

Presence of Confirmation Bias in Judgement

After recognizing the presence of confirmation bias in the selection of information, the following question was whether participants' judgements were biased as well or whether the irrelevant pieces of information were simply ignored.

Table 5 shows what judgement participants made depending on the information on the cards turned over. The rows in the table represent all possible variations of information which could be obtained by turning the informative cards **[p; #]** and **[¬ q; #]**. Similarly, the columns represent a breakdown by the information content of the uninformative cards **[q; #]** and **[¬ p; #]**, and by the judgement made (True of False).

It is worth noting that any deductive mistakes in the participants' judgements were very rare once they had sufficient evidence for establishing the statements as true or false. From the results in the 4th (the row denoted as *[p; ¬ q] only*) to 8th row it is clear that participants recognized the significance of disconfirmation if they found it in the vast majority of cases. Specifically, they judged the statement as false in 419 out of 456 cases. Likewise, as it can be seen in the 9th row (denoted as *[p; q] + [¬ q; ¬ p]*), once participants had the sufficient evidence needed for judging the statement as true, they did so in 228 out of 238 cases. This also suggests that most participants understood the task.

Table 5 - Relation between judgements and the cards turned over

		Uninformative cards turned over							
		Neither [q; #] nor [¬ p; #]		[q; p]		[q; ¬ p]		[¬ p; #] but not [q; #]	
		True	False	True	False	True	False	True	False
Informative cards turned over	none	44	43	40	8	16	38	20	9
	[p; q] only	190	16	145	2	74	118	17	12
	[¬ q; ¬ p] only	30	12	9	1	7	8	14	5
	[p; ¬ q] only	5	50	7	68	0	0	0	2
	[¬ q; p] only	5	22	0	4	0	5	1	5
	[p; q] + [¬ q; p]	11	151	1	33	1	20	0	2
	[p; ¬ q] + [¬ q; p]	0	0	0	0	0	0	0	0
	[p; ¬ q] + [¬ q; ¬ p]	4	41	2	15	0	0	0	1
	[p; q] + [¬ q; ¬ p]	168	3	14	0	41	6	5	1
	Total	457	338	218	131	139	195	57	37
	percent true	57 %		62 %		42 %		61 %	

Note: The table shows what judgement participants made depending on the information on the cards turned over. The rows in the table represent all possible variations of information which could be obtained by turning the informative cards [p; #] and [¬ q; #]. Similarly, the columns represent a breakdown by the information content of the uninformative cards [q; #] and [¬ p; #], and by the judgement made (True of False). From the results in the 4th (the row denoted as [p; ¬ q] only) to 8th row it is clear that participants recognized the significance of disconfirmation if they found it in the vast majority of cases. Likewise, as it can be seen in the 9th row (denoted as [p; q] + [¬ q; ¬ p]), once participants had the sufficient evidence needed for judging the statement as true. The last row of table 5 shows in what percentage of cases participants judged the statement as true with respect to the information revealed by uninformative cards. There is a clear difference in true judgement frequencies between the participants who found [q; p] (62 %) and those who found [q; ¬ p] (42 %), which shows that an uninformative card impacted participants' judgement.

Source: *Own data*

The impact of the uninformative cards can be assessed by looking at the participants who selected card **q**, and at how their judgment was affected by what was revealed after turning the card over. The last row of table 5 shows in what percentage of cases participants judged the statement as true with respect to the information revealed by uninformative cards. The difference in true judgement frequencies between the participants who found **[q; p]** (62 %) and those who found **[q; ¬ p]** (42 %) is clear at the first glance.

The impact of the information on judgement was assessed and the results are presented in the Table 6. It has been confirmed by a chi-squared test that the impact of the card [**q; #]** was significant at the 1% significance level. Therefore, it can be concluded that confirmation bias affected not only the selection of information, but also the way in which the information were evaluated.

Table 6 - Impact of information on judgement

Information	Number of observations of judging:		X^2 test p-value
	True	False	
[p; q]	686	392	<0,001
[p; ¬ q]	19	182	
[q; p]	236	154	<0,001
[q; ¬ p]	158	219	
[¬ p; q]	13	32	<0,001
[¬ p; ¬ q]	81	52	
[¬ q; p]	20	260	<0,001
[¬q; ¬ p]	306	106	

Note: The table shows the impact of information revealed on judgement. A chi-squared test confirmed that the impact of the card [q; #] was significant at 1% level of significance. Thus, it shows that confirmation bias affected not only card selection, but also judgement.

