The Impact of Authority Figures in our Beliefs: Experimental Studies on the "Author´s Bias"

ABSTRACT: We've explored how our judgements tend to be biased by the opinions of prestigious figures. In the first study, subjects were randomly assigned to two groups, and they were asked to put a score of agreement (1-10) to several statements. In one group ("anonymous condition") subjects didn't know the author of the statement, while in the other group ("author condition") the statement was attributed to a prestigious person. We've found that the statements attributed to prestigious persons received higher scores than the anonymous statements. In the second study, a similar procedure was employed, and we've found that subjects assigned lower scores to the statements that were attributed to politicians that they were unwilling to vote. In the third study, we've explored people's opinions about two controversial events in Argentina that were interpreted in different ways by politicians of the governing party and of the opposition party, and we've found that most subjects agreed with the views adopted by their preferred party, even if they had little to do with politics. In the fourth study, subjects were randomly assigned to two groups (one group received previous information about the bias, while the other group didn't receive that information), and we've found that when people were aware of this "author's bias", its influence decreased. The implications of this bias in relation to people's judgements about justice will be analysed.

INTRODUCTION

Since its origins, philosophy has encouraged us to evaluate every thought for its intrinsic value, and not merely because it is supported by figures of authority, or by many people. The history of culture can be seen as a process in which the authority principle (religious, political, etc.) was gradually replaced by the free examination of ideas.

The philosophical thinking about this topic has a long history. Aristotle has discused the different forms of arguments in three of his books (*Topics, On Sophistical Refutations* and *Rhetoric*). Aristotle's *Rhetoric* described three modes of persuasion: ethos, pathos, and logos. Ethos is an appeal to the authority or credibility of the speaker, pathos is an appeal to the audience's emotions, and logos is an appeal to reason. Locke, in his *Essays Concerning Human Understanding*, proposed a distinction between four sorts of arguments that men ordinarily use: argumentum ad verecundiam, ad ignorantiam, ad hominem, and ad judicium. Hume, in his *An Enquiry Concerning Human Understanding*, suggested that "No man can safely be trusted with unlimited authority", and has contributed with a famous argument against the credibility of miracles. More recently, Sagan has written that "One of the great commandments of science is 'Mistrust arguments from authority.' (...) Too many of such arguments have proved too painfully to be wrong. Authorities must prove their contentions like everybody else" (Sagan, 1996, p. 31).

Traditional philosophical reflection on the topic was mainly based on the analysis of arguments, but empirical research is also relevant to answer some questions related to this topic (e.g., why do speakers commit fallacious appeals to authority, why are listeners persuaded by irrelevant appeals to authority, how education might change those propensities). Several philosophers have argued about the relevance of integrating empirical and philosophical methods (Mortensen & Nagel, 2014). The philosophical research programs that are characterized by this integration of methods can be included in the category of "experimental philosophy", if we understand this concept in a broad sense (e.g., Schwitzgebel, 2012; Rose & Danks, 2013; Sytsma & Livengood, 2015).

There are several areas of empirical research that can be considered relevant to this topic, including the literature about source credibility (Pornpitakpan, 2004), heuristics and biases (Tversky & Kahneman, 1975; Gilovich et al, 2002; Gigerenzer & Selten, 2002),

critical thinking (Pithers & Soden, 2000; Lai, 2011) and attitude change (Crano & Prislin, 2011). Several models of persuasion (e.g., Petty & Cacioppo, 1986; Chaiken et al, 1989) are based on dual-processing theories of cognition. These theories (e.g., Stanovich & West, 2002; Kahneman, 2011) propose two fundamentally different modes of cognition: Type 1 (fast, effortless, experiential, intuitive, reflexive processes) and Type 2 (slow, effortful, controlled, analytic, rule-based processes, that require high cognitive effort and systematic reasoning).

These kinds of biases in the processes of speakers and listeners might be related to both phylogenetic (evolutionary) and ontogenetic (developmental) causes. In their theory of cultural evolution, Richerson and Boyd (1985) proposed a number of possible transmission biases that take place when people are fascinated by high status people and try to adopt their behaviors (for example, “lifestyles of the rich and famous”). The kind of features that we consider "prestigious" is determined by culture, social class, and group membership, but that doesn't mean that it is arbitrary. The list of biases has been refined over the years (e.g., Henrich & McElreath, 2003). Social learning sometimes involves copying behaviors from a model (someone observed behaving in certain way), and it has several potential biases, including success bias (copying from those who are perceived to be better off), status bias (copying from those with higher status), homophily (copying from those most like ourselves), and conformist bias (copying behaviors that more people are performing). A study about the prestige-biased cultural learning showed that children of 3- and 4-year-old learned more from an adult model to whom bystanders had previously preferentially attended for 10 seconds (Chudek et al., 2012).

Prestige figures are familiar to us, and previous research has shown that familiarity influences our judgements and choices. The faces that we find most attractive are those that are more familiar (Moreland & Zajonc, 1982). The tendency to be influenced by familiarity might have been selected in the history of the species, because for our ancestors, familiarity was evidence of safety.

Another bias related to prestige is the halo effect, where our overall impression of a person influences how we feel about his/her character and what he/she says. (Thorndike, 1920). We think "He is nice!", and this thought impacts in our assessment of other traits

("He is also smart!"). One example of the halo effect is our overall impression of celebrities.

Some studies have explored specific instances of the human tendency to trust authority, and the conditions that promote a more autonomous reflection. Milgram's study found that 65% of the participants obeyed an authority figure, even when they thought that obedience caused serious injury and distress to a victim (Milgram, 1963). Other studies have explored the effects of arguments that intend to weaken the perceived authority of a source (van Eemeren et al., 2009; Oaksford and Hahn, 2013).

