COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS, AND INFORMATICS

WHO WILL SAVE INNOCENT KITTIES WITH COUNTERFACTUAL PRIMING?

Master Thesis

2018 **Miroslava Galasová, MBA**

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS, AND INFORMATICS

WHO WILL SAVE INNOCENT KITTIES WITH COUNTERFACTUAL PRIMING?

DEBIASING WITH SELF-BASED VERSUS OTHER-BASED COUNTERFACTUAL PRIMING

Master Thesis

Study Program: Cognitive Science

Field of Study: Cognitive Science

Department: FMFI.KAI - Department of Applied Informatics

Supervisor: Mgr. Lenka Kostovičová, PhD.

Bratislava 2018

Miroslava Galasová, MBA

36672722

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	M.B.A. Miroslava Galasová
Study programme:	Cognitive Science (Single degree study, master II. deg., full time form)
Field of Study:	Cognitive Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Who Will Save Innocent Kitties with Counterfactual Priming?

Annotation: Decision making processes are often influenced by cognitive biases that create barriers to optimal choices. Counterfactual priming is a debiasing method which effectiveness has been proven in several types of biases. We will approach the topic by using ecologically valid scenarios that should motivate participants to authentic reasoning.

Aim: 1. Design and run an experiment with a purpose to test effectiveness of counterfactual priming in three cognitive biases – confirmation bias, attribution error, and sunk cost fallacy.
2. Find out whether "self"-based or "other"-based counterfactual priming impact our decision making differently through ecologically valid tasks.

Literature: De Brigard, F., Spreng, R., Mitchell, J., & Schacter, D. (2015). Neural activity associated with self, other, and object-based counterfactual thinking. Neuroimage, 109, 12-26.
Dudeková, K., Kostovičová, L., & Konečný, M. (2017). Redukcia kognitívnych odchýlok u finančných profesionálov prostredníctvom kontrafaktov. In: Kognícia a umelý život XVII, (pp. 46–51). Bratislava.
Kray, L., & Galinsky, A. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. Organizational Behavior And Human Decision Processes, 91(1), 69-81.

Supervisor:	Mgr. Lenka Kostovičová, PhD.
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	prof. Ing. Igor Farkaš, Dr.
Assigned:	15.10.2016
Approved:	15.11.2016

prof. Ing. Igor Farkaš, Dr.
Guarantor of Study Programme

.... Student

.... Supervisor

36672722

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	M.B.A. Miroslava Galasová
Študijný program:	kognitívna veda (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor:	kognitívna veda
Typ záverečnej práce:	diplomová
Jazyk záverečnej práce:	anglický
Sekundárny jazyk:	slovenský

Názov: Who Will Save Innocent Kitties with Counterfactual Priming?
Kto zachráni nevinné mačiatka kontrafaktovým primingom?

Anotácia: Rozhodovacie procesy su často ovplyvnené kognitívnymi odchýlkami, ktoré vytvárajú bariéry pri výbere optimálnych možností. Kontrafaktový priming je metóda na redukciu kognitívnych odchýlok, ktorej účinok bol preukázaný pri niekoľkých odchýlkach. Pri výskume použijeme ekologicky validne scenáre, ktoré by mali participantov motivovať k autentickému usudzovaniu.

Cieľ: 1. Nadizajnujte a zrealizujte experiment s cieľom otestovať efektívnosť kontrafaktového primingu pri troch kognitívnych odchýlkach – sklon k seba-potvrdzovaniu, atribučný omyl a efekt utopených nákladov.
2. Cez ekologicky valídne úlohy zistite, či kontrafaktový priming zameraný na "ja" alebo "ostatných" spôsobuje rozdiely v rozhodovaní.

Literatúra:
De Brigard, F., Spreng, R., Mitchell, J., & Schacter, D. (2015). Neural activity associated with self, other, and object-based counterfactual thinking. Neuroimage, 109, 12-26.
Dudeková, K., Kostovičová, L., & Konečný, M. (2017). Redukcia kognitívnych odchýlok u finančných profesionálov prostredníctvom kontrafaktov. In: Kognícia a umelý život XVII, (pp. 46–51). Bratislava.
Kray, L., & Galinsky, A. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. Organizational Behavior And Human Decision Processes, 91(1), 69-81.

Vedúci:	Mgr. Lenka Kostovičová, PhD.
Katedra:	FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry:	prof. Ing. Igor Farkaš, Dr.
Dátum zadania:	15.10.2016

Dátum schválenia: 15.11.2016

prof. Ing. Igor Farkaš, Dr.
garant študijného programu

.... študent

.... vedúci práce

Declaration in lieu of oath

This is to confirm my master thesis was independently composed/authored by myself, using solely the referred sources and support.

In Bratislava, 15.1.2018

....

Miroslava Galasová, MBA

Acknowledgments

I am endlessly grateful to Lenka Kostovičová for 179 exchanged emails, everlasting support, and optimism. Many thanks to my family and my landlord for their patience and love. Finally, thank you, Martin P., for extensive sharing of my survey :)

Abstract

How can you save an innocent kitty? Easy – do not make dumb decisions. However, how can we avoid dumb decisions when cognitive biases lurk to prevail our rationality? In this case, counterfactual priming can be the right choice.

Cognitive biases create barriers to optimal choices and can lead to serious consequences. Therefore, we should try to reduce them. There are several debiasing strategies available. We decided for counterfactual priming that seems to be the promising one. In previous studies, counterfactual priming was created by scenarios in which the main character was an unfamiliar person. However, neuroscientific evidence suggests that when we imagine "if only" situations in which we are the main actors (e. g. "If only I was five minutes earlier."), different brain regions are activated than when we imagine "if only" scenarios with someone else. Thus, we assumed that counterfactual priming with self-based scenario will have different effect on reducing cognitive biases than other-based scenario.

We conducted a between subject experiment (N = 266) in which participants solved a task either on attribution error, confirmation bias, or sunk cost fallacy. Then, they were primed by self-based or other-based counterfactual scenarios, and after the priming, they solved one of the two remaining tasks. Moreover, we were interested whether actively open-minded thinking (AOT) and age category (under and above 25) affect susceptibility to the three cognitive biases. We found out that counterfactual priming reduced attribution error; however, only priming by "self" led to a significant difference compared to the control group. Priming by "others", on the other hand, slightly reduced confirmation bias. Sunk cost fallacy was not reduced by counterfactual priming at all. Nevertheless, we observed a moderation effect of AOT and age category on the intervention. Counterfactual priming in general and counterfactual priming by "self" reduced attribution error among people who reached medium and high scores in dogmatism. In contrast, priming by "others" affected only younger participants. Among older participants this type of counterfactual priming led to higher susceptibility to attribution error and sunk cost fallacy than priming by "self".

We conclude that the effect of counterfactual priming on certain cognitive biases depends on the character in priming scenarios, cognitive capabilities (such as AOT), and age of participants. These findings can be used in proposing further practical interventions to optimize individual and group decisions … and for saving innocent kitties.

Abstrakt

Ako dokážete zachrániť nevinné mačiatko? Jednoduché – nerobte hlúpe rozhodnutia. Ale ako sa môžeme vyhnúť hlúpym rozhodnutiam, keď na nás číhajú kognitívne odchýlky, ktoré sa chystajú prevalcovať našu racionalitu? V tomto prípade môže byť dobrým riešením kontrafaktový priming.

Kognitívne odchýlky tvoria bariéry optimálnym voľbám and môžu mať vážne následky. Preto by sme sa mali snažiť tieto odchýlky redukovať. Existuje niekoľko spôsobov na redukciu kognitívnych odchýlok. My sme sa rozhodli pre kontrafaktový priming. V predchádzajúcich štúdiách bol kontrafaktový priming tvorený scenármi, v ktorých hlavnou postavou bola neznáma osoba. Avšak neurovedecké zistenia naznačujú, že sa nám aktivujú v mozgu iné oblasti, keď si prestavujeme „keby som len..." situácie, v ktorých sme v hlavnej úlohe my (ako napr. „Keby som len šla o päť minúť skôr..."), než keď si predstavujeme v tejto situácii niekoho iného. Preto sme predpokladali, že kontrafaktový priming zameraný na „ja" bude mať rozdielny vplyv na redukciu kognitívnych odchýlok než kontrafaktový priming zameraný na „ostatných".

Vytvorili sme medzi-subjektový experiment (N = 266), v ktorom participanti riešili úlohu zameranú na atribučnú chybu, sklon k seba-potvrdzovaniu alebo sklon k utopeným nákladom. Následne boli participanti primovaní kontrafaktovým scenárom zameraným na „ja" alebo „ostatní", a po primingu mali opäť vyriešiť jednu zo zvyšných úloh. Ďalej nás zaujímalo, či má aj aktívne otvorené myslenie (AOM) a veková kategória (do a nad 25 rokov) vplyv na náchylnosť k spomínaným trom kognitívnym odchýlkam. Zistili sme, že kotrafaktový priming redukoval sklon k atribučnej chybe, ale iba primovanie zamerané na „ja" viedlo k významnému rozdielu v porovnaní s kontrolnou skupinou. Na druhej strane primovanie zamerané na „ostatných" mierne znížilo sklon k seba-potvrdzovaniu. Sklon k utopeným nákladom však kontrafaktový priming neznížil vôbec. Napriek tomu sme pozororovali, že AOM a veková kategória mali moderujúci efekt na priming. Efekt kontrafaktového primingu všeobecne a priming „ja" na atribučnú chybu fungoval len u ľudí so strednou a vysokou úrovňou dogmatizmu. Navyše priming zameraný na „ostatných" fungoval pri atribučnej chybe iba u mladých ľudí. U starších viedol k väčšiemu sklonu k utopeným nákladom a seba-potvrdzovaniu než priming zameraný na „ja".

Konštatujeme, že efekt kotrafaktového primingu na niektoré kognitívne odchýlky závisí od protagonistu kontrafaktového primingu, kognitívnych dispozícií (ako AOM) a veku participantov. Tieto zistenia môžu byť ďalej využité pri návrhoch praktických intervencií na optimalizáciu individuálnych a skupinových rozhodnutí a na záchranu malých mačiatok.

Table of Contents

Introduction

Every time we make a wrong (or simply dumb) decision, an innocent kitty dies. Saving innocent kitties is an English idiom which tries to, at least partly, prevent us from thoughtless choosing. Therefore, we thoughtfully chose this idiom because a goal of this study is to find out whether we can somehow push people to reason more objectively and considerately so that they can save these fictional innocent kitties.

Many studies that are mentioned below have discovered that decision-making processes often subject to heuristics and cognitive biases that can mislead us from rational or optimal decisions. However, during last decades, decision-making processes have become intensively observed and studied. Among others, to improve our decision making, scientists have started to propose various debiasing strategies. One of these promising debiasing strategies is counterfactual thinking that uses mental simulations to generate "if only" alternative scenarios. We exploit this strategy in our study to see whether it has a reducing effect on three cognitive biases - attribution error, confirmation bias, and sunk cost fallacy. Moreover, based on neuroscientific evidence we assume that a main character in the counterfactual priming can influence this effect. Therefore, we chose two counterfactual scenarios in which we used a scenario focused on "self" and "others". That means that in the first scenario, participants should imagine themselves when reading the scenario and generating "if only" alternatives, while in the other-based scenario, they imagined a fictional character – Thomas. We have also decided to include Actively Open-minded Thinking (AOT) in testing which seems to be related to certain resistance to cognitive biases.

This master thesis is divided into five sections – basic theories, methods, results, discussion, and conclusion. In the section A little bit of theory, you can find evolvement of decision-making from first normative theories and their classifications to establishment of neuroeconomics. We also describe basic neural correlates of decision-making and certain medical technologies thanks to which we can observe functioning of the brain and even modulate its activity. Further, you find a more detailed description of heuristics, cognitive biases, mainly attribution error, confirmation bias, and sunk cost fallacy. We also describe effects of counterfactual priming and actively open-minded thinking. Finally, this section is concluded by our goals, hypotheses, and research questions. The following section methods summarizes materials and further procedure of our work during the experiment. Eventually, sections results, discussion, and conclusion finalize and debate our findings.

1. A little bit of theory on decision making

Decision making, despite the fact it creates a significant part of our lives, became studied just a few decades ago. We can mainly thank to economists and psychologists, who first discovered the pleasure to ponder or observe decision making processes and came with first theories explaining behavior or consumers. Decision making, since it is not an easy process, consists of two elements – judgement and decision making – in short JDM. When we judge, we consider, compare, and evaluate options. After this part is finished, we can proceed to a second part that is choosing an option or deciding (Bačová, 2010; Gonzalez, 2014). Judgement and decision making are complements that should not be torn apart. This event would have tragic (sometimes even comic) consequences. Just imagine thinking, comparing, weighing in vicious circle without a possibility to cut it off with a decision. Moreover, there are people whose ability to judge is severely damaged and thus, their decisions are often wrong. The cause of these deficits majorly lies in the brain that is a crucial for decision making. Mainly frontal lobes that settled just behind our forehead allow us to judge and choose somewhat wisely and informingly. Here, again, psychologists and economists come to join neurobiologists not only to give birth to neuroeconomics but to discover mechanisms responsible for judgement and decision making.

The history of decision making theories most probably starts in ancient Greece. Aristotle the Peripatetic, who got nicknamed for wandering about the Lyceum while teaching, was interested in human deduction and logical reasoning (Skořepa, 2011). For the next more than a thousand years, pondering about our decisions stayed hibernated. Potentially the interest about this topic reappeared in the beginning of the 17th century when Hamlet of a Shakespeare's play said those notorious words "to be, or not be" (Shakespeare & Timmins, 1860). One way or another, in the following decade mathematicians such as Girolamo Cardano, Luca Paciolli, Blaise Pascal and Pierre de Fermat discovered pleasure in dealing with probabilities that play a role in decision making. Probably the most famous first economist Adam Smith also contributed to the field of decision making when he wrote *The Theory of Moral Sentiment* (1759) and *The Wealth of Nations* (1776). In these works, he focused on choice behavior of consumers and producer and also incorporated psychological factors that affected it (Bavoľár, 2012; Camerer, Fehr, Glimcher, & Poldrack, 2009). Finally, it was Daniel Bernoulli, inspired by his cousin Nicolas Bernoulli, who offered the first theory explaining our decision making.

In 1740's, Daniel Bernoulli presented a paradoxical behavior of people who played a lottery. This paradox, later named St. Petersburg paradox, showed that people were willing to pay only a small amount of money to win an infinite payoff (Buchanan & O'Connell, 2006; Skořepa, 2011). At that time, it was believed that people wanted to maximize their gain, and therefore, they would have been ready to pay any price for the win. However, this assumption was not the right one, and thus, finding an explanation was desired. Daniel Bernoulli suggested that people are more interested in the expected utility of the gain than in the height of the gain. As a result of this observation, he established the term "marginal utility". According to Bernoulli, marginal utility reflected the "moral value of the money" (Baláž, 2010). Marginal utility means that if were hungry and someone buys us a sandwich, we would mark the utility of this sandwich very high because it would ease our hunger. However, if another person gave us another sandwich while we already had one in our stomach, we would not praise the second sandwich as much as the first one. Consequently, our happiness with sandwiches (and their marginal utility) would decrease with every other sandwich we get that moment because we cannot eat it all. A declining trend of marginal utility was also confirmed by Bernoulli's successors.