Source: *Own data*

In contrast with the findings of Jones and Sugden (2001), the impact of **¬ p** card on judgment has been also found significant at 1% level of significance. A possible explanation for this pattern could be that some participants understood the logical relation of the considered statement as an equivalence "*iff p, then q*" (meaning if, and only if). However, the actual logical relation was an implication "*if p, then q*". This explanation can be corroborated by the fact that finding **[q; ¬ p]** led participants to judge the statements as false at significantly higher frequency (see Table 6). On the other hand, in order to test equivalence, participants would need to turn over all the cards, which was not the case.

Effect of the Training Intervention

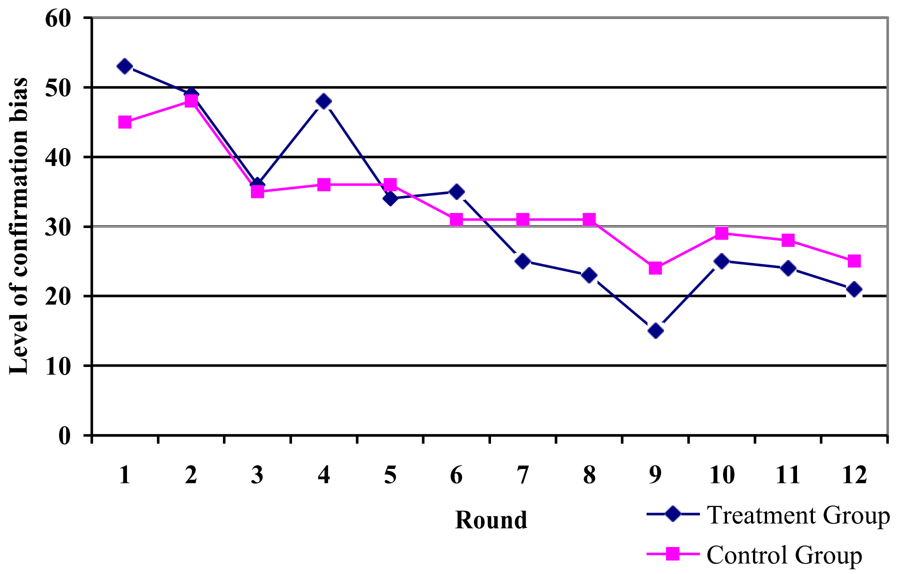
The training intervention was meant to mitigate confirmation bias. It means that its effect should be reflected on the level of confirmation bias. As it can be seen on the following graph, the trend of the level of confirmation bias was decreasing during the course of the experiment for both groups.

In order to distinguish the effect of the training from the effect of learning throughout the experiment, I carried out difference-in-difference analysis. The equation for OLS method reads as:

$$lnCBlevel_i = \alpha + \beta_0 Group_i + \beta_1 After_i + \beta_2 Group_i * After_i + \epsilon_i$$

The dependent variable, ln$CBlevel_i$, is a natural logarithm of a value of the level of confirmation bias of an individual *i*. All of the independent variables are dummy variables. The coefficient β_0 was to show whether there was any difference in the level of confirmation bias between the treatment and the control group as $Group_i$ takes the value of 1 for treatment group, 0 otherwise. As $After_i$ takes the value 1 only for the last 6 rounds, the coefficient β_1 captures the effect of learning the first and the last 6 rounds. And the coefficient of the interaction variable $Group_i*After_i$, β_2 then captures the effect of training intervention.

Figure 3 - Level of confirmation bias



Note: The figure shows the evolution of the level of confirmation bias throughout the game for both treatment and control group.

Source: *Own data*

The results of the analysis show that the training intervention decreased the level of confirmation bias by 33,7 % at the 5% level of significance (t-test, p = 0,025 for the corresponding regression coefficient).

They also show that there was no significant difference in the level of confirmation bias between the groups before the intervention, and that the effect of learning was comparable with the effect of the training intervention, as it leads to the 31 % decrease of the level of confirmation bias at the 1% level of significance.