The influence of authority has also been explored in academic contexts. Both students and researchers rate arguments as being more persuasive when they are associated with an expert mathematician than when the author is anonymous, but this effect only occurs when the argument is uncertain (Inglis and Mejía-Ramos, 2009). Nonexperts persons judge abstracts as being of higher quality when they include meaningless mathematical equations, but this "nonsense math effect" was not found in participants with degrees in mathematics, science, technology or medicine (Eriksson, 2012). Weisberg et al. (2008, 2015) found that neuroscience information is intuitively compelling, even when it is irrelevant. Neuroscience experts were not seduced by irrelevant information that came from their own domain of expertise, which suggests that increased education can be an effective antidote for the bias.

Dan Sperber has proposed the "guru effect model", which refers to people's tendency to judge profound what they fail to grasp, just because it is professed by a prestigious figure or “guru” (Sperber, 2010). He suggested that when people of no particular authority express their thoughts in an obscure manner, we often think that their words don't worth much effort of interpretation, but when we trust in someone's authority, we may be affected by the “confirmation bias”, paying more attention to confirming than to disconfirming evidence. If the text is obscure, the more open to a variety of construals, the greater risk of “confirmation bias”. According to Sperber, the guru effect has a social benefit: belonging, being capable of appreciating "the importance of a difficult great thinker", who is overestimated not in spite but because his oscurity.

Very few papers have explored the "guru effect model" so far. The study of Pennycook et al. (2015) is not strictly an exploration of the guru effect, but of "bullshit",

which was defined by philosopher Harry Frankfurt as something that is designed to impress, without any concern for the truth. Anyway, there's a link between both topics, since the guru effect might be one of the factors that increase people's receptivity to bullshit. Pennycook et al. (2015) created a Bullshit Receptivity Scale using quotes from Deepak Chopra, and studied the individual differences in the receptivity of pseudo-profound statements in relation with other dispositions. They proposed two mechanisms that might explain people's receptivity to bullshit. The first one is a Type 1 process that consists in a default assumption of the statement's truth. According to Gilbert's model (1991), humans must believe something in order to comprehend it, so there's a general bias towards accepting statements as true. The second mechanism is a Type 2 process of deliberative reasoning to assess the truth or meaningfulness of the statement. Some individuals may have a stronger tendency to believe in statements that seem profound regardless of their content, and a weaker tendency to engage in the evaluative process. Also, some situational variables (e.g., the prestige of the "bullshitter") might increase or decrease those individual differences.

Martin et al. (2016) have tested the prediction (derived from the guru effect model) that "source credibility can cause attitudinal change by biasing the interpretation of pragmatically ambiguous material". They assigned participants to a condition in the 2 (Source credibility: high versus low) x 2 (Involvement: high versus low) x 2 (Passage version: original versus simplified) study design. Participants in the high involvement condition were told that they would take a quiz at the end of the study and that high scorers would receive bonuses. This instruction was intended to motivate them to increase their effort. Participants in the high credibility condition were told that the passages came from "the award winning book *Modernity*, written by the world-renowned philosopher and bestselling author Alex Wells". Participants in the low credibility condition were told that the passages came from "the essay *Modernity* written by college student Alex Wells for an introductory philosophy class". In the study, the prediction was not confirmed (text difficulty was not found to have a significant effect on attitudes). The authors proposed several hypotheses that might explain these results without rejecting the guru effect model (e.g., the difference between the passages might have been very small, source credibility might require a more active involvement with a social group). We think that this study is

valuable as a first step in the empirical exploration of the guru effect model, but future studies should explore a wider spectrum of levels of text ambiguity and kinds of social influence. When a group has a high reverence towards a "guru", its members will usually reinforce positive attitudes and punish negative attitudes towards the prestigious figure. These social contingencies might be sometimes consciously delivered, but they're often the unplanned effects of other dispositions. The condition used in this study (informing that the author is a "world-renowned philosopher") doesn't seem to be enough to explore the kinds of social contingencies that we've just described.

In the sequence of studies of this paper, we will offer an empirical exploration of how our ideas, political preferences and judgements of behaviors tend to be biased by the opinions of prestigious figures. This research is part of a broader project that tries to integrate philosophical and empirical approaches related to the topic of fallacies and persuasion. We will explore a positive author's bias (study 1), a negative author's bias (study 2), a correlation between belief and political affiliation (study 3), and the possibility of decreasing the author bias with a specific instruction (study 4).

STUDY 1 - MATERIALS AND METHOD

In this study, subjects (adult men and women) were randomly assigned to two groups, and they were asked to rate their agreement (1-10) with several statements. In one group ("anonymous condition") subjects didn't know the author of the statement, while in another group ("author condition") the statement was attributed to a prestigious person.

The quotes that were used in this study are the following ones:

Q1: "We win justice quickest by rendering justice to the other party." (Mahatma Gandhi)

Q2: "Knowledge without justice ought to be called cunning rather than wisdom." (Plato)

Q3: "Life is like riding a bicycle. To keep your balance, you must keep moving." (Albert Einstein)

Q4: "Failure does not necessarily show incompetence. Oedipus had value, was intelligent and persevering but was a victim of circumstances." (intentionally misattributed to the argentine writer Jorge Luis Borges)

Q5: "What is the truth? Tough question, but I have solved it as far as I'm concerned saying that's what your inner voice tells you." (Mahatma Gandhi)

A first group of 140 subjects (70 men, 70 women) were randomly assigned to the anonymous and author conditions, and were asked to rank their agreement with the quotes 1 and 2.