Despite certain critique of the St. Petersburg's paradox, the Bernoulli explanation became a background for the **theory of expected utility** that was proposed by John von Neumann and Oskar Morgenstern in 1940's and its update, the **subjective expected utility** theory, presented by Leonard J. Savage (Baláž, 2009; Camerer et al., 2009; Skořepa, 2011; Von Neumann and Morgenstern, 1944). Both these theories are normative and expect that we aim to maximize our subjective utility when deciding. We should reach this maximum by a specific combination of our preferences and probabilities that are strongly influenced by our beliefs (Bačová, 2010). The theories of expected utility and subjective expected utility contributed significantly to the field of decision making and helped to explain risk aversive behavior (Baláž, 2010). Nevertheless, first successes of the theories have faded away by often observed deviations from rational decisions.

Normative theories were confronted by several paradoxes. One of them was demonstrated by Maurice Allais at a conference in Paris in 1952. Maurice Allais spread a paper with two questions among his colleagues (Allais, 1953). In the first task, respondents could choose from two options. In the option A, they could win:

- 1 million dollars with 100% certainty

In the option B, they could win:

- 1 million dollars with 89% chance or

- nothing with 1% chance or
- 5 million dollars with 10% chance

In general, respondents chose the option A – a sure gain of 1 million dollars over the option B. The second task also had two options. The option C offered:

- 89% chance to win nothing or
- 11% chance to win 1 million dollars

In the option D, they could win:

- nothing with 90% chance or
- 5 million dollars with 10% chance (Allais, p. 527)

In this task, respondents usually chose the option D that is to win nothing with 90% chance or 10 million dollars with 10% chance. However, choosing the combination of option A in the first task and option D in the second task is a violation of the expected utility theory. Specifically, it violates the independence axiom that is one of four principles defined for this theory (Allais, 1953). The expected utility theory assumes that we neglect the same chances in both tasks. Thus, we should neglect the 89% chance of winning 1 million dollars in both options (A and B) in the first task and 89% chance of winning nothing in both options (C and D) in the second task. If we do that, options A and C and options B and D become the same, and therefore, it is irrational to choose A in the first task while choosing D in the second. Nonetheless, how many of us would naturally come to this logical conclusion? The Allais paradox can be really confusing. Interestingly, Leonard Savage, the author of the subjective utility theory, was also addressed by Maurice Allais in Paris to answer the two tasks (Camerer et al., 2009; Heukelom, 2007), and therefore, he had to admit that expected utility theory has shortcomings that cannot be easily overcome.

The period of 1950's was one of the crucial milestones in the field of decision making. A series of paradoxes, including the one of Maurice Allais, indicated that people always do not go for rational choices, and therefore, they do not behave as 'homo economicus'. Herbert Simon, "a major contributor to cognitive science", also identified with this idea (Camerer et al., p. 3, 2009). He, too, argued against the theory of subjective expected utility; however, his objection differed in its nature from mentioned paradoxes. Simon (1955) argued that organisms are physiologically and psychologically constrained. Next, organisms always cannot get to perfect information as economists often assume in their theories. Moreover, information gathering and dealing with complex problems may cost us a lot of time. Therefore, we use approximations (or heuristics) to simplify demanding decision-making processes. Consequently, due to limited resources (time and cognition) we do not have to go for the optimal option; however, we can go for an option that first satisfies our needs. Of course, choosing a satisfying option over the optimal is not rational, and thus, this conclusion is not in accordance with the expected utility theory. Instead, Herbert Simon introduced the term "approximate" or "limited" (p. 113) rationality that points out at the shortcomings of our 'cognitive apparatus'.

Under the influence of growing objections that indicated inconsistency in rational decision making against the utility theories, scientists started to look for further ways in research of decision making. According to Heukelom (2007), researchers divided into two branches. The first branch continued to study decision making processes from the economic point of view (e. g. Milton Friedman), and the second branch focused on psychological aspects (e. g. Ward Edwards). On the other hand, Goldstein and Hogarth (1997) differentiate four initiatives that were formed in 1950's. The first tried to expand the field or research in decision making. The second focused on an option to find more principles that would explain deviations in human decision making. They also started to pay attention to judgement processes that had been neglected. The third initiative was to broaden existing methods in research. The final one focused on updating of old models together with bringing new models into the field of decision making (Bačová, 2010). Eventually, criticism of expected utility theory led to very fruitful period in this area of research. Old paradigms were shifting and strictly logical conclusions were changing under the influence of true, sometimes irrational, human nature.

In 1970's, psychologists Daniel Kahneman and Amos Tversky (1979) introduced a new theory - **prospect theory**. Unlike normative theories that prescribed how a rational man should behave, prospect theory was based on careful observation of human decision-making processes during a series of experiments. The tasks in these experiments were also different from those used in normative theories because they were adjusted to real life experience (Heukelom, 2007). Moreover, Kahneman and Tversky returned to the idea of marginal utility defined by Bernoulli and renamed it to 'experienced utility'. For a better picture, Morgenstern and von Neumann also operated with a term of utility, later called 'decision utility' in expected utility theory; however, their utility represented only financial means. Kahneman, on the other hand, argued that considering only financial gain or loss in decision making is insufficient. He advocated that individual preferences are those which significantly influence decision making processes (Heukelom, 2007).

Subjective preferences cause that people have different reference points when looking at the same thing. This view can be extended to how people think of future gains and losses (***Figure 1***). Therefore, Kahneman and Tversky (1979) drew attention to relativity of assessing a value. For example, a struggling family can consider a hundred-euro bill as a critical cut over their budget. For a wealthy single man, this bill can be easily neglected. The reason for various reactions to the same situation depends on different reference points that actors have. If another person was added, a new reference point would appear. Thus, our reference point is the one that creates a margin between gains and losses. In a hypothetical value (utility) function presented by Kahneman and Tversky (1979), a function for gain is concave but a function for losses is convex. This difference represents our decreasing sensitivity towards higher gains or losses. Another feature of this function is steeper slope for losses and less steep slope for gains. This means that in general people are more sensitive to losses than to gains because losses are usually painful. Kahneman and Tversky (1979) supported their hypothesis by referencing to several studies that came with results that fit the proposed value function (e. g. Barnes and Reinmuth, 1976; Halter and Dean, 1971).

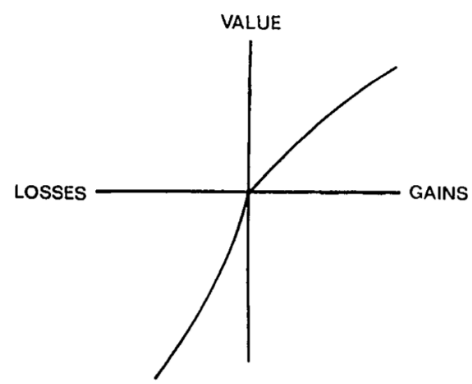


Figure 1. A value function proposed by Kahneman and Tversky (1979) in the article *Prospect Theory: An Analysis of Decision under Risk*.

Having in mind losing money, Kahneman and Tversky (1979) also observed an interesting behavior among respondents playing a simulated lottery. Results showed that people usually behave in a risk aversive way and try to minimize the risk of losing. However, when a loss is inevitable, respondents of the experiment rather chose an option with higher financial loss with 25% probability than an option with lower loss but 50% probability. Kahneman and Tversky called this phenomenon “reflection effect” due to a change in behavior from risk aversive to risk seeking. Kahneman and Tversky also observed that people are not able to realistically assess probabilities that are too high or too low.

Consequently, we tend to either overlook or overweight events with very low probability. On the other hand, events with very high probability we often take as a sure thing or we overstate the difference between sureness and real probability. Based on this observation, Kahneman and Tversky created a convex weighting function because our subjective decision weights did not correspond with the real probabilities.

Prospect theory formulated by Kahneman and Tversky in 1979 was a breakthrough in the field of decision making. Inspired by Herbert Simon (1955) and bounded rationality, prospect theory zoomed on irrational decisions as limits of our cognitive abilities and clarified paradoxical behavior of people dealing with gains, losses, and risks. Moreover, it also showed that we are not able to realistically and rationally assess probabilities that lie in extremes of very high or very low probabilities. By explaining anomalies in human decision making, this theory became a successor of expected utility theory that based on its normative character did not reflect the true decisions that people made. However, despite many positives, prospect theory was also criticized because it violated stochastic dominance and the number of outcomes was limited to a very small number. As a result, Tversky and Kahneman (1992) came with an upgrade of prospect theory called **cumulative prospect theory**.

The most important change was that the new theory focused on cumulative probabilities rather than on individual probabilities. Another feature of this upgrade led to changes in weighting function which shape changed from convex to inverted S-shape. This displayed overweighting and underweighting of events with very low and very high probability respectively. Finally, both prospect theories became respected and valued for their explanation power of certain decision-making processes. They also provided a nutritive substrate for further theoretical and practical extensions. Moreover, growing interest in neuroscience and new imaging techniques in medicine provided attractive possibilities for observation of decision making processes *in situ*, in a brain, and gave rise of the new scientific field – neuroeconomics.

1.1 Neuroeconomics and neuroscience behind decision-making

Neuroeconomics is a young interdisciplinary scientific field that combines knowledge of psychology, neuroscience, and economics. However, its origins are, of course, rooted deeper in the history. In 19th century psychologists and neurologists became interested in connections between brain and behavior (Camerer et al., 2009). A famous case of Phineas Gage, a man whose scull and frontal lobes were pierced by a steel bar, supported the idea that behavior is a product of a brain. Family of Phineas Gage observed that he displayed a significant change in behavior after this injury. Moreover, in further examination, specialists also found out that his decision-making ability was also severely impaired (Macmillan, 2002). In 1970's, Kahneman and Tversky (1979) worked with healthy participants; however, inspired by Herbert Simon and his theory of limited cognitive ability, they focused on suboptimal choices in decision making influenced by risk and loss. In the end 20^{th} century, Antonio Damasio, Daniel Tranel, Hanna Damasio (1990) and Antoine Bechara (1997) also observed decision making under risk; however, their participants were also patients who suffered a brain injury or other brain damage in the area of frontal lobes. Based on series of experiments these researchers found out that these patients were not able to make optimal choices due to the brain damage. Moreover, their results supported three hypotheses. First, frontal lobes play a crucial role in decision making. Second, emotions have a significant impact on decision making. Third, certain physiological changes in brain can be observed in behavior.

The work of Damasio and Bechara was another step in better understanding of decision making processes and thanks to new technologies in medicine, researchers could go even further. Functional neuroimaging tools, such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), finally allowed to see not only what how our brain looks like but what is happening in it (Camerer et al., 2009). Thanks to fMRI and PET, scientists can observe areas of the brain that related to various processes like decision making. In addition, fMRI technique compared to positron emission tomography (PET) is not invasive in the meaning that no radioactive substance needs to be administered to a patient. Thus, fMRI provides a (relatively) safe option for observation of brain functioning.

As a result, new technologies helped cognitive psychologists and behavioral economists observe decision making processes. Neuroscientists and physiologists who were interested in decision making and economics could use economic theories as a manual that helped them to determine neural circuits for decision making. They also believed that studying neural mechanisms of decision making could explain anomalies in decision making in neoclassical economic theories. Eventually, by identifying the reasons for these anomalies, they thought, it would be possible to remove them (Camerer et al., 2009). Moreover, neuroscientific approach can also support previous conclusions that were articulated after behavioral experiments. For example, this happened when a Breiter and colleagues (2001) conducted an experiment inspired by Kahneman and Tversky. Participants monitored by fMRI played a lottery game in which they had three options. One of the options was a zero-dollar ‘gain’. Experimenters were changing values of the other two, non-zero wins to observe how brains of participants reacted to the zero-dollar gain. According to Kahneman and Tversky (1979), zero-dollar gain should be perceived positively, when the other two options are worse than the zero-dollar win. On the other hand, when the two options are above zero, then, the zero-dollar gain is perceived negatively. Eventually, fMRI results confirmed conclusions of prospect theory because an activation of ventral striatum, an important brain structure in reward processing and decision making, was in line with expected negative and positive reaction to zero-dollar gain.

By observation of blood flow in the brain via fMRI we can identify which regions of the brain turn on when we are fulfilling a task. Therefore, we can recognize which brain structures correlates with decision-making processes. What we cannot observe from functional brain imaging is what exactly a turned-on structure causes while we are e. g. deciding. In other words, we see which part is active but we do not know what its role in judgement and decision-making processes. Here, new technologies in neuroscience have offered tools that seem to help in this issue. One of them is transcranial magnetic stimulation (TMS) which is a non-invasive method by which we can stimulate specific regions in the brain through electric impulses. Though, if we know from fMRI which areas relate to certain task processing, we can observe how people behave when these specific areas are temporarily disrupted by TMS (Camerer et al., 2009). Thanks to this approach, researchers Knoch, Pascual-Leone, Meyer, Treyer, and Fehr (2006) could find out that although fMRI from an earlier study conducted by Sanfey, Rilling, Aronson, Nystrom, and Cohen (2003) displayed activity in both right and left dorsolateral prefrontal cortex when people decided about unfair bids, only the right dorsolateral prefrontal cortex stimulated by TMS caused that participants were willing to accept more unfair bids (***Figure 2***). In addition, the fMRI study of Sanfey and his colleagues (2003) again proved importance of emotions in decision making when anterior insula, a structure connected with emotions, got activated when unfair bids were proposed and rejected respectively.

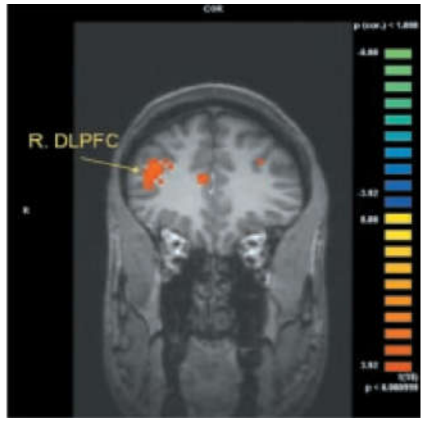


Figure 2: Right dorsolateral prefrontal cortex (R. DLPFC) activated by fair and unfair bids (Sanfey et al, 2003). Knoch and colleagues (2006) used TMS to find out that impaired right DLPFC causes that people are willing to accept unfair bids more often.

Previous example with TMS showed that when we change neural activity in certain brain areas, we can achieve a change in behavior, as well. TMS, however, is not the only mean that can do that. There are also other variables that can modulate our behavior. For example, oxytocin and dopamine can be such modulators. Oxytocin is a (neuropeptide and a peptide) hormone naturally produced in our brains and except other functions, it is responsible for social bonding. Michael Kosfeld and his colleagues from the University of Zurich assumed that oxytocin may be a 'precursor' for trust among people (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005). Moreover, while this applies not only to personal relationships but political, economic, and business relations, they have hypothesized that oxytocin can influence decision making by increasing trust among concerned parties. Therefore, they conducted an experiment in which participants in the experiment group inhaled oxytocin and then they played a trust game. Results of the experiments showed that participants who inhaled oxytocin (investors) expressed higher trust towards trustees despite their knowledge that later, these trustees could cheat on them. However, Kosfeld and colleagues (2005) were also interested whether higher trust among investors means that oxytocin lowers risk aversive behavior or it only rises trust. Thus, they conducted another experiment in which researchers observed that risk aversion behavior did not change under the influence of administered oxytocin. A reason for this outcome can be explained by omitting social interactions in the risk experiment. In other words, while participants in the first experiment were people in both cases (investors and trustees), in the second experiment, the investors did not interact with people but instead they interacted with a random mechanism that assessed a risk that investors undertake.