Table 7 - Effect of the training intervention

Variable	Coefficient
constant	3,639*** (0,070)
Group	0,094 (0,099)
After	−0,312*** (0,099)
GroupAfter	−0,337** (0,139)
R^2	0,731
No. of observations	24

Note: OLS standard errors in parenthesis. The coefficient on the interaction term GroupAfter shows that the intervention effectively decreased the level of confirmation bias by 33,7% at 5% level of significance. Furthermore, it does not show any significant prior difference in the level of confirmation bias between treatment and control group. It also shows that the level of confirmation bias was decreasing throughout the game, suggesting some degree of learning.
** denotes significance at 5% significance level.
*** denotes significance at 1% significance level.

Source: *Own data*

Discussion

Even though the training effectively and significantly decreased the level of confirmation bias, this was not reflected in a higher success rate in verifying the statement. I provide the explanation that since the training video was not designed to provide the optimal strategy for the Wason's selection task but it only presented the concept of confirmation bias and the debiasing strategy in general terms, the participants learned to avoid a biased selection of information but did not managed to infer the optimal strategy from that. The participants simply adjusted their strategies, but not enough to come to the rational strategy of choosing the (p; ¬ q) combination. Instead, the effect of the training was broken down between the slight increase in the frequency of turning over only the card **¬ q**, and the mild decrease in the frequency of turning over only the card **p**. None of these changes in frequency were statistically significant but they happened to be so once added up. As these two strategies are parts of the set $f(A_i)$ and A_i, respectively, and the level of confirmation bias is measured as the difference between A_i and $f(A_i)$ sets, subtracting the increased value of $f(A_i)$ from the decreased value of A_i resulted in the statistically significant decrease in the level of confirmation bias. The question remains whether participants would apply the normative strategy effectively had it been available to them.

Another area worth attention is the ecological validity of the results. The Wason's selection task is quite a specific task which might not well represent the usual setting of decision-making situations in the sense that many of our judgements and decisions can be elicited automatically while during the WTS participants asked to make a judgement with a certain level of deliberation. Therefore, the question is whether the effect of the training intervention is domain-specific or it is transferable to other tasks as well. This might be the subject of further research.

A relevant population for this research is the whole population, across all ages and all social groups. As it can be seen from Table 1, the average age of the participants was only around 18. It can be reasonably assumed that older participants could exhibit better performance due to greater amount of experience in decision making and judgement, which would naturally reduce the training potential.

On the other hand, participants were students of a grammar school which prepares students for further studies at university and their skills might be thus assumed to be above average within their age group. This would imply that there might be other groups for which the potential training effect is even larger. The external validity and the effects on other population groups specifically is subject to further research.

Given the payoff scheme and the guaranteed minimum award, it is conceivable that some participants only skimmed through the test without any effort and, thus, attained an award without much deliberation. Although the payoff scheme was designed to encourage participants to make their best, the peak at the lower end of the payoff distribution (see Figure 1) suggests that some participants could have indeed chosen this strategy. Nevertheless, even if this was the case, the estimate of the training effect would not be affected. It is because these participants' card selection could be reasonably assumed to be random, thus not affecting the level of confirmation bias.

Even though the 5-minute-long training managed to decrease the level of confirmation bias significantly, it has not eliminated it entirely. The question now is whether more comprehensive and perhaps continual training would deliver a complete elimination of the bias.

Conclusion

To address the distorting effects of cognitive biases, various debiasing methods can be deployed. Three main approaches have been outlined, namely increasing incentives, introducing decision devices, and training people in corrective strategies. However, these approaches vary in their effectiveness and each of them might be suitable for a different context.

This work suggests that training people to recognize and avoid cognitive biases can be effective. The results of the experiment clearly confirmed the presence of confirmation bias among participants, both in the selection of information and also in their processing. They also showed that even only a 5-minute-long video containing an explanation of confirmation bias and its impact on our decision, and a recipe for avoiding a biased decision can help reduce confirmation bias by approximately 30 % compared to a proper control group, which is an effect comparable to the one reported by Morewedge et al.