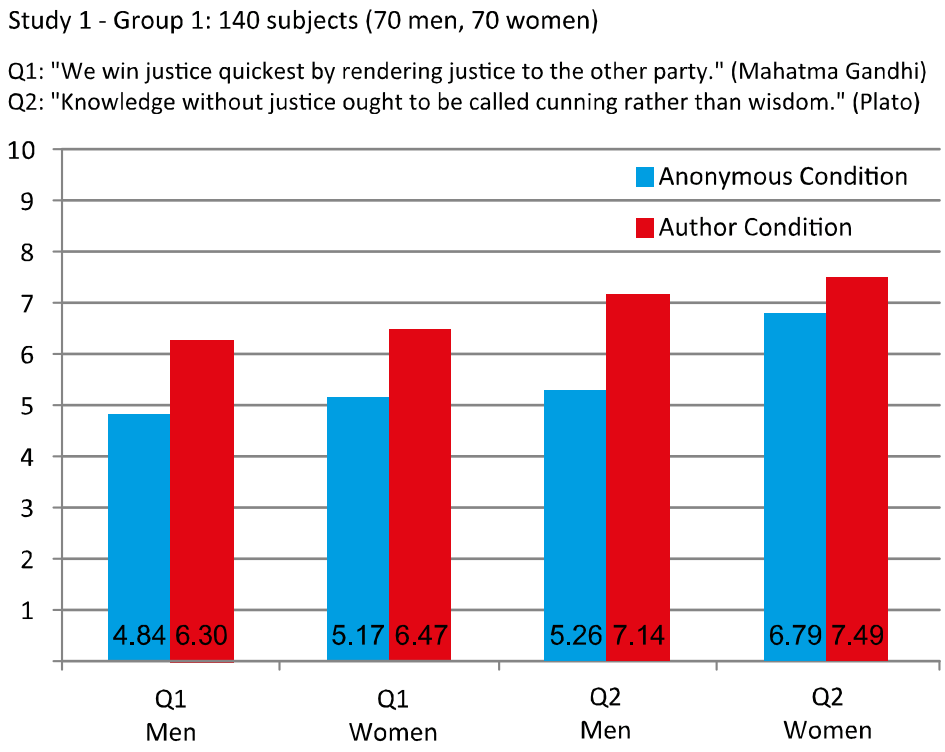
A second group of 100 subjects were randomly assigned to the anonymous and author conditions, and were asked to rank their agreement with the quotes 3 to 5.

RESULTS

STUDY 1 - Group 1:

A one-way ANOVA was conducted to compare the effects of the conditions (anonymous, author) on the scores assigned to the quotes (Q1, Q2). Significative differences were found. Subjects assigned higher scores in the author condition than in the anonymous condition, when they assessed the quotes Q1 ($F(1,278) = 16{,}023$, $p < .001$) and Q2 ($F(1,128) = 13{,}569$, $p < .001$).

The results of this study are shown in the following graphic.



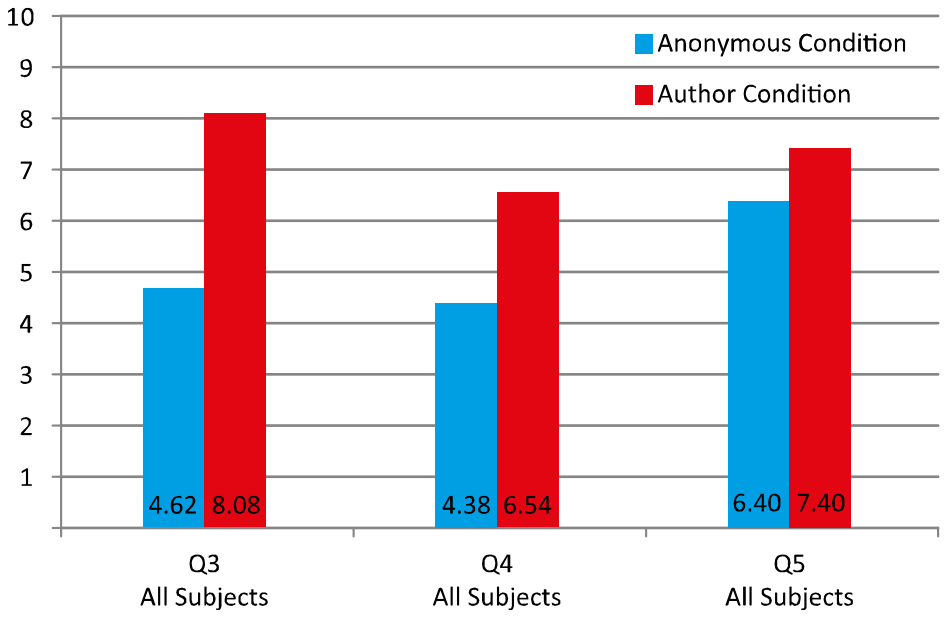
STUDY 1 - Group 2:

A one-way ANOVA was conducted to compare the effects of the conditions (anonymous, author) on the scores assigned to the quotes (Q3, Q4, Q5). Significative differences were found. Subjects assigned higher scores in the author condition than in the anonymous condition, when they assessed the quotes Q3 ($F(1,98)=158.150$, $p< .001$), Q4 ($F(1,98)=105.644$, $p< .001$) and Q5 ($F(1.98)=27.841$, $p< .001$).

The results of this study are shown in the following graphic.

Study 1 - Group 2: 100 subjects

Q3: "Life is like riding a bicycle. To keep your balance, you must keep moving." (Albert Einstein)
Q4: "Failure does not necessarily show incompetence. Oedipus had value, was intelligent and persevering but was a victim of circumstances." (intentionally misattributed to Jorge Luis Borges)
Q5: "What is the truth? Tough question, but I have solved it as far as I'm concerned saying that's what your inner voice tells you." (Mahatma Gandhi)



ANALYSIS

Study 1 suggests that there's a positive "author's bias": a tendency to overestimate some ideas when they are signed by a valued author. This phenomenon might be involved in the persuasive effects of the ad verecundiam fallacy, an argument that claims that a position is true because an authority suggests so.

The strongest effect was the one of Einstein's signature, whose average score was almost the double in the author condition compared with the anonymous one.

Women scored higher than men in every sentence, but the author's bias was found in both genders.