The next important modulator in decision making processes is a neurotransmitter dopamine. Dopamine plays an important role in a reward system. That means that when we make a decision that proves to be good, we should feel satisfied because dopamine has been released in our brain. The feeling of satisfaction should motivate us to repeat the same decision in similar situation in future. This process, of course, is much more complicated; however, for this paper, the basic knowledge should be sufficient. For example, Robert Rogers (2011) mentions in his work tenths of experiments done on the role of dopamine in decision making. Interesting results from studies revealed that drug addicts have problems to decide when dealing with probabilistic results (Rogers et al., 1999b). Interestingly, even though the incidence is minor, patients who undergo dopamine treatment (e. g. parkinsonian patients who are treated by levodopa) may incline to gambling or other types of compulsive behavior (Dodd et al., 2005). Nevertheless, this treatment helps them to reduce other symptoms of parkinsonism such as tremor that is also caused by insufficient production of dopamine. Moreover, a treatment by a drug called Ritalin that stimulates the central nervous system and is used by people suffering from dementia, helps these patients to decrease a tendency of risk seeking behavior that they often display (Rahman et al., 2005).

Dopamine is a very important element that influences many cognitive processes in brain such as decision making, reward predicting, learning, and many others. For example, an experiment showed that when researchers administered substances which increased the level of dopamine to young adults, they could choose more successfully than those whom a drug that decreased the levels of dopamine was administered (Pessiglione, Seymour, Flandin, Dolan, Frith, 2006). Pessiglione and colleagues explain this difference by lower prediction error among participants whose dopamine levels were artificially increased. Prediction error is a term denoting a difference between outcomes that we predict before an event happens and real outcomes after the event. Consequently, we can assume that dopamine affects the ability and speed of learning.

An important stage in lives of many people is also significantly marked by dopamine. This stage is puberty. In puberty, the body is coming through many changes to achieve a final adult stage and levels of dopamine change abruptly during this period, and thus, perceiving rewards or punishments is quite hyperbolized. Geier, Terwilliger, Teslovich, Velanova, and Luna, (2009) concluded based on their experiment that these exaggerated reactions together with insufficient ability to control these abruptions can lead adolescents towards risky behavior. Insufficient control is most probably underlined by the fact that not all brain structures are fully developed in the adolescence. Specifically, prefrontal cortex is still forming circa till the age of twenty-five (Sapolsky, 2014). Interestingly, frontal cortex is shaped by not only by genes but by environment and experience (Jerison, n.d.; Sapolsky, 2014). Therefore, experience, that is usually limited among teenagers compared to adults, also influence decision-making processes. These three mentioned factors can explain reasons for adolescent risky behavior. Thus, when parents scowl to their teenagers for impulsive and risky decisions, they should consider these factors too and hope that with growing age and maturing frontal lobes, the behavior of their offspring will become wiser.

While frontal lobes and their connection to decision making were previously often mentioned, in this section, we discuss structures involved in decision making in a little bit more detail. First, the key structure for decision making appears to be the **ventromedial prefrontal cortex** (***Figure 3***). Studies of Bechara and colleagues (1997), Damasio (1990), Fellows and Farah (2007), and others credibly supported its crucial role in basic decision-making processes. Moreover, it seems that right ventromedial prefrontal cortex correlates more with decision making, emotions, and social relations than the left side (Martínez-Selva, Sánchez-Navarro, Bechara, & Román, 2006). The second structure, that appears to play a significant role in decision-making together with hippocampus is already mentioned **dorsolateral prefrontal cortex** (***Figure 3***). This area relates to working memory, planning, and other executive functions that are utilized in certain types of decision-making processes. Moreover, as it has been mentioned above, an activity of dorsolateral prefrontal cortex may influence our decision when dealing with unfair bids (Knoch, et al., 2006).

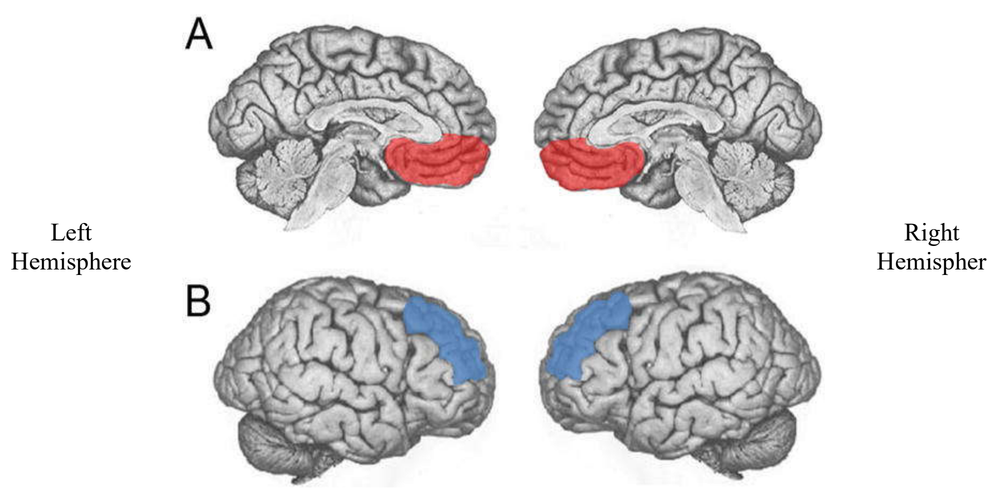


Figure 3: A displays ventromedial prefrontal cortex. B displays dorsolateral prefrontal cortex (Koenigs & Grafman, 2009).

Another important part connected to emotions and decision making is amygdala (***Figure 4***). Amygdala is from the evolution point view a very old structure resembling an almond. Amygdala activates mainly when we are in dangerous and stressful situations. Due to emotions stuck to certain situations, amygdala can trigger automatic reactions in choosing. That means that we somehow omit the conscious part of deciding executed by frontal cortices and simply jump to a decision based on emotions (Martínez-Selva et al., 2006). This decision, however, does not have to be the best. Nevertheless, people with amygdala damaged by a lesion displayed indifferent behavior to rewards and punishments; therefore, their ability to decide is noticeably limited (Gupta, Koscik, Bechara, & Tranel, 2011). Moreover, since these patients cannot recognize the difference between a good and a bad decision, they are unable to learn for future decisions. As a result, they would go for wrong or right choises without realizing it over and over.

Finally, based on the review of Martínez-Selva and colleagues (2006), the last (but not the least) structure that influences decision making is anterior cingulate cortex (***Figure 4***). This part of brain seems to be active when we expect a result of a choice we have made, and interestingly, it becomes more active when we think that the future result will be negative. In addition, when Sanfey and colleagues (2003) conducted an experiment with Ultimatum game, the fMRI results showed that anterior cingulate cortex together with anterior insula and dorsolateral prefrontal cortex were more active, when participants were confronted with unfair bids. Even more interesting fact is that activation in these areas was significantly different when unfair bids were offered by people (higher activity) than when they were offered by an artificial medium – a computer. Consequently, Sanfey and colleagues made a conclusion that people reacts to unfair behavior from other people with higher sensitivity.

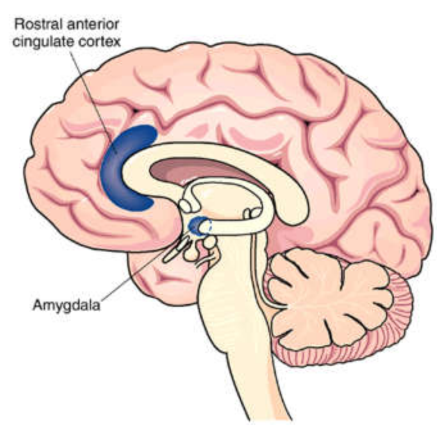


Figure 4: Localization of anterior cingulate cortex and amygdala in brain (Schacter and Addis, 2007).

Eventually, despite many structures that are involved in decision making, we can conclude that both the right side of the frontal and prefrontal cortex are crucial for these processes. Moreover, we pointed out that these structures are influenced by various neurotransmitters and hormones which levels impacts our decision making. We can share all these findings due to advances in neuroscience and medicine that allowed us to dig deeper in observing of what is going on in our heads and where it is happening. Then, thanks to these findings we can get better insight in decision-making processes and test theories such those created by Nicola and Daniel Bernoulli, von Neuman, Morgenstern, Kahneman, and Tversky. In addition, neuroscientific findings can be very helpful in understanding heuristics and cognitive biases that impacts our decisions. First, however, we should take a closer look on what heuristics and cognitive biases are.

1.2 Heuristics

Judgement and decision making is not a free lunch process. It always cost us relative portion energy and relative portion of time and still, an "optimal" decision is not a sure thing. The time is what people often lack not only these days but seconds were also very precious when our ancestors needed to decide whether they would fight that huge mammoth in front of them or run away. However, as our human world has been becoming more complex, also our decision-making processes became more demanding. Therefore, our brain has developed two ways of thinking –fast and slow. The book *Thinking, fast and slow* (2012) by Daniel Kahneman presents these two ways as – "System 1" (the fast and unconscious) and "System 2" (the slow and conscious). He borrowed these terms from Keith Stanovich and Richard West (p. 658, 2000). System 1 is fast, unconscious, automatic, personalized, and in certain situations it leads us to biased thinking and decision making. It uses various heuristics that may resemble a "standby" regime in terms of saving time and energy. System 2, on the other hand, is slow, asks for attention (conscious), and thus, it costs us more energy and time than System 1. System 2 is also analytical, rational, and depersonalized. Therefore, it should eliminate heuristic thinking and immunize us against cognitive biases (Kahneman, 2012; Stanovich & West, 2000). Both systems demonstrate advantages and disadvantages which make them a vital complement of our beings. We, however, will focus more on System 1 which is a fruitful source of heuristics and cognitive biases.

Heuristics are usually automatic shortcuts in our thinking that should ease and fasten our decision-making processes, and thus, they are a part of the System 1 together with cognitive biases. Heuristics also help us to frame complex problems in a simpler view.

Heuristics and biases introduced by Kahneman and Tversky (1974) created a specific approach towards the science of decision making when they clearly displayed irrationality in our decision making and contrasted the normative theories that expected rational choices. Heuristics observed during experiments of Kahneman and Tversky were **representativeness**, **availability**, and **anchoring and adjustment**. For example, representativeness is about a probability that we assess a certain object or a person into a certain group based on certain characteristics and may often lead to a conjunction fallacy. A very typical (but criticized (Stanovich & West, 2000)) example of representative heuristics by Tversky and Kahneman is a case of Linda, who is described as "31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations" (p. 297, 1983). Participants had two options from which they could choose one. The first option says that "Linda is a bank teller" and the second says that "Linda is a bank teller and is active in the feminist movement" (p. 297, 1983). Participants mostly chose the second option over the first even though a mathematically calculated probability that Linda is the bank teller and the feminist at the same time is low. Tversky and Kahneman (1983) explained this behavior as a result representativeness heuristics that led participants to "conjunct" two improbable statements in one based on a description of Linda. In other words, the description of Linda's character might cause a faulty conviction that we have enough information to conclude that she is both a bank teller and a feminist.

Heuristics, unconscious shortcuts or rules of thumb, may lead us to wrong conclusions but would the evolution allow such a process to survive if it always misled us? Gerd Gigerenzer and Peter M. Todd (1999) developed a program of "fast and frugal heuristics" in which they argued that heuristics are not a source of irrational decisions but rather a beneficial part of human reasoning (Albar & Jetter, 2009; Gonzalez, 2014). They promote heuristics as useful tools of decision making but they conditioned their usefulness by ecological rationality. That means that our minds and thoughts adapted to the environment in which we live and strive. Moreover, ecological rationality might also have an evolutionary explanation because it seems that certain heuristics and biases are persistent among people (and monkeys) while others – ineffective ones – might faded away from our minds (Wilke & Mata, 2012). Thus, the environment and also our personal experience adjust heuristics we 'carry in mind' and influence their effectiveness. Moreover, according to Passer and Smith (2004), experience we gain over the lifetime co-creates cognitive maps in our brains. Based on these maps we are able to predict events in our lives. Of course, this ability to predict works only to certain extent unless we are fortune tellers. Nevertheless, our experience and knowledge should lead us to a fact that when we see dark clouds and lightning strikes, we should go someplace safe because a storm is coming. Eventually, heuristics should have strong rational background in our ongoing lived experience and environment.

1.3 Cognitive Biases

Cognitive biases demonstrated a clear deviation between normative (rational) and observed (real) behavior, and so, they have become a critical tool for opponents of normative theories (Kahneman & Tversky 1979; Tversky & Kahneman 1974, Staw, 1981; Šinský, 2010, etc.) The term "cognitive bias" was firstly coined by Daniel Kahneman and Amos Tversky in 1970's when they observed that people, under certain conditions, often make certain irrational decisions (Wilke & Mata, 2012). They observed that cognitive biases are repetitive systematic errors that affect our judgement, choice, and behavior, and lead us to faulty decisions (Arnott, 2006; Tversky & Kahneman, 1974; Wilke & Mata, 2012). Tversky and Kahneman attributed these irrational decisions to using heuristics in our judgmental and decision-making processes. Therefore, they related cognitive biases to unconscious and fast System 1. However, Arkes (1991) suggests that slow, rational, and analytical System 2 leads to a biased behavior, too. Eventually, heuristics do not seem to be the only cause of cognitive biases.

The origin of cognitive biases is most probably mixed. Certain cognitive biases can relate to deeply rooted heuristics such as anchoring or representativeness. The other group of biases seem to relate to our emotions. Neuroscientific evidence also suggests that sensitivity to biases is correlated with emotions and the way how these emotions are processed in our brains (Fujino et al., 2006; Sanfey et al., 2003). The next group of cognitive biases build up during our lives – from the moment we are born, till we die. This kind of biases can arise from implicit learning or repetitive actions and habits that get under our skin (Stanovich, 2011). Moreover, there are various factors that can influence our susceptibility to biases. According to Croskerry (2014), lack of sleep, cognitive overload, and exhaustion can increase our proneness to cognitive biases. In addition, culture, intelligence, our gender, or a character has an impact on our susceptibility to biases (Croskerry, 2014; Ficková, 2011; Mason & Morris, 2010; Strachanová, 2015).

Cognitive biases relate to various thinking processes and this led several scientists to categorize known biases in groups. First, Tversky and Kahneman (1974) identified thirteen biases which they divided into three groups based on a heuristic that people used. Thus, there are cognitive biases connected with heuristics of representativeness, availability, and anchoring. Later, Kotterman, and Remus (1986) proposed another classification because they observed that subjects in decision making were sensitive to how information is presented and how it is processed (Šinský, 2010). Nevertheless, soon, it was clear that these classifications were insufficient due to a growing number of observed biases. For example, David Arnott (2002, 2006) mentioned 37 well-documented biases and presented their wider classification. Based on previous research, Arnott grouped biases in to six categories – "memory, statistical, confidence, adjustment, presentation and situation biases" (p. 2, 2002). Till now, Wikipedia lists 160 different biases that mess up not only with our decision making but with our beliefs, memory and impacts our social interactions (*List of cognitive biases*, 2017).

The list of cognitive biases is quite long and we cannot test them all. Therefore, we choose three well known cognitive biases in our experiment – attribution error, confirmation bias, and sunk cost fallacy.