Note that the effect of the training was detected immediately after the training and no postponed measurements have taken place as a part of this experiment. Nevertheless, Morewedge et al. (2015) tested the persistence of the training effects and they still found the effects on a comparable level even 8 weeks after the training. However, given the different duration of the training videos, it could further be investigated how, and if, the duration of the training impacts the persistence of the effects in any way.

The coupling evidence of imperfections in human decision-making and its costs imposes challenges not only on the economic theory but also on the society as a whole. To cope with the flaws in our judgement, it is necessary to have some viable measures available. Thus, the research of the effectiveness of various debiasing methods is a way to secure that resources are not wasted on some ineffective measures.

This work contributes to this endeavour. As this area of research is still in its early stage, we miss the sufficient comparison to assess which debiasing measure is the most effective. However, the results presented here suggest that education might be among the effective tools for improving human decision-making. As we face the serious implications of cognitive biases, spreading the knowledge about biases and equipping people with mitigation strategies bring the prospect of improvement.

References

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486–498.

Arnott, D. (2006). Cognitive biases and decision support systems development: A design science approach. *Information Systems Journal*, *16*(1), 55–78.

Camerer, C. F., Hogarth, R. M., Robertson, V. J., & Baker, K. G. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk & Uncertainty*, *19*(1–3), 7–42.

Clegg, B. A., Martey, R. M., Stromer-Galley, J., Kenski, K., Saulnier, T., Folkestad, J. E., …

Strzalkowski, T. (2014). Game-based Training to Mitigate Three Forms of Cognitive Bias. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2014)*, (14180), 1–12.

Clemen, R. T., & Reilly, T. (2001). Making Hard Decisions with Decision Tools®. *Isbn: 0-534-36597-3*.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*(3), 253–292.

Gilovich, T., Economics, B., Griffin, D., Kahneman, D., Higgins, E., Affairs, P., & Affairs, P. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York, NY: Cambridge University Press.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: a methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*(3), 383-403-451.

Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision*, *50*(1), 59–99.

Kahneman, D. (2003b). A perspective on judgment and choice: Mapping bounded rationality. *Behavioral Science*, *58*(9), 697–720.

Kahneman, D. (2011). *Thinking , Fast and Slow*. *Book*. New York: Farrar, Straus and Giroux.

Kahneman, D., Lovallo, D., & Sibony, O. (2011). Before you make that big decision... *Harvard Business Review*, *89*(6).

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Klayman, J., & Brown, K. (1993). Debias the environment instead of the judge: an alternative approach to reducing error in diagnostic (and other) judgment. *Cognition*, *49*(1), 97–122.

Larrick, R. P. (2008). Debiasing. In *Blackwell Handbook of Judgment and Decision Making* (pp. 316–338).

Lunn, P. (2013). The Role of Decision-Making Biases in Ireland's Banking Crisis. *Irish Political Studies*, *28*(4), 563–590.

Minsky, H. P. (1982). *Can "it" Happen Again?: Essays on Instability and Finance*. *M.E. Sharpe* (Vol. 25).

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing Decisions: Improved Decision Making With a Single Training

Intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129–140.

Nickerson, R. S. (1998). Confirmation Bias : A Ubiquitous Phenomenon in Many Guises, *2*(2), 175–220.

Pohl, R. F. (2004). *A handbook on fallacies and biases in thinking, judgement and memory*. New York, NY: Psychology Press.

Protivínský, T. (2013). *Konfirmační zkreslení a vliv zpětné vazby na jeho překonání*. Brno.

Shefrin, H. (2010). How psychological pitfalls generated the global financial crisis. *Leavey School of Business*, (10), 1–42.

Slovic, P., Fischhoff, B., Lichtenstein, S., Corrigan, B., & Combs, B. (1977). Preference for Insuring against Probable Small Losses : Insurance Implications. *Journal of Risk and Insurance*, *44*(2), 237–258.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: how (and how well) people use others' opinions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(3), 780–805.

Stapel, D. A., Martin, L. L., & Schwarz, N. (1998). The smell of bias: What instigates correction processes in social judgments? *Personality and Social Psychology Bulletin*, *24*(8), 797–806.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273–281.