STUDY 2 - MATERIALS AND METHOD

Study 1 suggests that we overestimate an idea when it comes from a valued author. We wondered if the reverse is also true: do we underestimate the quotes when we don't value their authors?

In order to explore this issue, 86 subjects were randomly assigned to two groups (anonymous versus author condition), and they were asked to rate their agreement (1-10) with the following statements:

Q6: "If you don't have a friend or a relative, you don't enter even disguised as a monkey to the Judicial System." (Cristina Fernández de Kirchner)

Q7: "In Argentina we should remember and take as an example those who are able to give everything for the values they have." (Cristina Fernández de Kirchner)

Both sentences were sent to Twitter on April 9th and March 24th of 2013 by Cristina Fernández de Kirchner, President of Argentina when this study was done.

After rating both quotes, we explored the political preferences of the participants, asking them for what party would they vote in the next parliamentary elections. From the 86 subjects, 72 said that they wouldn't vote for the official party. Only the data of this group (N=72) was analysed.

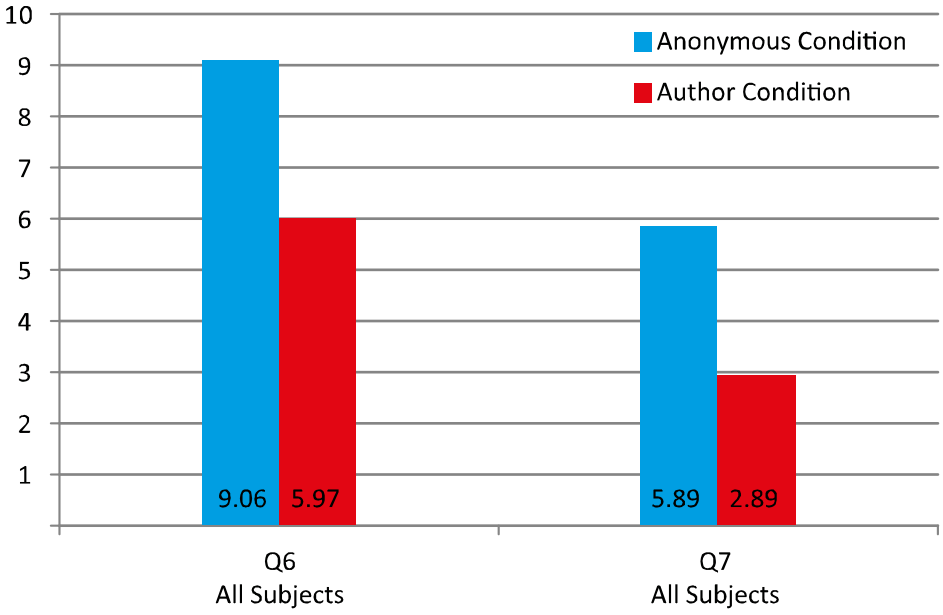
RESULTS

A one-way ANOVA was conducted to compare the effects of the conditions (anonymous, author) on the scores assigned to the quotes (Q6, Q7). Significative differences were found. Subjects assigned lower scores in the author condition than in the anonymous condition, when they assessed the quotes Q6 ($F(1,70)= 100.779$, $p< .001$) and Q7 ($F(1,70)= 102.692$, $p< .001$).

The results of this study are shown in the following graphic.

Study 2 - 72 subjects (who won't vote for the governing party)

Q6: "If you don´t have a friend or a relative, you don´t enter even disguised as a monkey to the Judicial System." (Cristina Fernández de Kirchner, governing party)
Q7: "In Argentina we should remember and take as an example those who are able to give everything for the values they have." (Cristina Fernández de Kirchner, governing party)



ANALYSIS

Study 1 suggests that there's a negative "author's bias": a tendency to underestimate some ideas when they are signed by a disvalued author. This phenomenon might be involved in the persuasive effects of the ad hominem fallacy, in which an argument is rejected by attacking the person associated with it, rather than the substance of the argument itself.

The bias was even stronger than in study 1, perhaps because it involves political attitudes that elicit a stronger emotional reaction.

We can now define the "author's bias" as a distortion that consists in attributing more or less value to an idea, depending on the person who generated it. It's a tendency to overestimate statements that are signed by a valued author, and to underestimate the ones that are signed by a disvalued author.

Kahneman and Tversky (1974) defined a "cognitive bias" as a distortion that concentrates on small and irrelevant fragments of information. We have limited information and limited time, so we must use heuristics (shortcuts that not always are accurate). The heuristic behind the author's bias might be expressed with the following statements: "If the author is valued (by me or by my group), the statement must be true", and "If the author is not valued (by me or by my group), the statement must be false".

STUDY 3 - MATERIALS AND METHOD

When we interpret political and judicial events, is our judgment biased depending on the perspective adopted by the politicians that we appreciate or reject?

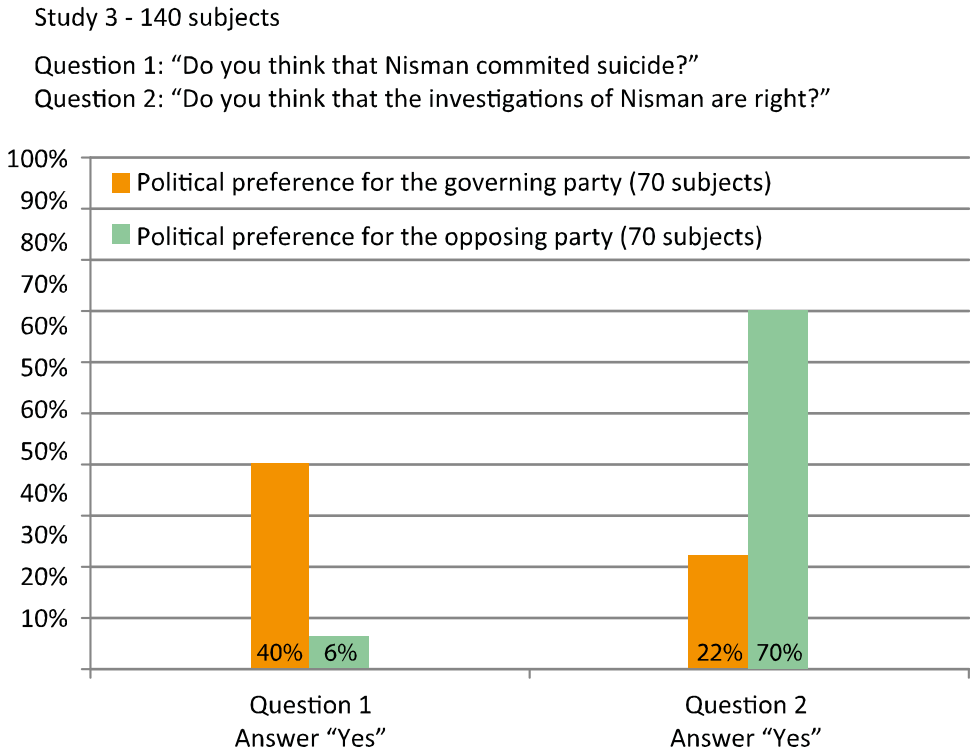
In order to explore this issue, we developed a new study. Days after the death of prosecutor Alberto Nisman, who was investigating Cristina Fernández de Kirchner, the President of Argentina at that moment (January, 2015), we wondered it there could be a bias in the analysis of the causes that led to his death. Four days before Nisman died, he had accused President Cristina Kirchner to cover-up, obstruct justice and protect the perpetrators of the bombing of the Jewish center AMIA (Argentina Israelite Mutual Association) in 1994. The attack, which left eithty-five people dead and hundreds injured, was the worst in the country's recent history. Nisman investigated the case for more than ten years, and believed that the Iranian government and Hezbollah were behind it.

This research took place the first days after Nisman died, when the President spoke in favor of the possibility that he had commited suicide. However, it was suspected that he was murdered.

In the study we worked with 140 adults, of which 70 had a political preference for the governing party, and 70 had a political preference for the opposing party. We asked them the following questions: (1) "Do you think that the prosecutor Nisman committed suicide?", (2) Do you think that the investigations of Nisman are right? and (3) Which candidate will you vote in the next election?

RESULTS

The results are shown in the following graphic.



ANALYSIS

This study is correlational, but it allows us to appreciate some differences in people's judgements about an event that had significative political repercussions in Argentina. Voting intention was used as a proxy for the participant's political preferences. Future studies might include other measures of political values and beliefs.

We think it's interesting to note that the answer to the first question ("Did he commit suicide or was the murdered?") correlates with political preferences, even when it's a statement of fact (and not of value). We can ask whether there are, and whether there should be, differences in the heuristics used to judge statements of fact versus statements of value. This broad question should be explored in future research.

STUDY 4 - MATERIALS AND METHOD

Can the "author's bias" be reduced if people are aware about its existence? We explored this possibility in a new study in which we assigned participants (100 adults) to a condition in the 2 (author versus anonymous) x 2 (information versus no-information) study design. Participants were asked to rate their agreement, in a scale from 1 to 10, with the following statements:

Q8: "The secret of human existence is not only to live, but to know what to live for." (Fyodor Dostoyevsky)

Q9: "When a man and a woman marry their novel finishes and their story starts." (Oscar Wilde)

In the author condition, participants knew the author of the quote. In the information condition, participants were told "We have done several studies that suggest that we tend to overestimate the value of ideas when they belong to prestigious figures. Try to estimate how much do you agree with these quotes, regardless the author to whom they're attributed".

After rating the quotes, the participants assigned to the no-information condition were also asked to rank, in a scale from 1 to 10, how much did they think that knowing the author of a quote could influence the score that participants gave to that quote. This question intended to explore if participants in the no-information condition were aware of this bias.

RESULTS

A 2x2 factorial ANOVA was conducted to compare the effects of the conditions (author/anonymous) x (information/no-information) on the scores assigned to the quotes (Q8, Q9). Significative differences were found in the information condition, but not in the no-information condition. In the information condition, subjects assigned higher scores to

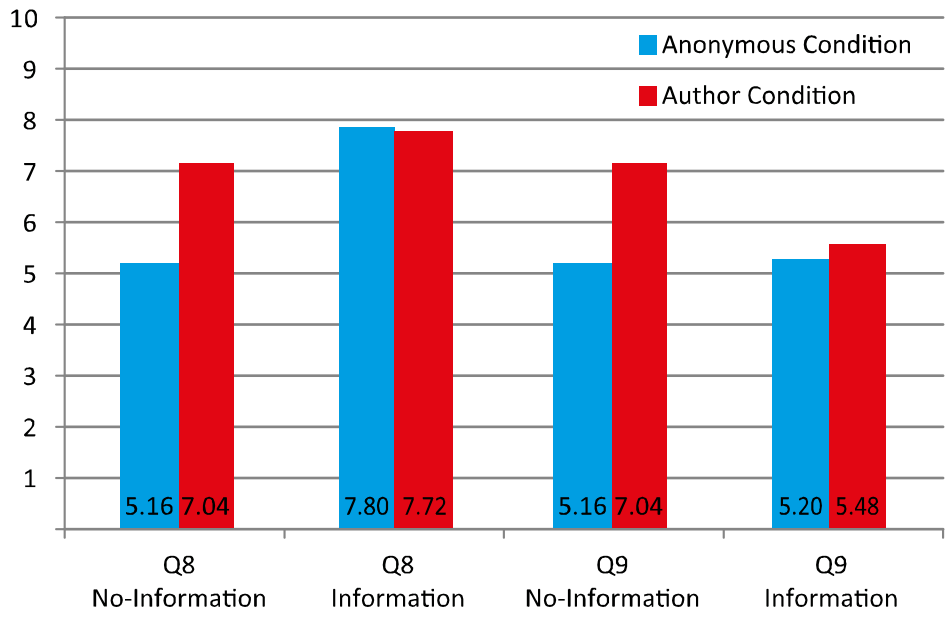
the same quote in the author condition than in the anonymous condition, for the quote Q8 ($F(1,48)=26.40$, $p< .001$) and Q9 ($F(1,48)=48.143$, $p< .001$). In the no-information condition, there were no significative differences between the scores in the author condition and the anonymous condition, for the quotes Q8 ($F(1,48)=0.072$, $p< .789$) and Q9 ($F(1,48)=0.601$, $p= .442$).

Participants answered the question about "how much did they think that knowing the author of a quote could influence the score" with an average of 3,48 in the author condition and 2,78 in the anonymous condition.

The results of this study are shown in the following graphic.

Study 4 - 100 subjects

Q8: "The secret of human existence is not only to live, but to know what to live for." (Fyodor Dostoyevsky)
Q9: "When a man and a woman marry their novel finishes and their story starts." (Oscar Wilde)



ANALYSIS

Results of Study 4 suggest that participants had a very low awareness of this kind of bias, and that information about the author's bias might reduce its effects (at least in some occasions). We think that the information about the bias might have been effective to

increase Type 2 processes and diminish the influence of Type 1 processes, but this conjecture should be explored in future research.

CONCLUSIONS

This sequence of studies has shown some evidence related to the "author's bias", a tendency to overestimate statements when they are signed by an author that is valued by the audience, and to underestimate statements when they are signed by an author that is disvalued by the audience.

This bias might be related to many other phenomena, such as problems in communication between different groups, the tendency to adopt ideas of prestigious persons, the violence elicited by political disagreement, people's interest in prestigious figure's opinions about topics in which they're not experts, the emergence of apocryphal quotes and forgeries, and the difficulty of assessing ideas by their intrinsic value.

Apocryphal quotes are very frequent in the web. For example, the ones attributed to Einstein, like "Insanity is doing the same thing over and over again and expecting different results". This might be a good idea, but Einstein never said it. The same quote appears signed by Benjamin Franklin and by Mark Silber. The attribution probably has a persuasive effect, even if there is no logical connection between an idea and the person who suggests it. A good anonymous quote is like a virtuous violinist playing in the subway: it has no framework, and therefore, it's less valued.

The appraisal we assign to a signature and to the name of an author is also related to forgeries. Michelangelo sold to Pope Julius II several sculptures telling him they were Greek, when he himself created them (Vasari, 1991). One of the most famous examples of forgery is the one of Han van Meegeren, who falsified some paintings of Johannes Vermeer, and signed his own paintings as if they belonged to Vermeer. He was sentenced to death for selling paitings of Vermeer to a nazi, but showed evidence that he himself had falsified them, and then was sentenced to prison, dying from a heart attack before concluding his sentence.

Frameworks not only give value, they can also provide emotional content. This may have positive effects, for example, when we are in front of an original painting that we like, and we are moved when we know that it belongs to our favourite painter. But this effect in other cases can bias our thought and lead us to bad decisions, for example, if we agree with political ideas and actions only because our favorite politician has suggested them.

Which mechanisms can explain this bias and which variables make an author more or less trustworthy for a specific audience? Future studies could explore them: e.g., whether the author is judged to be competent in an area, whether he is respected by other members of the community, whether his political attitudes are liked or disliked, etc.

We suggested that the mechanism might be explained with dual-processing theories, and this is another question that remains to be explored. We might distinguish between Type 1 processes (e.g., the audience might be guided by emotional processes or habit, without any thoughtful analysis of the source reliability) and Type 2 processes (e.g., the audience might think "this claim is probably true, because its source is reliable"). For example, when we listen to the name of a person that we don't like, emotion may blind us, and we cannot objectively assess what he says. In this example, a Type 1 process (emotion) inhibits a Type 2 process (objective analysis). This kind of emotional reaction is in conflict with the ideal of rationality, which asserts that being rational is not only being able to argue but to be persuaded by another person, whoever he is, when he gives good reasons.

Anyway, it's important to remember that heuristics used to assess source reliability are not necessarily irrational. During the last decades, argumentation research and expertise studies have explored the problem of how non-expert persons can figure out which statements from purported experts deserve their trust (Goldman, 2001; Goodwin, 2011). Walton (1997) has proposed six general types of critical questions that are useful to evaluate appeals to authority: Expertise Question (How credible is E as an expert source?), Field Question (Is E an expert in the field that A is in?), Opinion Question (What did E assert that implies A?), Trustworthiness Question (Is E personally reliable as a source?), Consistency Question (Is A consistent with what other experts assert?), Backup Evidence Question (Is A’s assertion based on evidence?).

Type 1 and Type 2 processes involved in the author's bias require an explanation that includes both phylogenetic (evolutionary) and ontogenetic (developmental) aspects.

Humans share with other primates several features, including the susceptibility to hierarchies. In nonhuman animals, dominance is usually expressed in the form of agression (power to coerce), while in humans, status often comes from non-agonistic sources, in particular, from excellence in valued domains of activity (power to persuade). The capacity to produce and evaluate arguments (Type 2 processes) allows us to monitor the risks of deception and manipulation involved in accepting the authority of communicators. There might be individual differences in the tendency to respect (or to defy) authority, and those initial differences might be shaped by the social and personal contingencies that affect each individual's learning and development.

Human groups have the ability to socially interconnect and learn from one another over generations. Social learning is guided by different cues, such as prestige, success, sex, dialect, and ethnicity (Sperber, 1985, 1996). Prestige-related heuristics have probably evolved to improve the quality of information acquired via cultural transmission. As we don't have infinite time nor access to all the information, to trust prestigious people seems to be a good heuristic. Because figuring out which ideas and behaviors are good is costly and difficult, selection favored a copying bias, which makes prestigious individuals more influential. Being in touch with them seems to be a proxy for having good ideas.

There's a preference for models who seem "popular" (Henrich & Gil-White, 2001). But there's a high flexibility in the selection of which behaviors or actions lead to the attribution of "high prestige". People’s preferences and motivations are not fixed as innate dispositions, they can be modified by social learning (Bandura, 1977).

Status or prestige also informs us about a person’s value as an ally. It enables more and broader alliances (social capital). Humans live embedded in multiple social hierarchies—at root, dominance relations—which must be weighted and added to appraise an individual’s value in alliances (Diamond, 2016).

Prestige-related heuristics are often beneficial as a guide of social learning, but they also have several downsides and risks:

(1) Ideas, articles for journals or candidates for a job are not necessarily judged for their own merits.

(2) Advantages are justified appealing to “quality”, but often are due to the framework (the author, an elite institution).

(3) Prestige is not necessarily evidence of any other positive trait.

(4) Gender and ethnic biases (e.g., women are less quoted in top journals ; Healy, 2013). Conference abstracts were assessed as having a greater scientific quality when they were signed by male authors, specially if the topic was male-typical (Knobloch et al, 2013).

(5) Positive changes for society are ignored if they are not proposed by prestigious figures.

(6) Status is a positional good—a zero-sum game—what is won by some is lost by others. (Diamond, 2016)

(7) Often we choose the prestigious to give us the answers we want to hear.

(8) We overestimate the value of certain ideas, books and "authors" because we relate them to prestigious "goods" (an institution, advertising, etc).

(9) We overestimate minor works of (good and bad) prestigious authors.

(10) Successful individuals are influential even beyond their domain of expertise (e.g., people might believe in a pseudoscience just because a prestigious person says that it works).

(11) We don't focus on outcomes. As Hanson (2016) points out: "We should help people to doubt and distrust the prestigious, so they can be more open to focus on outcomes".

Future research might explore: (1) which mechanisms are involved in the author's bias, (2) how can we distinguish between beneficial and harmful instances of reliance in a source, (3) how can we avoid the harmful instances of the author's bias.

Part of the answer to point 3 might come from changes in Type 1 and Type 2 processes, for example: realizing that we are vulnerable to the bias, remembering that people we value are fallible and people we don't like can say something true, assessing different valuable features that exceed prestige, imagining that an idea of our favourite politician is said by a politician from another party, asking the opinion of a devil's advocate, and omitting prestige-related information from job applications (De Cruz, 2016).

We said in the beginning that the history of culture can be seen as a process in which the principle of authority was gradually replaced by the free examination of ideas. It is the project of Socrates, the one of modernity, the one that invites us not to base our ideas merely on tradition, on what the majority or an authority says.

REFERENCES

Albarracin, D., Johnson, B. T., & Zanna, M. P. (Eds.). (2014). The Handbook of Attitudes. Psychology Press.

Bandura, Albert (1977). Social Learning Theory. Oxford, England: Prentice-Hall.

Bloom, P. (2010). How Pleasure Works: The new science of why we like what we like. Random House.

Chaiken, S., A. Liberman, and A. Eagly. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In Unintended thought, eds. J.S. Uleman, and J.A. Bargh, 212–252. New York: Guilford Press.

Chudek, M. H. S., Birch, S. and J. Henrich (2012) Prestige-Biased Cultural Learning: Bystander's Differential Attention to Potential Models Influences Children's Learning. Evolution and Human Behavior, 33, 46-56.

Comesaña, J. M. (2001). Lógica informal: falacias y argumentos filosóficos. Eudeba.

Correia, V. (2011). Biases and fallacies: The role of motivated irrationality in fallacious reasoning. Cogency: Journal of reasoning and argumentation, 3(1), 107-126.

Crano, W. D., & Prislin, R. (Eds.). (2011). Attitudes and attitude change. Psychology Press.

De Cruz, Helen. Powerpoint of "Prestige: The First and Final Hurdle for a More Inclusive Philosophy, Academia.edu Talks, 2016

Debiec, J., & Sullivan, R. M. (2014). Intergenerational transmission of emotional trauma through amygdala-dependent mother-to-infant transfer of specific fear. Proceedings of the National Academy of Sciences, 111(33), 12222-12227.

Diamond, Stephen R. (2013) Status inflation and deflation: Prestige, the essence of status, permits broad alliances. Is status a positional good?, Judicial Coherence blogspot, http://juridicalcoherence.blogspot.com.ar/2013/10/211-status-inflation-and-deflation.html

Dostoyevsky, F., & Carabine, K. (2000). Crime and Punishment. Wordsworth Editions.

Eriksson, K. (2012). The nonsense math effect. Judgment and decision making, 7(6), 746.

Frankfurt, H. G., & Bischoff, M. (2005). On Bullshit (p. 553). Princeton: Princeton University Press.

Gigerenzer, G., & Selten, R. (2002). Bounded Rationality: The Adaptive Toolbox. MIT press.

Gilbert, D. T. (1991). How mental systems believe. American Psychologist, 46, 107–119.

Gilbert, D. T. (2002). Inferential correction. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 167–184). New York: Cambridge University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press.

Goldman, A. I. (2001). Experts: which ones should you trust? Philosophy and Phenomenological Research, 63(1), 85-110.

Goodwin, J. (2011). Accounting for the Appeal to the Authority of Experts. Argumentation, 25(3), 285-296.

Hansen, H. V. (2002). The straw thing of fallacy theory: the standard definition of'fallacy'. Argumentation, 16(2), 133-155.

Hanson, Robin (2016), "Against Prestige", blog Overcoming Bias.

Healy, Kieran (2013), Gender and Citation in Four General-Interest Philosophy Journals, 1993-2013, blog, https://kieranhealy.org/blog/archives/2015/02/25/gender-and-citation-in-four-general-interest-philosophy-journals-1993-2013/

Henrich, J. & McElreath, R (2003). The evolution of cultural evolution. Evolutionary Anthropology 12, 123-135.

Hume, D., & Beauchamp, T. L. (2000). An Enquiry Concerning Human Understanding: A critical edition (Vol. 3). Oxford University Press.

Inglis, M., & Mejia-Ramos, J. P. (2009). The effect of authority on the persuasiveness of mathematical arguments. Cognition and Instruction, 27(1), 25-50.

Kahneman, D. (2011). Thinking, Fast and Slow. Macmillan.

Klostermaier, K. (2013). A World without Violence?. Peace on Earth: The Role of Religion in Peace and Conflict Studies.

Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The Matilda Effect in science communication an experiment on gender bias in publication quality perceptions and collaboration interest. Science Communication, 35(5), 603-625.

Krawczyk, M., & Smyk, M. (2016). Author′ s gender affects rating of academic articles: Evidence from an incentivized, deception-free laboratory experiment. European Economic Review.

Kruglanski, A.W., and E.P. Thompson. (1999). Persuasion by a single route: a view from the unimodel. Psychological Inquiry 10: 83–109.

Lai, E. R. (2011). Critical thinking: A literature review. Pearson's Research Reports, 6, 40-41.

Malani, A. (2006). Identifying placebo effects with data from clinical trials. Journal of Political Economy, 114(2), 236-256.

Martin, J. S., Summerville, A., & Wickline, V. B. (2016). Persuasion and Pragmatics: An Empirical Test of the Guru Effect Model. Review of Philosophy and Psychology, 1-16.

Martin, J. S., Summerville, A., & Wickline, V. B. (2016). Persuasion and Pragmatics: An Empirical Test of the Guru Effect Model. Review of Philosophy and Psychology, 1-16.

Massey, G. J. (1981). The fallacy behind fallacies. Midwest studies in philosophy, 6(1), 489-500.

Milgram, S. (1963). Behavioral study of obedience, Journal of Abnormal and Social Psychology 67, 371-378

Moreland, R. L., & Zajonc, R. B. (1982). Exposure effects in person perception: Familiarity, similarity, and attraction. Journal of Experimental Social Psychology, Volume 18, Issue 5, 395–415

Mortensen, K., & Nagel, J. (2014). Armchair-Friendly Experimental Philosophy. A Companion to Experimental Philosophy, 53-70.

Myers, D. G. (2004). Intuition: Its Powers and Perils. Yale University Press.

Oaksford, M., & Hahn, U. (2013). Why are we convinced by the ad hominem argument?: Bayesian source reliability and pragma-dialectical discussion rules. In Bayesian argumentation (pp. 39-58). Springer Netherlands.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. Judgment and Decision Making, 10(6), 549.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. Judgment and Decision Making, 10(6), 549.

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. Educational research, 42(3), 237-249.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: a critical review of five decades' evidence. Journal of Applied Social Psychology 34: 243–281.

Richerson, P. & Boyd, R. (1985). Culture and the Evolutionary Process, University of Chicago Press.

Richerson, P. J., & Boyd, R. (2008). Not by Genes Alone: How Culture Transformed Human Evolution. University of Chicago Press.

Rose, D., & Danks, D. (2013). In defense of a broad conception of experimental philosophy. Metaphilosophy, 44(4), 512-532.

Schwitzgebel, E. (2012). What Experimental Philosophy Might Be. http://schwitzsplinters.blogspot.com.ar/2012/09/what-experimental-philosophy-might-be.html

Sperber, D. (1985). Anthropology and psychology: Towards an epidemiology of representations. Man, 73-89.

Sperber, D. (1996). Explaining Culture: A Naturalistic Approach (p. 119). Oxford: Blackwell.

Sperber, D. (2010). The guru effect. Review of philosophy and psychology, 1(4), 583-592.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 421–439). New York: Cambridge University Press.

Sytsma, J., & Livengood, J. (2015). The Theory and Practice of Experimental Philosophy. Broadview Press.

Thorndike, E.L. (1920), A constant error in psychological ratings, Journal of Applied Psychology, Vol 4(1), 25-29.

Tolstoy, L. (2014). War and peace (Vol. 525). Dolce Stil Publishing.

Tversky, A., & Kahneman, D. (1975). Judgment Under Uncertainty: Heuristics and Biases. In Utility, probability, and human decision making (pp. 141-162). Springer Netherlands.

Van Eemeren, F. H., & Grootendorst, R. (1992). Argumentation, communication, and fallacies: A pragma-dialectical perspective. Lawrence Erlbaum Associates, Inc.

Van Eemeren, F. H., & Grootendorst, R. (2004). A systematic theory of argumentation: The pragma-dialectical approach (Vol. 14). Cambridge University Press.

Van Eemeren, F. H., Garssen, B., & Meuffels, B. (2009). Fallacies and Judgements of Reasonableness. Dordrecht: Springer.

Vasari, G. (1991). The Lives of the Artists. Oxford University Press, UK.

Walton, D. (1997). Appeal to Expert Opinion: Arguments from Authority. University Park.

Walton, D. (2003). A Pragmatic Theory of Fallacy.

Walton, D. (2010). Appeal to Expert Opinion: Arguments from Authority. Penn State Press.

Walton, D. (2010). Why fallacies appear to be better arguments than they are. Informal Logic, 30(2), 159-184.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. Journal of Cognitive Neuroscience, 20(3), 470–477.

Weisberg, D. S., Taylor, J. C., & Hopkins, E. J. (2015). Deconstructing the seductive allure of neuroscience explanations. Judgment and Decision Making, 10(5), 429.

Zagar, I., & Mohammed, D. (2011). Fallacies: do we “use” them or “commit” them? Or: is all our life just a collection of fallacies?