1.3.1 *Attribution error*

Attribution error belongs to the group of social biases, and thus, it may influence our social interactions. In 1958, Fritz Heider (1958) published a book *The Psychology of Interpersonal Relations* in which he settled basics for an attribution theory. From then, this theory has an influential role in the field of psychology. Heider also suggested that people tend to underestimate situational factors (e. g. accidents, social norms, random events, etc.) and overestimate dispositional factors (e. g. personality of a person, his/her reasons or attitudes when explaining a behavior of another person. For better understanding imagine you have a friend, Thomas, and once he tells you he went to a date with a new girl but she did not show up. We demonstrate a dispositional attribution error the moment we think "she must be a … very rude girl". On the other hand, if we thought "maybe she has got lost, or kidnapped, or her dog has died", we would demonstrate a situational attribution error. The term fundamental attribution error was coined by Lee Ross in his work in 1977. Moreover, the further research suggests that when people evaluate themselves, they reason in the opposite way (Ross, 1977). Thus, they tend to attribute successes they achieved to their personal characteristics while their missteps are rather associated with situational factors.

However, there are also papers which do not agree with previous conclusion. Miller, Smith, and Uleman (1981) impose a tough critique on previous research of attributional biases in which they claim that chosen methods for answering biased participants towards more desired answers. Moreover, the distinction between dispositional and situational factors was not clear. Miller, Smith, and Uleman used open-ended rather than closed-ended attributions that were used in previous studies because as they suggested, open-ended attributions allow researchers a deeper insight into internal mental processes of a participants.

A deeper insight into attribution error was brought by a more cognitive approach. A study done on autists indicates that their impaired ability to understand intentions of other people relates to their inability to 'mentalize' about them (Frith, Morton, & Leslie, 1991). Therefore, one can hypothesize that autists are immune or less prone attribution error (Moran, Jolly, and Mitchell, 2014). On the other hand, if they could name only situational factors, their behavior would be still biased, because we believe that a rationally behaving person would think of both – dispositional and situational attributions in the same proportion.

When we mentalize, we try to 'get in minds' of other people to see what underlies their behavior. Brain areas connected with mentalizing are also called "the social brain" (Adolphs, 2003) - (medial prefrontal cortex, posterior cingulate cortex, temporoparietal junction, and superior temporal sulcus). Thus, Moran and colleagues (2014) were interested whether it is possible to predict attribution error based on brain activity in these social brain areas. Their research found out that spontaneous activity in the medial prefrontal cortex, a part of the social brain, indicated when participants were about to make dispositional attributions. Moreover, they assume that humans are predisposed to think of dispositional factors first due to our social nature.

Finally, the research on attribution error is quite extended and it seems that also culture impacts about tendency to attribution error. Mason and Morris (2010) published a study in which they focused on cultural differences in susceptibility to attribution error. Interestingly, East Asians tend to name more situational factors while tendency of North-Americans is to name more dispositional factors in general. The reason why this happens is most probably hidden in the culture because it impacts "preconscious, spontaneous, automatic mental processes as well as conscious, deliberate, controlled reasoning" Mason and Morris (p. 293, 2010).

1.3.2 Confirmation Bias

Confirmation bias (or myside bias) is a tendency to search for information that is in line with our opinion and ignore facts that are against it. Probably the most famous experiment on confirmation bias were conducted in 1960 by a psychologist Peter Wason

(1960). In the experiment that is now called 2-4-6 task, he said three digits - 2, 4, 6 - to "young adults". Afterwards, he said that these three numbers follow a rule which they should have discovered. They were asked to propose other triplets of numbers to test their hypothesis about the rule. After they proposed the triplets, Wason gave them a feedback whether their numbers followed the rule or not. When participants felt confident about the rule, they could have announced it. Only six in twenty-nine participants could identify the rule in the first announcement. Then, the next ten came with the right rule in the second round. The most problematic issue appeared to be that many participants stayed fixed with the prescribed numbers (2, 4, 6) and generated only similar sequences (8, 10, 12). Thus, they only heard confirmatory responses from Wason which entrenched them in their false hypothesis. Their role, however, was to find a way to falsify their hypothesis. That means that should try to propose a sequence that was different from the prescribed one. In this case, coming up with a triplet 6, 4, 2 would be sufficient to get a negative feedback because the rule was "three numbers in increasing order of magnitude" (Wason, 1960, p. 130).

Another important insight into confirmation bias was done by Lord, Ross, and Lepper (1979). Based on previous research of Mahoney (1977), who observed that even "trained social scientists" are susceptible to confirmation bias, they hypothesized that people tend to adjust facts and data according to their prior beliefs. Thus, Lord, Ross, and Lepper made an experiment in which they recruited forty-eight students from which a half supported a death penalty and another half was against death penalty. Then, each participant read two fictitious studies and evaluated their content. One of the studies demonstrated efficacy of death penalty by lowering number of murder rates and the other claimed that states in which death penalty was institutionalized had higher number murder rates. Results of the experiment showed that students were positively biased towards the study which supported their previous opinion. Consequently, those who were against the capital punishment considered the study which doubted its efficacy a as more convincing than the other study. On the other hand, proponents of death penalty were more convinced by the study which confirmed the effect of death penalty on capital punishment. Finally, researches done by Wason (1960), Lord and colleagues (1979), and many others again supported the thesis that humans are not rational "data-driven processors" (p. 2108). They, however, often rely on their intuitions, beliefs, and previous experience.

Furthermore, confirmation bias also relates to memory because we often remember mainly facts that accord with our beliefs. For example, a similar experiment to the death penalty was done by Frost and colleagues (2015). They, however, focused on memory of participants. The topic of death penalty was exchanged by gun control and participants should express their attitude towards this topic. Then, in the first and second phase, they could choose which from twelve abstracts or posts from social media they preferred to read in an extended version. In the third phase, participants were shown thirty-two titles of abstracts and social media posts and should tell whether they recognize them or not and whether the titles were in the previous materials or no. Results confirmed that participants recalled titles that agreed with their opinion on gun control more often than they did for titles with which they disagreed.

In addition, an interesting research suggests that our susceptibility to confirmation bias can be genetically predetermined. Doll, Hutchinson, and Frank (2011) published a study in which they observed the role of two types of dopaminergic genes that also play a role in reinforcement learning and can lead to confirmation bias. During the experiment, participants delivered their saliva for genetic testing in which dopaminergic genes were identified. Then, participants were asked to complete a cognitive task before which they read a set of instructions. These instructions, however, were misleading, and thus, participants were imprinted by wrong prior beliefs. Results show that participants whose ability to learn reinforced probabilities was higher and followed instructions were more prone to confirmations bias. Higher ability of reinforced learning seems to be influenced by COMT gene which relates to "prefrontal dopaminergic function" and indicates proneness to confirmation bias. The second types of genes are related to "striatal dopaminergic function" and impact instant learning from real outcomes that contrast confirmation bias. Eventually, it is important to mention that building on prior beliefs and experience is a natural way how people learn.

Moreover, despite many negative connotations related to confirmation bias, evidence shows that this bias may be an important cue to social interactions. Dardenne and Leyens (1995) hypothesized that asking matching questions can be a demonstration of empathetic and adaptive behavior. Therefore, they conducted a study in which they wanted to reveal whether people with higher social skills are also more prone to ask questions that match their previous beliefs. In other words, they prefer to ask confirmative questions rather than non-confirmative and diagnostic questions. Participants were asked to choose four questions that they wanted to ask an interviewee to find out whether the interviewee "possesses a particular personality train" (p. 1233). The interviewee was either a professor or an undergraduate student and was presented as an extrovert or an introvert. After participants chose questions, they filled in a personality Self-Monitoring test that was originally developed by Snyder

(1974). Eventually, on average, participants, regardless their result in the Self-Monitoring test, asked more confirmatory questions than non-confirmatory questions. Moreover, those who scored highly in self-monitoring appeared to be more outgoing, socially skilled, and empathetic, and at the same time, they asked more questions that accorded with previous description of the interviewee's personality (extroversive or introversive). Consequently, one may assume that they were also more susceptible to confirmation bias even though the study of Dardenne and Leyens did not consider this as a faulty behavior. However, another study that is mentioned below and focuses on sunk cost fallacy also supports the idea that people who are careful and highly adaptive to their social environment had a higher proneness to the sunk cost fallacy.

1.3.3 Sunk Cost Fallacy

Sunk cost fallacy is another cognitive bias that is best described by an example when we eat a bad lunch not because we are hungry but because we spent a considerable amount of money on it. In other words, we succumb to sunk cost fallacy when we keep investing money, time, energy, or other resources in projects, relationships, etc., which final value is much lower than all incurred costs. However, this behavior is against normative theories because they suggest that we forget about the past losses and invest only to such options which future gains are higher than future costs of the chosen option. On the other hand, for those who consider erroneous decisions to be a result of limited rationality (e.g. Tversky & Kahneman, 1974), sunk cost fallacy is another example that support their hypothesis.

Arkes and Blumer (1985) spent quite a considerable time experimenting with sunk cost fallacy. They conducted 10 experiments in which they observed how people decide under various conditions. One of the experiments contained this scenario:

> "Assume that you have spent \$100 on a ticket for a weekend ski trip to Michigan. Several weeks later you buy a \$50 ticket for a weekend ski trip to Wisconsin. You think you will enjoy the Wisconsin ski trip more than the Michigan ski trip. As you are putting your just-purchased Wisconsin ski trip ticket in your wallet, you notice that the Michigan ski trip and the Wisconsin ski trip are for the same weekend! It's too late to sell either ticket, and you cannot return either one. You must use one ticket and not the other. Which ski trip will you go on?"

In this scenario, 54% of asked college students chose the trip to Michigan even though they would not enjoy it as much as the trip to Wisconsin. Therefore, Arkes and Blumer concluded that more than a half of the participants behaved irrationally when they went for an option which utility was not as high as the utility of the second option – the trip to Wisconsin.

However, according to Barry Staw (1981), sunk cost fallacy is not often based only on one bad decision. He argues that we usually make a series of decisions before we get to the final decision that is probably the fatal one. Michael Roberto (2002), who analyzed a tragic expedition to Mount Everest in May 1996 in which five of twenty-three people died, also concluded that this tragedy was a result of a series of suboptimal decisions … and bad luck. Among the five victims, two were highly skilled and experienced mountain guides and climbers. The next three were regular members of the expedition who paid a considerable amount of money to get on the top of the Mount Everest. One of them, Doug Hansen, had already tried once, unsuccessfully, to get on the top and he was also the last member of the expedition who reached the top but never returned. He said: "I've put too much of myself into this mountain to quit now, without giving it everything I've got." (Roberto, 2002). Thus, Doug Hanses, as many others, became a victim of the sunk cost fallacy due to escalation of commitment that goes hand in hand with the mentioned bias and makes resistance to sunk cost fallacy even harder.

Escalation commitment motivates people to 'throw good money after bad'. In other words, people continue in a bad course of action even though they see that previous decisions have been wrong and led to negative consequences. In essential, it seems to be the reason for the following sunk cost fallacy. Staw and Fox (1977), however, found out that when participants experienced a series of negative outcomes after they were highly committed to an action for a longer period, their commitment declined. This behavior may represent a loss of hope or a lesson learnt from the trial and error approach. The opposite, an increase in commitment, happened when participants faced several negative outcomes but their commitment was low until then. As if they had hoped that with higher commitment, they would have been able to reverse negative results. According to Staw (1981), this behavior can be explained by seeking consistency in decisions we make. In addition, previous experiment of Staw (1976) showed that when participants were those who made previous decisions which led to negative outcomes, in general, they persisted in investing. On the other hand, they were not willing to continue in a project with negative outcomes when previous decisions were made by someone else.

One's personality can also indicate an inclination towards sunk cost fallacy. Fujino and colleagues (2016) tested whether a character of a person can be a predisposition for a higher tendency to succumb sunk cost fallacy. They conducted a neurophenomenological study in which they used a personality test and combined it with fMRI. Results show that participants whose character was high in agreeableness and conscientiousness were more prone to sunk cost fallacy. Though, these two traits have positive influence in the social context because they relate to higher risk aversion and "adherence to social rules" (p. 3). However, they lead to a paradoxical behavior concerning the sunk cost fallacy. Moreover, based on fMRI scans, the researchers relate these two traits and inclination to sunk cost fallacy to higher neural activity in left insula when participants were making decisions about sunk costs. Insula, among others, also processes emotions and especially negative emotions. Finally, it seems that deciding about sunk cost fallacy causes an emotional conflict in people.

1.4 Counterfactual Priming

Counterfactual priming is one of debiasing methods that should reduce certain cognitive biases. It is facilitated by counterfactual thinking during which we produce alternatives to a certain situation. For example, we produce counterfactual thoughts when we contemplate about our failures and simulate options that could lead to a different outcome than the real one. During counterfactual thinking we usually ponder "if only I did not come late, I could get the job" or "at least I was not in a crashed bus, because then, I could have been injured". When we simulate these various options, we use **simulation heuristic** defined by Kahneman and Tversky (1982). They suggested that this heuristic is activated by people who deal with demanding questions in which they have to combine various mental processes to create a simulation. Thus, while other heuristics are usually unconsciously and automatically activated, the simulation heuristic has to be consciously kicked off. Once activated, our ability to create counterfactual thoughts increases.

Counterfactual thinking can also take us two ways – upwards or downwards. When a bad event has happened to us and we think of a positive alternative, it is an upward counterfactual. An example of an upwards counterfactual is already mentioned "if only I caught the bus, …". According to Galinsky and Moskowitz (2000), upwards counterfactuals can strengthen negative emotions because we compare a positive hypothetical alternative with negative reality – that e.g. we have come late. On the other hand, when we use downward counterfactuals, we use a more negative hypothetical alternative. For example, we did not catch the bus and we came late. However, while you were on the later bus, you saw that the earlier bus had an accident and was stuck among other crashed cars. Thus, "at least I missed the bus ….". Thus, this downward counterfactual should make us feel better because we compare a more negative hypothetical result to the real and positive one.

Counterfactual thoughts can influence our emotions; however, they can also impact cognitive biases. Galinsky and Moskowitz (2000) conducted three experiments in which they observed that counterfactual priming had debiasing effects on certain cognitive tasks such as Duncker candle problem and confirmation bias. Nevertheless, it had a biasing impact on Wason card test that was a performance task in which participants tended to behave as in the "2, 4, 6" task and tried to rather confirm their hypothesis than disconfirm. In the following research, Kray and Galinsky (2003) discovered that counterfactual mind-set prevented groupthink and motivate groups to be more deliberate in decision-making than groups that were not primed.

While culture has an impact on biases such the case of cultural differences in attribution error, it may also influence perceiving of counterfactual priming. Since our research is done in Slovakia, here, we list a number or Slovak studies with results on counterfactual priming. First, Strachanova (2017) observed in her work that counterfactual priming had a reducing effect on confirmation bias and attribution error. She also assumed that counterfactual priming can by mediated by analytical thinking; however, this relationship was not confirmed. In addition, counterfactual priming did not reduce the status-quo bias – an emotional bias that results in the preference of original but disadvantageous position rather than moving to a new "equilibrium". On the other hand, the research of Dudeková, Kostovičová, and Konečný (2017) done on financial professionals supported that counterfactual priming decreased status-quo bias and loss aversion while it had no effect on confirmation bias. However, it is important to mention that neither experimental group, nor the control group had a tendency towards confirmation bias. Therefore, evidence could be the reason why counterfactual priming did not affect confirmation bias.

In our research, among others, we would like to find out whether counterfactual priming is dependent on a protagonist in the scenario. Inspired by the study of De Brigard, Spreng, Mitchell, and Schacter (2015), we wanted to know whether counterfactual priming focused on "self" or "others" would impact reduction of biases and number of counterfactual thoughts. De Brigard and colleagues conducted an experiment in which they observed episodic counterfactual thinking under fMRI. Participants generated "if only" alternatives about themselves, close friends/family members, an unknown person (familiar and unfamiliar), and an object. First, the fMRI analysis indicates that when we think about people different brain areas are activated than when we think about objects. So-called 'default network'[1] of the brain, which is related to mental simulations and counterfactual thinking, was active when participants generated counterfactuals about people (***Figure 5***). However, when participants thought about objects, this default network was activated only partially. This outcome also supports a hypothesis that the default network plays an important role in underpinning social interactions.

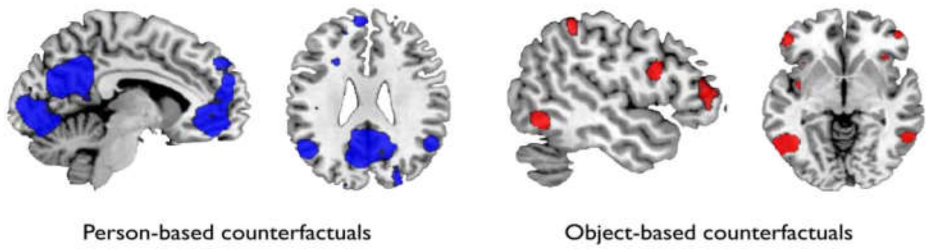


Figure 5: Differences in brain activation during person-based and object-based counterfactual thinking. Blue regions are in line with activation of default network (De Brigard et al., 2015).

Moreover, results de Brigard and colleagues (2015) suggest that when we simulate counterfactual thoughts about us, slightly different brain areas are activated than when we think about others, and the activity of rostral anterior cingulate cortex appears to be the point of difference. According to De Brigard and colleagues, when participants generated counterfactual thoughts about themselves, the rostral anterior cingulate cortex demonstrated higher activation than when counterfactuals about others were generated (***Figure 6***). Nevertheless, the higher activation of this brain area could have been induced by upward counterfactuals that were used in the study. Anterior cingulate cortex, as aforementioned, relates to regret emotion, and regret usually associates upward counterfactuals (De Brigard et al., 2015). Moreover, this suggests that we are more sensitive to our negative experience than to negative experience of others. Other brain regions which activity differed during self-based and other-based counterfactuals were anterior right hippocampus (relates to memory) and medial prefrontal cortex (MPFC). When counterfactuals about self and close friends/family members were created, ventral MPFC was preferentially activated. On the other hand, when counterfactuals about unknown people were generated, lateral and dorsal MPFC were preferentially activated.

[1] According to a study done by Buckner, Andrews-Hanna, and Schacter (2008), the 'default network' is created by ventral medial prefrontal cortex (vMPFC), posterior cingulate cortex (PCC), inferior parietal lobule (IPL), lateral temporal cortex (LTC), dorsal medial prefrontal cortex (dMPFC), and the hippocampal formation.

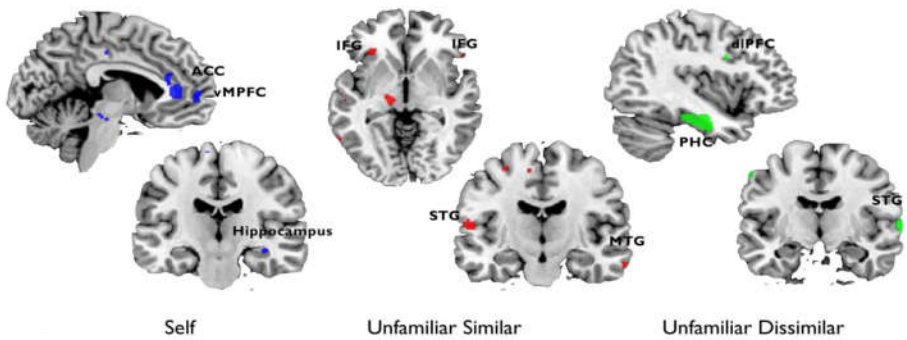


Figure 6: Blue regions represent activation in anterior cingulate cortex and ventral medial prefrontal cortex during counterfactual thinking using self. Red and green areas display activation of brain areas during other and unfamiliar-based counterfactual thinking (De Brigard et al., 2015).

Eventually, we assume that results of our behavioral experiment will be in accord with the neuroscientific evidence provided by De Brigard and colleagues (2015).

1.5 Actively open-minded thinking

Actively open-minded thinking (AOT) is a type of reasoning originally described by Baron (1993). People who score high in AOT are assumed to be more reluctant to certain biases (e. g. confirmation bias) and prior beliefs (Sá, West, and Stanovich, 1999). Moreover, based on results of Haran, Ritov, and Mellers (2013), AOT has also other positive impacts in reasoning of humans. They found out that people high in AOT tended to seek more information when it was available during fulfilling a task. Furthermore, their overall performance in tasks was higher than among those who scored low on AOT. In addition, when making predictions, people high in AOT relied more on the information they previously gathered, were more successful in estimates, and demonstrated lower over-confidence. On the other hand, their reliance on the gained information can lead to mistakes when the information is misleading. Based on these findings, we assume that people who score high in AOT will be naturally more resistant to at least confirmation bias even without counterfactual priming. Moreover, previous results may suggest that these people will be more sensitive to counterfactual priming.

1.6 The goal of the study and hypotheses

Decision making processes are often influenced by cognitive biases that create barriers to optimal choices. Counterfactual priming is a promising debiasing method which effectiveness has been proven in certain biases. Our goal is to verify its effect on the three biases – confirmation bias, attribution effect, and sunk cost fallacy. Moreover, we aim to find out whether there is a difference between self-based and other-based counterfactual priming in an impact on counterfactual thinking and following decision-making because neuroscientific evidence suggests this option. Furthermore, since frontal lobes, which role in decision-making is crucial, develop till the age of twenty-five, we want to observe whether there is a difference in susceptibility to cognitive biases between participants who are younger and older than twenty-five years.

Hypothesis 1: Counterfactual priming has a decreasing effect on attribution error.

Hypothesis 2: Counterfactual priming has a decreasing effect on confirmation bias.

Research Question 1: Does counterfactual priming affect sunk-cost fallacy??

Research Question 2: Are there any differences between the effects of self-based and other-based counterfactual priming on cognitive biases?

Research Question 3: Are there any relationships between actively open-minded thinking and susceptibility to cognitive biases?

Research Question 4: Are the effects of counterfactual priming on cognitive biases moderated by actively open-minded thinking?

Research Question 5: Are there any differences between people below 25 years and above 25 years in susceptibility to cognitive biases?

Research Question 6: Are the effects of counterfactual priming on cognitive biases moderated by age category (25- vs 25+)?

2. Methods

2.1 Pilot Study

Before the final experiment we wanted to test tasks on the three biases to see whether they evoke error responses, and thus, we tested efficacy of the tasks. We created a questionnaire that contained four questions focused on basic sociodemographic information and six scenarios out of which two scenarios should have induced confirmation bias, the other two should have induced attribution error, and the last two should have induced sunk cost fallacy. The pilot questionnaire was finished by 51 participants (37.3% men, 60.8% women, 1.9% other). Based on results from the pilot study we decided to use those scenarios which induction of cognitive biases was higher so that an effect of the intervention could impact a larger range.

Concerning the confirmation bias, we decided to go for an option with Tereza and Speed Dating event because 42% of all participants appeared to succumb to this bias. In the attribution error, we decided to use the scenario with Adam and his unpredictable colleague. We chose this option due to significant disproportion of situational and dispositional reasons that participants listed in the task. Finally, for the sunk cost fallacy, we chose the scenario about further investment into a project of electric cars which was not included in the pilot study. The reason for this decision was that our pre-tested sunk cost fallacy tasks did not provide sufficient confidence about their efficacy. Therefore, we chose the mentioned task that was previously successfully tested by Strachanová (2017). In all three scenarios, we kept the main storyline; however, we made small adjustments to increase ecological validity. All scenarios are described in more detail in the section *2.4 Materials*.

2.2 Participants

We powered the experiment to detect at least a medium-sized intervention effect ($\alpha = .05$, $1-\beta = .95$, two-sided tests). Two hundred and seventy-seven people participated in our research. We excluded one of them due to missing responses. Our final sample consisted of 276 participants among whom 57.2% (n = 158) were women and 42.8% (n = 118) were men. Age of the participants ranged from 15 to 66 years (Mdn = 31, IQR = 13), with roughly one third being in the age category 18 to 27 years, another third between 28 and 36 years, and the participants in the last third were at least 37 years old. Nearly 3% of people (n = 7) had either primary education or finished some vocational high school, 29.3% (n = 81) graduated from high school with a certificate of school-leaving examination, and 68.1% (n

= 188) of the sample had a university degree. Approximately a half of the sample consisted of people who studied social sciences or humanities (*n* = 144), one third were students and graduates of natural or technical disciplines (*n* = 92), and the rest of them belonged to other specializations (*n* = 40). In addition to 27 full-time students, the most frequently represented occupation categories were information technologies (*n* = 28), education (*n* = 23) and health care (*n* = 19).

2.3 Design and procedure of the final experiment

Our goal was to determine and compare effectiveness of the counterfactual priming containing "self" and "others" on the three cognitive biases – confirmation bias, attribution error, and sunk cost fallacy. To achieve this, we used a 3x2x2 between-subject experimental design which participants were randomly divided into three groups after they answered basic sociodemographic questions (***Figure** 7*). A participant in each of the three groups had to answer one of the critical problems that were supposed to induce either confirmation bias, attribution error, or sunk cost fallacy. When participants answered the first task, they completed a shortened (17-item) version of Actively opened-minded thinking test. Afterwards, they were randomly primed by one of the two counterfactual scenarios focused on "self" (*n* = 142) or "others" (*n* = 134). At last, participants had to solve the final critical problem inducing a biased behavior. However, the last task could not be identical with the task which participants solved before the intervention. For example, when they solved the task on confirmation bias first, then, the next task had to be either on attribution error or sunk-cost fallacy.

Thus, a strategy for the task distribution was as follows: Approximately a third of participants solved a problem focused on contribution bias, attribution error, and sunk cost fallacy before the intervention. By this, they became a control group for each bias. Then, they were subjected to intervention in the form of counterfactual priming focused on "self" or "others". After priming, participants became either an experimental group 1 (primed by self-based scenario) or experimental group 2 (primed by "others" scenario). All tasks were randomly but evenly distributed; therefore, each task was completed by approximately one third of participants.

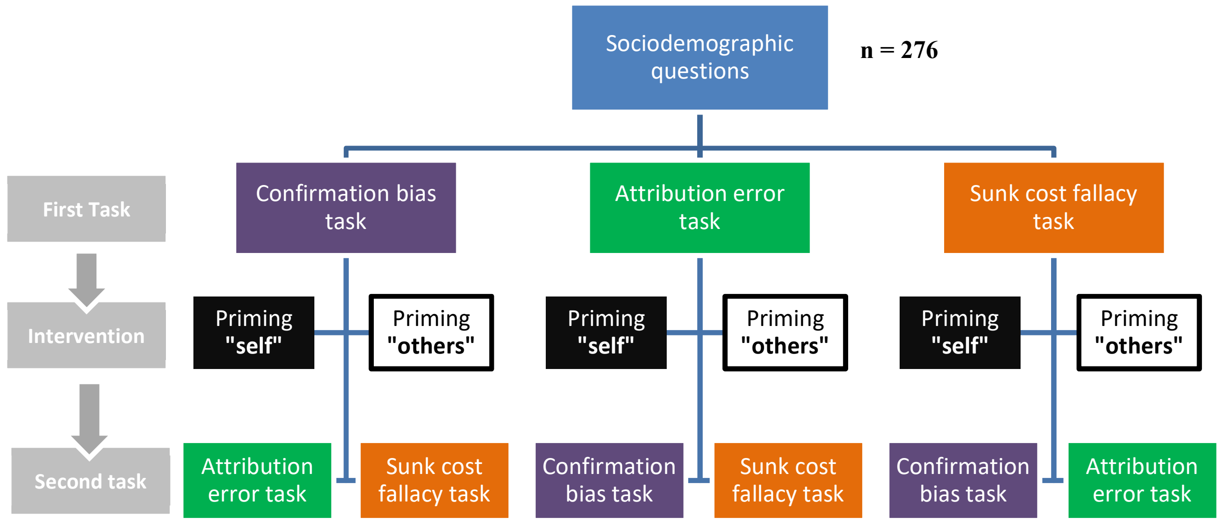


Figure 7: The scheme of the experimental design.

2.4 Materials

All materials used in the study were previously tested except the scenario on counterfactual priming. In the following sections, we describe the used assignments.

2.4.1 Counterfactual priming

Even though majority of previous studies (Dudeková, et al., 2017; Galinsky & Moskowitz, 2000; Kray & Galinsky, 2003, Strachanová, 2017) used counterfactual priming focused on the past event, we used two scenarios which referred to future – one focused on "self" and another focused on "others". We believed that even future, although there is no point of regret connected with the real past event, could kick off simulative thinking because people had to generate alternative ends of a fictive story.

In the self-based scenario, participants should imagine that they have been unemployed for last six months and except other expenses, they have to pay a mortgage for their flat. Thus, now, they are quite broke and need a job. However, they have a second-round interview today in a prosperous company. Moreover, they have certain hints that they are favorited among other candidates. In the morning, as chance would have it, they receive a call from another company that is searching for candidates on a similar position but with a higher salary and interesting benefits. The problem is that the managing director of this company is very busy and do not have time for personal interview but today and at the same time as the former interview in the first company has been scheduled. Participants do not have much time for thinking and they know that both positions are not sure. Therefore, they should come up with alternatives about what they could gain or lose if they canceled their interview in the first company. After participants read this scenario, they should start to generate several alternative ends of this story.

The other-based scenario had the same formulation except participants themselves were substituted by an unknown character – Thomas. Participants again should generate thoughts about possible alternatives that could happen if he rejected the former interview.

2.4.2 Attribution error

For the attribution error, we used a scenario with Adam who is about to cooperate with a new colleague on a project. The colleague of Adam is hardworking, diligent, and punctual; however, as Adam has heard from other colleagues, he can be very unpleasant and his moods are difficult to predict. Adam yet has not had a chance to know his colleague better. Therefore, he ponders what reasons can lead his new colleague to such behavior.

In this task, participants should write four potential reasons that would explain changing behavior of the Adam's colleague. Then, their answers will be recoded depending on a type of the answer. If the answer relates to personality of the Adam's colleague, we mark it as a dispositional factor (e. g. "He is simply moody."). When it refers to external circumstances, we mark it as situational factor (e. g. "Maybe he has problems in his family.").

2.4.3 Confirmation bias

Confirmation bias scenario is about introverted Tereza who, under the influence of her friend, goes to a speed dating and is inspired by Snyder and Swann (1978) and Galinsky and Moskowitz (2000). At Speed Dating she can meet various men, however, she has only a minute to know them. Before the event starts, Tereza sees a sympathetic man in a hallway, who stands alone, reads newspaper and looks a bit off color. She assumes based on his behavior that he might be an introvert like she is. She knows that she will not have much time for questions when the event starts, therefore, she wants to prepare few in advance.

At this point, participants should choose four in ten offered questions to help Tereza estimate the real character of the man. Four of the offered questions are formulated in a way which confirms a hypothesis of Tereza – that is - the man is rather an introvert (e. g. Do you like to be at home in the evening?). The other four questions disconfirm her hypothesis about the man (e. g. Do you like meeting new people?). The last two questions were catchers which neither confirm nor disconfirm the hypothesis (e. g. Do you like cooking?).

2.4.4 Sunk cost fallacy

In the sunk cost fallacy, we chose the scenario about further investment into a project of electric cars that was inspired by Arkes and Blumer (1985) and previously successfully tested by Strachanová (2017). In this scenario, an R&D director should decide in which project his employer (an automotive company) should invest 10 million EUR to become number one in the market. There are two projects between which he decides. The first is an investment into a new model of an electric car. The second is a project of a hydrogen model. The R&D director decides for the electric car and when 80 % of the total sum is invested, he finds out that another automotive company has just revealed a new electric car that seems to be better in all aspects. Consequently, a chance that the company will become the first in this market segment is quite unlikely. However, there are still 2 million EUR which are not invested and the company should decide, whether it will invest the remaining money into the started project.

In this last cognitive bias task, participants should choose on an 11-point Likert scale whether the company should invest remaining two million EUR into the project of the electric car or no. The scale starts with 0 ("The company **definitely should not invest** the money into the project.") and ends with 10 ("The company **should definitely invest** the money into the project.")

2.4.5 Actively open-minded thinking

The original version of AOT developed by Sá, West, and Stanovich (1999) has 41 items with 8 subscales which assess a type of reasoning. Haran and colleagues (2013), however, used only 7-item scale. Since 7-item scale may be quite risky for our needs due to lower reliability, we decided to employ a 17-item version with 4 subscales – dogmatism, fact resistance, liberalism, and belief personification – recently developed and validated by Svedholm-Häkkinen and Lindeman, (2017).

Besides computing a total score of the actively open-minded thinking (17 items), we also quantified 4 AOT factors. The higher was the score, the higher was the ability to think open-mindedly. At the same time, the higher was the level of AOT, the lower is a tendency towards dogmatism (6 items), fact resistance (5 items), and belief personification (3 items). In contrast, the higher the level of AOT, the greater a tendency towards liberalism (3 items).

The overall scale had a high level of internal consistency, α = .80. For the first two factors, the reliability was sufficient: **dogmatism: α = .69**, **fact resistance: α = .75**, the remaining two, however, were not reliable enough: liberalism: α = .26, belief personification: α = .55. Consequently, the latter two subscales were excluded from further analyses.

3. Results

3.1 Counterfactuals

Distributions of numbers of counterfactual ideas produced in the two groups ("self" vs "others"; ***Figure 8***) were non-normal. Participants in the "self" group listed 0 to 9 counterfactuals (*Mdn* = 3, *IQR* = 2). Roughly half of participants mentioned 2 counterfactual thoughts or less, and only 8% (*n* = 11) generated more than 5 thoughts. Participants of the "others" group listed 0 to 12 counterfactuals (*Mdn* = 3, *IQR* = 3) and roughly half of them mentioned 3 counterfactual thoughts or less. One fifth (*n* = 25) generated more than 5 thoughts. **Difference between the two groups was significant: priming "others"** (*MRank* = 150.7) **led to substantially higher number of counterfactuals than priming "self"** (*MRank* = 127.0), *M-W U* = 7876.5, $p = .012$, $r_m = .15$.

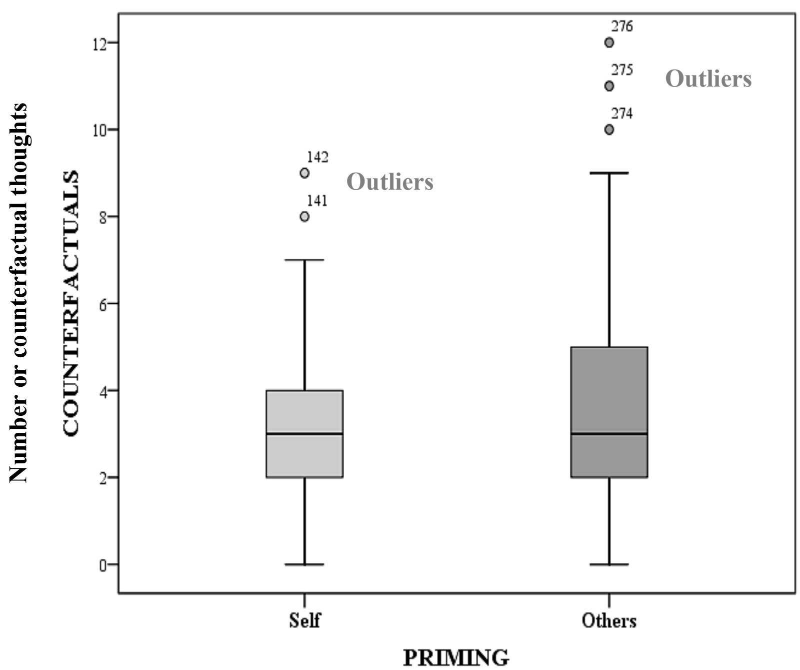


Figure 8: Counterfactuals produced in the two groups with different types of priming

3.2 Attribution error

First, we identified one extreme outlier and excluded the respective answer from following analyses. Descriptive characteristics are shown in ***Table 1.*** Neither numbers of the two factors (internal and external), nor the difference between them was in line with the assumptions of normal distribution. Therefore, we used nonparametric tests.

Table 1. Attribution error - descriptive statistics: ranges, medians and interquartile ranges

	n	**# Internal factors [IF]**	**# External factors [EF]**	**| IF - EF | = Attribution error**
Control group [CG]	89	0-5, 3 (2)	0-4, 2 (2)	0-4, 1 (1)
Experimental group [EG]	91	0-7, 3 (3)	0-6, 2 (2)	0-6, 1 (2)
Experimental group 1 [EG1]	38	0-7, 2.5 (3)	0-6, 2 (2)	0-4, 1 (2)
Experimental group 2 [EG2]	52	0-7, 3 (2)	0-5, 2 (2)	0-6, 1 (1.75)

Control (CG) and experimental (EG) group did not differ in the number of IF reasons (*M-W U* = 3594.0, $p = .223$, $r_m = .09$) and number of EF reasons (*M-W U* = 3942.0, $p = .851$, $r_m = .01$). Yet, they differed significantly in the level of the attribution error. Specifically, members of CG (*MRank* = 98.5) committed the attribution error to much greater extent than the members of EG (*MRank* = 81.6), *M-W U* = 3250.5, $p = .024$, $r_m = .17$). Thus, **counterfactual priming substantially decreased the attribution error, with small effect size.**

As for the comparison of the three groups – control group (CG), experimental group 1 (EG1 = priming “self”), and experimental group 2 (EG2 = priming “others”), the differences in the numbers of IF reasons and EF reasons were small but insignificant, IF: $H(2) = 2.06$, $p = .357$, $r_m = .11$; EF: $H(2) = 4.46$, $p = .107$, $r_m = .16$. Nevertheless, the differences in total scores were significant, $H(2) = 7.62$, $p = .022$, $r_m = .21$. Thereafter, we proceeded with pairwise tests. First, we found that EG2 (“others”) did not significantly differ from CG in any of the three variables. However, members of EG1 (“self”; *MRank* = 50.8) showed substantially lower level of attribution error than members of CG (*MRank* = 69.6), *M-W U* = 1190.0, $p = .006$, $r_m = .24$. In addition, number of EF reasons was significantly higher among EG2 (“others”; *MRank* = 50.1) than among EG1 (“self”; *MRank* = 39.3), *M-W U* = 751.0, $p = .047$, $r_m = .21$. Consequently, **counterfactual priming with self-based scenario substantially decreased the attribution error, with small to medium effect size.** On the other hand, **other-based counterfactual priming facilitated reflection of situational factors to a greater extent compared to counterfactual priming “self”.**

3.3 Confirmation bias

The number of introversion-related questions (IRQ) and number of extraversion-related questions (ERQ) in each group ranged from 0 to 4. Thus, the difference between the two numbers, i.e. the level of confirmation bias (CB), in each group ranged from -4 to 4. All the three indicators were normally distributed. Descriptive characteristics are summarized in ***Table 2.***

Table 2. Confirmation bias - descriptive statistics: means and standard deviations

	n	**IRQ**	**ERQ**	**CB**
Control group [CG]	103	2.1 (0.9)	1.6 (0.9)	0.5 (1.7)
Experimental group [EG]	91	1.9 (0.9)	1.7 (0.9)	0.2 (1.7)
Experimental group 1 [EG1]	54	2.0 (0.8)	1.7 (0.9)	0.3 (1.6)
Experimental group 2 [EG2]	37	1.8 (0.9)	1.9 (0.9)	-0.1 (1.7)

It is important to point out that our sample, in general, was not prone to confirmation bias. Roughly 58% did not choose more IRQ than ERQ, and only 8% chose three or four IRQ.

Differences between CG and EG in IRQ: $t(192) = 1.29$, $p = .198$, $d = 0.19$, ERQ: $t(192) = -1.39$, $p = .167$, $d = 0.20$, and CB: $t(192) = 1.43$, $p = .156$, $d = 0.21$, were small but insignificant. Specifically, members of CG chose slightly more IRQ and slightly less ERQ. As a result, their tendency toward CB was mildly higher than in EG. **Counterfactual priming slightly decreased the confirmation bias, with small effect size but without statistical significance of the result.**

As for the comparison of the three groups, the differences in IRQ: $F(2, 191) = 1.15$, $p = .319$, $\eta^2 = .02$, ERQ: $F(2, 191) = 1.58$, $p = .209$, $\eta^2 = .02$, and CB: $F(2, 191) = 1.53$, $p = .219$, $\eta^2 = .02$, were small but insignificant. Then, we proceeded with pairwise post-hoc tests, while most of them provided statistically insignificant results. However, we found two marginally significant comparisons. First, between CG and EG2 (“others”), EG2 produced slightly more questions against the prior belief (i.e. ERQ), $MDiff = -0.3$, $p = .077$, 95% CI [-0.7, <0.1], $d = 0.34$, and exhibited slightly lower level of the bias, $MDiff = 0.6$, $p = .083$, 95% CI [-0.1, 1.2], $d = 0.33$. Eventually, **other-based counterfactual priming slightly**

decreased the confirmation bias, with small effect size but without statistical significance on the result.

3.4 Sunk cost fallacy

People responded on an 11-point Likert scale. The answers were non-normally distributed and ranged from 0 to 10 in all the groups. Control group (CG; n = 84, *Mdn* = 8, *IQR* = 6, *MRank* = 89.5) and experimental group (EG; n = 94, *Mdn* = 7, *IQR* = 5, *MRank* = 89.6) exhibited almost identical level of sunk cost fallacy, *M-W U* = 3943.5, p = .989, r_m < .01. **Counterfactual priming failed to reduce the sunk cost fallacy.**

Distributions of the responses in the CG, the EG1, and EG2, are depicted in ***Figure 9***. Overall comparison found insignificant differences among the three groups, $H(2)$ = 1.21, p = .545, r_m = .08. Similarly, two pairwise tests between control group and experimental groups brought no significant findings, CG-EG1: *M-W U* = 1936.5, p = .567, r_m = .05, CG-EG2: *M-W U* = 1764.0, p = .529, r_m = .06. Nevertheless, there was a small difference between the two experimental groups, EG1-EG2: *M-W U* = 954.0, p = .257, r_m = .12. **Counterfactual priming "others" led to slightly more answers in line with the sunk cost fallacy compared to counterfactual priming "self".**

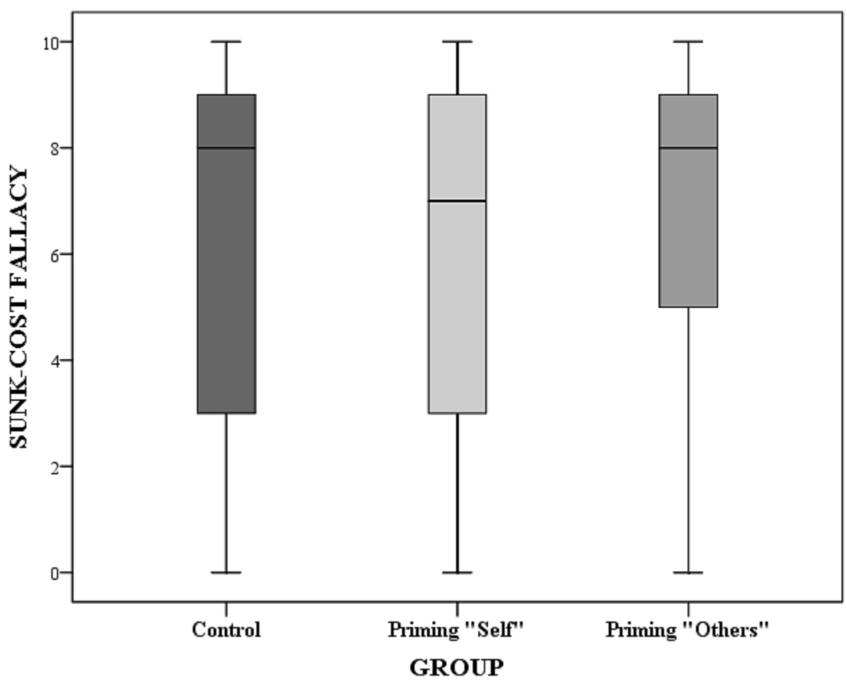


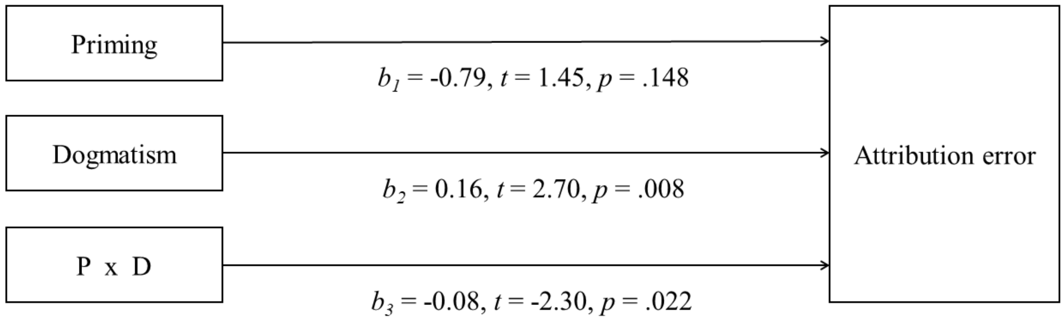
Figure 9: Differences in answers to sunk-cost fallacy problem. Differences between groups were insignificant; even though participants primed by other-based scenario displayed a slightly higher tendency towards sunk-cost fallacy than participants primed by self-based scenario

3.5 The role of actively open-minded thinking

First, we conducted a series of correlation analyses to observe relations among AOT and other variables. Second, we focused on moderation analyses (Hayes, 2013) to examine whether the effects of intervention depend on actively open-minded thinking (AOT) and its factors (dogmatism - DG & fact resistance - FR). Concerning mutual correlations, we found that production on counterfactuals was positively linked to AOT, $r_S = .23$, $p < .001$, and negatively linked to dogmatism, $r_S = -.17$, $p = .004$, fact resistance, $r_S = -.18$, $p = .003$, and age, $r_S = -.14$, $p = .015$. Moreover, the level of attribution error positively correlated with dogmatism, $r_S = .25$, $p = .017$. No significant correlation was found between AOT and confirmation bias and sunk cost fallacy.

In case of moderation effect of AOT, we found out that the effect of counterfactual priming as such and counterfactual priming by “self” on attribution error were moderated by dogmatism. As can be seen in ***Figure 10***, the effect of counterfactual priming on the attribution error is negative and increases with dogmatism. Specifically, counterfactual priming reduced attribution error among people with medium and high dogmatism. Yet, it did not affect people with a low level of dogmatism. The effect of dogmatism and the effect of the interaction of the intervention and dogmatism were significant. The model was significant and explained 7% of variation in attribution error. This pattern was also visible in counterfactual priming by “self” (***Figure 11***). Consequently, the effect of counterfactual priming “self” on the attribution error is negative and increases with dogmatism. Specifically, counterfactual priming “self” reduced attribution error among people with medium and high dogmatism. Yet, it did not affect people with a low level of dogmatism. The effect of dogmatism and the effect of the interaction of the intervention and dogmatism were significant. The model was significant and explained 11% of variation in attribution error. Finally, moderation analysis with counterfactual priming “others” led to similar results but this time without significant relations.

STATISTICAL DIAGRAM

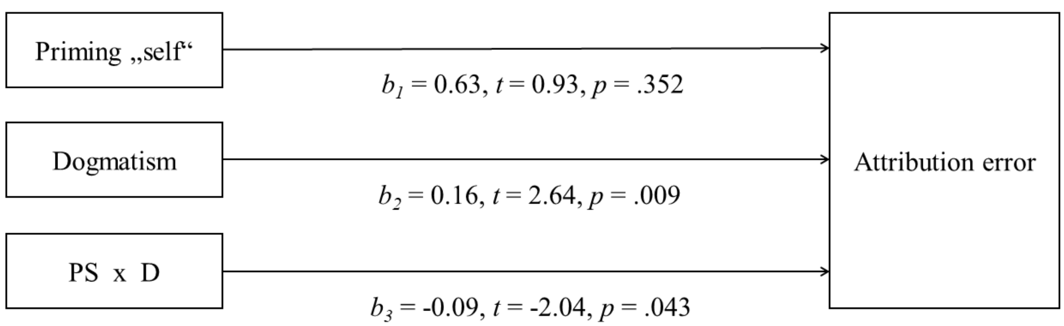


Conditional effects:

Dogmatism	**effect**	***t***	***p***	**95% *CI***
Low	0.02	0.08	.941	[-0.48, 0.52]
Medium	-0.40	-2.21	.029	[-0.75, -0.04]
High	-0.81	-3.18	.002	[-1.31, -0.31]

Figure 10: Dogmatism as a moderator of the effect of counterfactual priming on the attribution error

STATISTICAL DIAGRAM



Conditional effects:

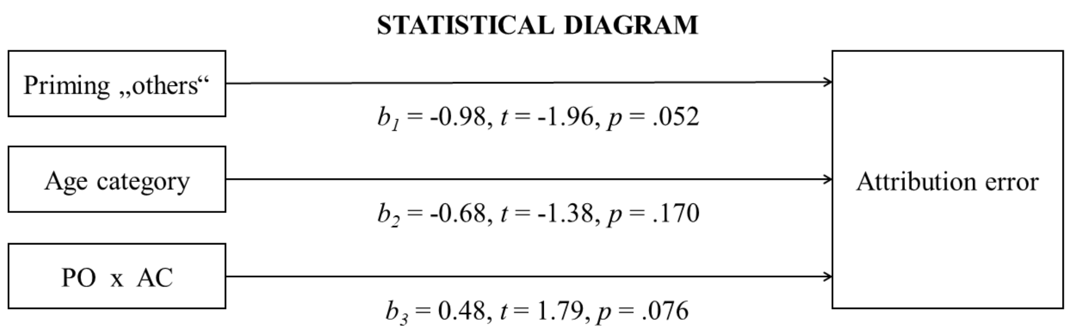
Dogmatism	**effect**	***t***	***p***	**95% *CI***
Low	-0.19	-0.58	.566	[-0.83, 0.46]
Medium	-0.63	-2.82	.006	[-1.08, -0.19]
High	-1.08	-3.61	<.001	[-1.67, -0.49]

Figure 11: Dogmatism as a moderator of the effect of counterfactual priming “self” on the attribution error

3.6 Age differences

We compared the results of participants younger and older than 25 years. We did not find any significant differences between these two groups. Thus, the two groups did not significantly differ in generating counterfactuals, AOT scores, and susceptibility to the three biases. However, we observed a moderation role of the age in the counterfactual priming.

Although the effect of priming by "self" was not moderated by age category, priming by "others" displayed such moderation (***Figure 12***). Priming by "others" was effective only in the case of younger participant, and it was substantially higher among younger than older participants. The effect of priming and the effect of the interaction of the intervention and age category were marginally significant. The model was significant but explained only 3% of variation in attribution error.



Conditional effects:

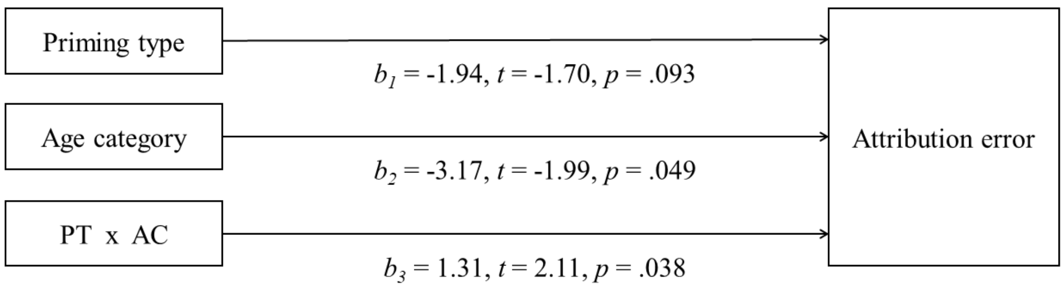
Age category	**effect**	***t***	***p***	**95% *CI***
25-	-0.49	-2.04	.043	[-0.97, -0.02]
25+	-0.01	-0.06	.952	[-0.25, 0.24]

Figure 12: Dogmatism as a moderator of the effect of counterfactual priming "self" on the attribution error

Moreover, we found out that different effects of counterfactual priming ("self" versus "others") were moderated by age category in the case of attribution error (***Figure 13***) and sunk cost fallacy (***Figure 14***). In case of attribution error, younger participants were slightly less (but non-significantly) susceptible to attribution error when they were primed by "others" than when primed by "self". Among older participants, however, the level of attribution error was substantially higher after priming by "others" compared to priming by "self". The effect of priming type was above statistical significance, whereas the effects of age category and their interaction were significant. Thus, the model was significant and explained 8% of variation in attribution error.

In case of sunk cost fallacy, younger participants were mildly (but non-significantly) more resistant to this bias after priming by "others" compared to priming by "self". On the other hand, among older participants, the level of sunk cost fallacy was substantially higher after priming "others" compared to priming "self". The effects of priming type, age category, and their interaction were significant. The model was significant and explained 9% of variation in sunk cost fallacy.

STATISTICAL DIAGRAM

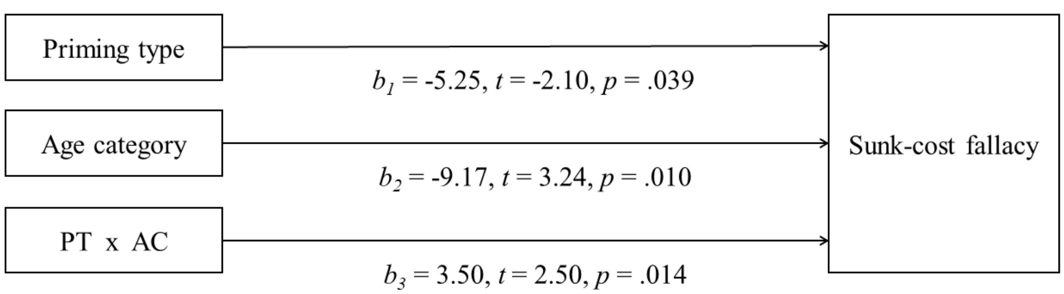


Conditional effects:

Age category	**effect**	***t***	***p***	**95% *CI***
25-	-0.62	-1.13	.261	[-1.72, 0.47]
25+	0.69	2.36	.020	[0.11, 1.27]

Figure 13: Age category as a moderator of the different effects of the two priming interventions on attribution error

STATISTICAL DIAGRAM



Conditional effects:

Age category	**effect**	***t***	***p***	**95% *CI***
25-	-1.75	-1.46	.148	[-4.14, 0.63]
25+	1.75	2.43	.017	[0.32, 3.17]

Figure 14: Age category as a moderator of the different effects of the two priming interventions on sunk cost fallacy.

3.7 Gender differences

As for the gender differences, answers of women were significantly more biased in the sunk cost fallacy problem (♂: *Mdn* = 7, *IQR* = 5, *MRank* = 35.1; ♀: *Mdn* = 8, *IQR* = 5.75, *MRank* = 49.3), *M-W U* = 582.5, p = .003, r_m = .30. Yet, women performed slightly but non-significantly better in the confirmations bias problem (♂: M = 0.9, SD = 1.6; ♀: M = 0.3, SD = 1.8), $t(101)$ = 1.64, p = .104, d = 0.33, and in the attribution error problem (♂: *Mdn* = 2, *IQR* = 2, *MRank* = 49.6; ♀: *Mdn* = 1, *IQR* = 1, *MRank* = 41.4), *M-W U* = 794.5, p = .124, r_m = .16.

4. Discussion

Saving innocent kitties is a noble act; however, when we link it to saving human lives, this research gains different dimensions. Croskerry (2014) warns that medical doctors may easily become a victim of biased behavior when dealing with patients. Constant stress and exhaustion can lead, except others, to confirmation bias. Consequently, a patient can be wrongly diagnosed and die. Roberto (2002) described how various cognitive biases led five people to death during the Mount Everest expedition. Of course, in those altitudes which have tremendous impact on living organisms that are used to different conditions is difficult to keep a rational mind. However, rationality and objective reasoning are inevitable when lives are at stake. Moreover, these qualities are important when dealing with difficult social issues such as discrimination, xenophobia, or racism. In these days, we face so much negative information from many sides and keeping calm might be demanding. Growing fear creeps in our minds and influences our thinking. As a result, we exchange our natural inclination to humanity for an isolation that may give us a false feeling of security. However, this should not be our new nature. We should be wise and think open-mindedly, and counterfactual priming appears to the tool that can help us in this endeavor.

Thus, the goal of our research was to find out whether counterfactual priming can decrease levels of the three cognitive biases – attribution error, confirmation bias, and sunk cost fallacy. Our results demonstrate that counterfactual priming decreased attribution error and slightly decreased confirmation bias that partly accords with Strachanová (2017). However, the effect on confirmation bias was not significant. At the same time, our participants, in general, did not demonstrate a high tendency towards this bias. Thus, there was not much to reduce. Similar conclusion was suggested by Dudeková and colleagues (2017) because the tendecy to confirmation bias was also low among participants in their study. The last cognitive bias – sunk cost fallacy – was not reduced by counterfactual priming at all. We assume that a problem can be hidden already in the task because its ecological validity can be low. Moreover, its formulation could not be adequate to sunk costs. In addition, it is important to mention that scenarios we used were hypothetical and referred to future. Thus, they were not like typical scenarios used in previous studies which referred to past (Dudeková et al., 2017, Galinsky & Moskowitz, 1999, Kray & Galinsky, 2003, Strachanová, 2017). Moreover, majority of participants probably did not have an experience with a situation like the one described by scenarios, and thus, their imagination might have been limited. Yet still, the situation used in the original scenario did not have to be familiar to participants either.

The next question we wanted to answer was whether a self-based and other-based character in counterfactual priming had different effects on these cognitive biases as neuroscientific evidence suggested (De Brigard et al. 2015). Our results indicate that self-based counterfactual priming significantly facilitated reduction of attribution error while other-based priming led to generating more situational factors. In addition, other-based scenario slightly but insignificantly facilitated reduction of confirmation bias. In sunk cost fallacy, however, other-based scenario led to mildly higher tendency towards this bias. Thus, the self-based scenario appeared to be more effective in attribution error and sunk cost fallacy while confirmation bias was insignificantly more sensitive to the other-based scenario.

Interestingly, we found out that counterfactual priming with other-based scenario led participants to generate more counterfactual alternatives than self-based scenario. We assumed that generating more alternatives would be connected with stronger simulation mind-set, and thus, it would reduce the tendency towards cognitive biases. Nevertheless, it seems that generating more alternatives does not lead to higher resistance to biases. Possible reason for the differences in the effectiveness of both scenarios can be in neural correlates. Maybe the brain regions activated during self-based scenario could create a different mental set-up which led to higher resistance to attribution error. Moreover, imaging ourselves could remind us how we would like to be treated if we were in shoes of Adam's colleague.

Besides the role of counterfactual priming, we were also interested in the role of actively open-minded thinking and its effect on cognitive biases. We found out that only two in four AOT factors had sufficient internal consistency – dogmatism and fact resistance. The other two – liberalism and belief personification – were, therefore, excluded. Our results appear to be in line with previous studies which support the negative correlation between AOT and cognitive biases (Haran, Ritov, & Mellers, 2013; Sá, West, and Stanovich, 1999; Svedholm-Häkkinen & Lindeman, 2017). In our research, we observed that the higher was the AOT, the higher number of counterfactual thoughts was produced by participants. In contrast, the lower number of counterfactual thoughts was produced, the higher scores participants gained in dogmatism, fact resistance, and the lower was their age. In addition, we found out that higher level of dogmatism also positively correlated with attribution error. Results also indicate that dogmatism moderated the effect of counterfactual priming as such and self-based counterfactual priming in the attribution error. Specifically, counterfactual priming reduced the attribution error among participants who had high or medium scores in dogmatism. In addition, self-based counterfactual priming decreased attribution error among the same group of participants.

Eventually, we were interested whether development of frontal lobes which lasts till the age of twenty-five (Sapolsky, 2014) influences reduction of cognitive biases. Results suggest that age category did not correlate with the number of generated counterfactuals, susceptibility to examined biases, or AOT. However, we found out that the effect of counterfactual priming was moderated by the age category. First, other-based priming on attribution error (compared to the control group) was effective only in the case of younger participants. Additionally, other-based priming led to higher attribution error and higher sunk-cost fallacy compared to self-based priming, but only among older participants.

Results about gender differences indicate that women were more prone to sunk cost fallacy than men what is in line with findings of Strachanová (2015). In her study, women were aslo more prone towards this bias. Moreover, they declared a higher regret when dealing with sunk cost fallacy than men. This also accord with Ficková (2011) who observed that women displayed higher regret when dealing with upward counterfactuals (Ficková, 2011). A potential explanation of this behavior is that higher regret of sunk costs motivate participants to continue in further investments.

To summarize our main findings, counterfactual priming reduced the attribution error (***research hypothesis 1***), but it failed to substantially decrease the confirmation bias (***research hypothesis 2***) and the sunk cost fallacy (***research question 1***). As for the differences between priming "self" and priming "others" (***research question 2***), only priming "self" substantially decreased the attribution error compared to the control group. Conversely, only priming "others" decreased the confirmation bias compared to the control group - however, the difference was marginally significant this time. Next, the level of counterfactual thinking was positively linked to actively open-minded thinking and negatively associated with dogmatism and fact resistance. Moreover, dogmatism was positively related to the attribution error (***research question 3***). As for the actively open-minded thinking as a moderator (***research question 4***), both the effects of counterfactual priming in general and priming "self" on attribution error were moderated by dogmatism. Priming reduced the attribution error among those with a medium and a high level of dogmatism but it did not affect those who scored low in dogmatism. Further, participants below 25 years and above 25 years did not differ in susceptibility to cognitive biases (***research question 5***), production of counterfactuals or actively open-minded thinking. Yet, the age category moderated the effects of counterfactual priming on cognitive biases (***research question 6***).

Despite interesting results, we have to admit that certain limitations occurred in our research. First, even though we tried to prepare ecologically valid tasks, it is highly possible that in the case of sunk cost fallacy we have not succeeded. Second, our sample did not represent a typical sample of Slovak population because 68.1% (n = 188) of our participants were graduates from universities. In contrast, roughly one fifth of the Slovak citizens attained a university degree according to OECD (2015); thus, this disproportion could have an impact on our results since education can influence reluctance to cognitive biases. Therefore, for the future research, we should target a more balanced sample of Slovak population. The next limitation potentially occurred due to too general scope of our research. In future, we should try to focus more on a specific area like Dudeková and colleagues (2017) did and detect incident cognitive biases. Dudeková and colleagues (2017) tested counterfactual priming on financial specialists. We assume that there are also other areas that can benefit from the knowledge about cognitive biases and debiasing strategies.

As suggested by Croskerry (2014), health care is a demanding area in which increasing the knowledge about cognitive biases and their reduction can save lives of people and conscience of doctors. In education, teachers can inform students about cognitive biases, discuss their effects on behavior, and teach effective methods for objective reasoning. Moreover, schools can support students in actively open-minded thinking to strengthen prevention against cognitive biases. Last but not least, programmers, who develop various systems, applications, and state-of-the-art decision-making tools for businesses, should be interested in this topic as well. Since programmers are humans we have to realize that they are also susceptible to cognitive biases, and these cognitive biases can be unconsciously transferred to software systems and potentially cause failures. 'Inspired' by humans, artificial intelligence systems can also develop their own biases. Kristian Hammond (2016), a blogger on techcrunch.com, points to five biases that occur in artificial intelligent systems– emergent bias, data drive bias, similarity bias, interaction bias, and bias of conflicting goals. Interestingly, emergent bias can even influence our behavior and increase our tendency towards confirmation bias by suggesting us only those posts on social networks which agree with our opinions. Eventually, this 'bug' seems to be a toll of the system.

We can conclude that research of decision-making processes and cognitive biases overlaps with many areas such as economics, psychology, neuroscience, informatics, health care, management, education, and many others. The interdisciplinary nature of this topic makes the research complex and brings motivating ideas. Moreover, it pushes people from these various areas to talk to each other and share their knowledge and thoughts. We believe that this approach is truly fruitful in objective examination of decision-making processes and cognitive biases.

Conclusion

We hope that our research contributed positively to the overall endeavor to examine cognitive biases and debiasing strategies. Counterfactual priming has confirmed to be the promising debiasing strategy in the case of attribution error. Moreover, counterfactual priming by “self” appears to help in the fight with cognitive biases more than priming by “others”. Consequently, we suggest that a higher number of generated counterfactuals does not prevent people from cognitive biases. Thus, there may be a different ‘mechanism’ that facilitates reduction of cognitive biases in priming by “self”. In addition, open-minded thinking demonstrated its strength in resistance to cognitive biases. Finally, we should learn to listen effectively, seek open and intelligent dialogues, and motivate people to talk, think, and argument wisely. These abilities, as we believe, could help us to nurture our humanity and fight unfounded fear.

"This research was supported by grant APVV-16-0153 – ‘Cognitive failures - individual predictors and intervention possibilities’.

References

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*, 165–178.

Albar, F. & Jetter, A. (2009). Heuristics in decision making. In *PICMET 2009: Technology Management in the Age of Fundamental Change* (pp. 578-584). Portland: Portland International Center for Management of Engineering and Technology.

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine [The behavior of the rational man in the face of risk: Critique of the postulates and axioms of the American school]. *Econometrica*, *21*, 503-546.

Arnott, D. (2002). *A taxonomy of decision biases* (Technical Report. No. 2002/01). Melbourne: Decision Support Systems Laboratory, Monash University. Retrieved from http://dsslab.sims.monash.edu.au/taxonomy.pdf

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

Arkes, H. R. (1991). Costs and benefits of judgement errors: Implications for debiasing. *Psychological Bulletin*, *110*, 486-98

Bačová, V. (2010). Psychológia posudzovania a rozhodovania [Psychology of judgement and decision making]. In V. Bačová (Ed.), *Rozhodovanie a usudzovanie. Pohľady psychológie a ekonómie 1* (pp. 14-47). Bratislava: Ústav experimentálnej psychológie SAV.

Baláž, V. (2009). *Riziko a neistota. Úvod do behaviorálnej ekonómie a financií* [Risk and uncertainty. Introduction to behavioral economics and finance]. Bratislava: VEDA.

Baláž, V. (2010). Prospektová teória a jej miesto v ekonomickom myslení [The prospect theory and its position in economic thinking]. In V. Bačová (Ed.), *Rozhodovanie a usudzovanie. Pohľady psychológie a ekonómie 1* (pp. 88-132). Bratislava: Ústav experimentálnej psychológie SAV.

Barnes, J. & Reinmuth, J. (1976). Comparing imputed and actual utility functions in a competitive bidding setting. *Decision Sciences*, *7*, 801-812.

Baron, J. (1993). Why teach thinking? – An essay. *Applied Psychology*, *42*, 191–237.

Bavoľár, J. (2012). Úvod do psychológie rozhodovania [Introduction to psychology of decision making]. (1st ed.). Košice: Univerzita Pavla Jozefa Šafárika v Košiciach.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293 – 1295.

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk (L. Sommer, Trans.) *Econometrica*, *22*; 23 – 36. (Original work published 1738)

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., & Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron, *30*, 619 – 639.

Buchanan, L. & O'Connell, A. (2006). A brief history of decision making. *Harvard Business Review*. Retrieved from https://hbr.org/2006/01/a-brief-history-of-decision-making

Buckner, R., Andrews-Hanna, J. and Schacter, D. (2008). The Brain's Default Network. *Annals of the New York Academy of Sciences*, *1124*, 1-38.

Camerer, C., Fehr, E., Glimcher, P., & Poldrack, R. (2009). Introduction: A brief history of neuroeconomics. In C. Camerer, E. Fehr, P. Glimcher & R. Poldrack, *Neuroeconomics: Decision making and the brain* (1st ed., 1-12). Academic Press.

Croskerry, P. (2014). ED cognition: any decision by anyone at any time. *Canadian Journal of Emergency Medicine*, *16*, 13-19.

Damasio, A., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, *41*, 81-94.

Dardenne, B. & Leyens, J. (1995). Confirmation bias as a social skill. *Personality and Social Psychology Bulletin*, *21*, 1229-1239.

De Brigard, F., Spreng, N. R., Mitchell, J. and Schacter, D. (2015). Neural activity associated with self, other, and object-based counterfactual thinking. *NeuroImage*, *109*, 12-26.

Dodd, M., Klos, K., Bower, J., Geda, Y., Josephs, K., & Ahlskog, J. (2005). Pathological gambling caused by drugs used to treat Parkinson disease. *Archives of Neurology*, *62*, 1377-1381.

Doll, B., Hutchison, K., & Frank, M. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *Journal of Neuroscience*, *31*, 6188-6198.

Dudeková, K., Kostovičová, L., & Konečný, M. (2017). Redukcia kognitívnych odchýlok u finančných profesionálov prostredníctvom kontrafaktov [Reducing cognitive biases among financial professionals via counterfactuals]. *Proceedings of Kognícia a umelý život XVII*, (46–51). Bratislava, Slovakia.

Fellows, L. & Farah, M. (2007). The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cerebral Cortex*, *17*, 2669-2674.

Ficková, E. (2011). Rodové odlišnosti v kontrafaktoch a múdrosti adolescentov [Gender differences in counterfactuals and wisdom of adolescents]. In I. Ruisel, & Z. Ruiselová, *Múdrosť v kontexte osobnosti: problémy, fakty, otázky* (pp. 109-137). Bratislava: Ústav experimentálnej psychológie SAV.

Frith, U., Morton, J., & Leslie, A. (1991). The cognitive basis of a biological disorder: autism. *Trends in Neurosciences*, *14*, 433-438.

Frost, P., Casey, B., Griffin, K., Raymundo, L., Farrell, C., & Carrigan, R. (2015). The influence of confirmation bias on memory and source monitoring. *The Journal of General Psychology*, *142*, 238-252.

Fujino, J., Fujimoto, S., Kodaka, F., Camerer, C., Kawada, R., Tsurumi, K., Tei, S., Isobe, M., Miyata, J., Sugihara, G., Yamada, M., Fukuyama, H., Murai, T., & Takahashi, H. (2016). Neural mechanisms and personality correlates of the sunk cost effect. *Scientific Reports*, *6*.

Galinsky, A., & Moskowitz, G. (2000). Counterfactuals as Behavioral Primes: Priming the Simulation Heuristic and Consideration of Alternatives. *Journal Of Experimental Social Psychology*, *36*, 384-409.

Geier, C., Terwilliger, R., Teslovich, T., Velanova, K., & Luna, B. (2009). Immaturities in reward processing and its influence on inhibitory control in adolescence. *Cerebral Cortex*, *20*, 1613-1629.

Gigerenzer, G. & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Goldstein, W. M. & Hogarth, R. M. (1997). Judgment and decision research: Some historical context. In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on judgment and decision making* (3–65). New York: Cambridge University Press.

Gonzalez, C. (2014). Decision-making: a cognitive science perspective. In S. Chipman, *The Oxford Handbook of Cognitive Science* (1st ed.).

Gupta, R., Koscik, T., Bechara, A., & Tranel, D. (2011). The amygdala and decision-making. *Neuropsychologia*, *49*, 760-766.

Halter, A. N. & Dean, G. W. (1971). Decisions under uncertainty. Cincinnati: South Western Publishing Co.

Hammond, K. (2016). *5 unexpected sources of bias in artificial intelligence* [Blog post]. TechCrunch. Retrieved from: https://techcrunch.com/2016/12/10/5-unexpected-sources-of-bias-in-artificial-intelligence/

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making*, *8*, 188–201.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons.

Heukelom, F. (2007). *Kahneman and Tversky and the origin of behavioral economics* (Tinbergen Institute Discussion Paper No. 07-003/1). Retrieved from University of Amsterdam, Tinbergen Institute.

Kahneman, D. (2012). *Thinking, fast and slow*. New York: Farrar, Straus, and Giroux.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, (201–208). New York: Cambridge Univ. Press.

Kray, L., & Galinsky, A. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational behavior and human decision processes*, *91*, 69-81.

Koenigs, M. & Grafman, J. (2009). The functional neuroanatomy of depression: Distinct roles for ventromedial and dorsolateral prefrontal cortex. *Behavioural Brain Research*, *201*, 239-243.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U. & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*, 673 – 676.

List of cognitive biases. (2017). In *Wikipedia*. Retrieved December 5, 2017, from https://en .wikipedia.org/wiki/List_of_cognitive_biases

Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098-2109.

Macmillan, M. (2002). An odd kind of fame: stories of phineas gage. Cambridge, MA: MIT Press.

Mahoney, M. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, *1*, pp.161-175.

Martínez-Selva, J. M., Sánchez-Navarro, J. P., Bechara, A., & Román F. (2006). Brain mechanisms involved in decision making. *Revista de neurologia*, *42*, 411-418

Mason, M. & Morris, M. (2010). Culture, attribution and automaticity: a social cognitive neuroscience view. *Social Cognitive and Affective Neuroscience*, *5*, 292-306.

Miller, F., Smith, E., & Uleman, J. (1981). Measurement and interpretation of situational and dispositional attributions. *Journal of Experimental Social Psychology*, *17*, 80-95.

Moran, J., Jolly, E., & Mitchell, J. (2014). Spontaneous mentalizing predicts the fundamental attribution error. *Journal of Cognitive Neuroscience*, *26*, 569-576.

OECD (2015), *Education at a Glance 2015: OECD Indicators*, OECD Publishing, Paris.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R., & Frith, C. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*, 1042-1045.

Rahman, S., Robbins, T., Hodges, J., Mehta, M., Nestor, P., Clark, L., & Sahakian, B. (2005). Methylphenidate ('Ritalin') can ameliorate abnormal risk-taking behavior in the frontal variant of frontotemporal dementia. *Neuropsychopharmacology*, *31*, 651-658.

Remus, W.E. & Kottemann, J.E. (1986). Toward intelligent decision support systems: an artificially intelligent statistician. *MIS Quarterly*, *10*, 403–418.

Roberto, M. (2002). Lessons from Everest: The Interaction of Cognitive Bias, Psychological Safety, and System Complexity. *California Management Review*, *45*, 136-158.

Rogers, R., Everitt, B., Baldacchino, A., Blackshaw, A., Swainson, R., Wynne, K., Baker, N., Hunter, J., Carthy, T., Booker, E., London, M., Deaking, J., Sahakian, B., & Robbins, T. (1999b). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: Evidence for monoaminergic mechanisms. *Neuropsychopharmacology*, *20*, 322-339.

Ross, L. (1977). The Intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in experimental social psychology*, *10*, 173-220. 10.1016/S0065-2601(08)60357-3.

Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *91*, 497–510

Schacter, D. & Addis, D. (2007). The optimistic brain. *Nature Neuroscience*, *10*, 1345-1347.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*, 1755-1758.

Sapolsky, R. (2014). Lecture 18: Agression II. [Video lecture]. Retrieved from Standford University Human Behavioral Biology https://www.youtube.com/watch?v=lX74CfRAZmo

Shakespeare, W. & Timmins, S. (1860). Hamlet. London: S. Low, Son, and Co.

Simon, H. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99.

Skořepa, M. (2011). *Decision making: a behavioral economic approach* (1st ed.). Basingstoke: Palgrave Macmillan.

Snyder, M. and Swann, W. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *36*, 1202-1212.

Stanovich, K. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.

Staw, B. (1976). Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance*, *16*, 27-44.

Staw, B. (1981). The escalation of commitment to a course of action. *Academy of Management Review*, *6*, 577-587.

Staw, B. and Fox, F. (1977). Escalation: The Determinants of Commitment to a Chosen Course of Action. *Human Relations*, *30*, 431-450.

Strachanová, D. (2015). *Kontrafaktuálne myslenie a utopené náklady v rizikovom rozhodovaní* [Counterfactual thinking and sunk cost in risky decision-making.]. (Bachelor Thesis). FSEV, Univerzita Komenského: Bratislava.

Strachanová, D. (2017). *Čo by bolo keby? Redukcia kognitívnych odchýlok prostredníctvom kontrafaktov* [What would be if? The debiasing effect of counterfactuals]. (Master Thesis). FSEV, Univerzita Komenského: Bratislava.

Svedholm-Häkkinen, A., & Lindeman, M. (2017). Actively open-minded thinking: development of a shortened scale and disentangling attitudes towards knowledge and people. *Thinking & Reasoning*, *24*, 21-40.

Šinský, M. (2010). Taxonómia sklonov k chybám [Taxonomy of cognitive biases]. In V. Bačová (Ed.), *Rozhodovanie a usudzovanie. Pohľady psychológie a ekonómie 1* (pp. 162-204). Bratislava: Ústav experimentálnej psychológie SAV.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124-1131.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293-315.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297-323.

Neumann, J. v., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

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

Wilke, A. and Mata, R. (2012). Cognitive Bias. In V.S. Ramachandran (ed.) *The Encyclopedia of Human Behavior*, *1*, 531-535. Academic Press.