Appendix

Table A.1 – Frequency of combinations of cards turned over by treatment group

Round	1	2	3	4	5	6	7	8	9	10	11	12
no cards	6	2	2	3	3	1	2	2	4	3	3	4
(p)*	22	14	12	10	8	8	8	8	5	5	7	3
(q)*	7	8	5	2	1	3	3	2	1	3	4	4
(¬ p)#	0	1	2	0	1	1	1	0	2	1	0	0
(¬ q)#	1	2	2	0	4	1	3	6	4	4	5	4
(p; q)*	22	29	19	32	25	18	13	13	9	17	12	12
(p; ¬ p)	0	2	4	0	1	2	0	3	1	0	1	0
(p; ¬ q)	1	4	14	14	14	21	16	17	27	21	21	26
(q; ¬ p)	3	1	0	1	1	0	1	3	1	1	2	2
(q; ¬ q)	3	2	0	1	1	0	3	1	1	1	1	2
(¬ p; ¬ q)#	0	1	2	0	1	0	2	1	0	1	2	1
(p; q; ¬ p)*	1	0	0	1	1	2	1	1	0	0	3	0
(p;q; ¬ q)*	2	2	6	3	6	7	7	7	8	7	5	8
(p; ¬ p; ¬ q)#	0	0	0	0	1	1	1	0	1	1	0	1
(q; ¬ p; ¬ q)#	0	0	0	0	0	0	0	1	1	0	0	0
all cards	0	0	0	1	0	3	7	3	3	3	2	1
total	68	68	68	68	68	68	68	68	68	68	68	68
A_i sets	54	53	42	48	41	38	32	31	23	32	31	27
$f(A_i)$ sets	1	4	6	0	7	3	7	8	8	7	7	6
level of CB	53	49	36	48	34	35	25	23	15	25	24	21

Note: A_i sets are marked by *; $f(A_i)$ sets are marked by #. The upper part of the table shows, the selection frequencies of all 16 possible card combinations in respective rounds for treatment group. The lower part summarizes the selection frequencies by defined categories and reports the level of confirbation bias (calculated as the difference between A_i sets and $f(A_i)$ sets frequencies).

Source: *Own data*

Table A.2 – Frequency of combinations of cards turned over by control group

Round	1	2	3	4	5	6	7	8	9	10	11	12
no cards	2	4	3	3	3	5	3	5	7	7	6	4
(p)*	19	16	11	12	11	11	16	12	9	9	12	13
(q)*	13	6	6	4	3	5	3	6	4	2	2	5
(¬ p)#	3	1	3	4	2	1	3	1	2	0	0	0
(¬ q)#	3	2	3	3	1	3	2	4	2	4	2	4
(p; q)*	21	26	20	22	21	10	15	13	12	15	12	6
(p; ¬ p)	0	2	2	0	0	2	1	2	1	2	3	2
(p; ¬ q)	5	8	11	13	12	19	14	15	24	18	19	24
(q; ¬ p)	0	0	2	0	2	2	0	2	0	0	0	2
(q; ¬ q)	2	2	5	0	2	1	1	0	1	1	3	0
(¬ p; ¬ q)#	2	0	0	1	3	0	1	2	1	1	2	1
(p; q; ¬ p)*	0	0	1	2	0	0	0	0	0	1	0	0
(p;q; ¬ q)*	0	3	3	4	8	9	5	7	5	8	7	6
(p; ¬ p; ¬ q)#	0	0	0	0	1	0	2	0	0	0	0	0
(q; ¬ p; ¬ q)#	0	0	0	0	0	0	0	0	1	1	1	0
all cards	0	0	0	2	1	2	4	1	1	1	1	3
total	70	70	70	70	70	70	70	70	70	70	70	70
A_i sets	53	51	41	44	43	35	39	38	30	35	33	30
$f(A_i)$ sets	8	3	6	8	7	4	8	7	6	6	5	5
level of CB	45	48	35	36	36	31	31	31	24	29	28	25

Note: A_i sets are marked by *; $f(A_i)$ sets are marked by #. The upper part of the table shows, the selection frequencies of all 16 possible card combinations in respective rounds for control group. The lower part summarizes the selection frequencies by defined categories and reports the level of confirbation bias (calculated as the difference between A_i sets and $f(A_i)$ sets frequencies).

Source: *Own data*

IES Working Paper Series

2020

1. Tomas Kucera: *Cognitive Bias Mitigation: How to Make Decision-Making Rational?*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz