

STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Thesis in Economics
Spring 2024

Correcting False Information: An Experimental Study on Charitable Giving

Evan Shokor Qasim (42415)

In light of mounting concerns over mis- and disinformation, in this paper I provide a novel preregistered incentivized non-deceptive economic experiment measuring if exposure to false negative information about a charity that is quickly corrected causes a lower mean donation to the charity than no exposure to false negative information. With a sample of 175 students, there is suggestive evidence ($p < 0.05$) of an effect in the opposite direction than hypothesized, to be interpreted with caution and not robust to additional testing. No statistically significant effect ($p < 0.005$) is found. The results indicate that exposure to false negative information that is quickly corrected does not cause a lower mean donation to the charity than no exposure to false negative information. Future experiments may benefit from hypothesizing effects in both directions and by aiming to find smaller effect-sizes.

Keywords: bounded rationality, charitable giving, continued influence effect, correction, false information

JEL: C99, D83, D91

Supervisor:	Anna Dreber Almenberg
Date submitted:	May 13, 2024
Date examined:	May 24, 2024
Discussants:	Sofia Berntsson and Malin Holmqvist
Examiner:	Magnus Johannesson

Acknowledgements

I want to express my deepest gratitude to Anna Dreber Almenberg. I also want to thank Marijane Luistro Jonsson for sharing her expertise and thoughts.

Contents

1	Introduction	4
2	Background	6
2.1	Misinformation, Disinformation, Fake News	6
2.2	The Impact of False Information on Society and Charitable Giving in Particular	7
2.3	Advancements in Technologies for Spreading False Information	9
3	Literature Review	10
3.1	Corrected False Information – Continued Influence	10
3.2	Drivers of Belief in False Information	11
3.3	Corrected False Information in Economics	12
3.4	Drivers of Charitable Giving	16
4	Theoretical Framework	18
5	Method	21
5.1	Preregistration	21
5.2	Choosing Charity Organizations	21
5.3	Choosing, Presenting, and Correcting an Artifact of False Negative Information	23
5.4	Experimental Design	28
5.5	Research Question, Testing, and Econometric Specification	32

5.6	Participants and Power Calculations	34
6	Results	36
6.1	Descriptive Statistics	36
6.2	Main Test Results	37
6.3	Robustness Test Results	38
7	Discussion	39
8	Conclusion	42
9	Limitations and Future Research	43
A	Appendix	58
A.1	Preregistration	58
A.2	Survey	60
A.2.1	Survey Distribution	66
A.3	Impure Altruism in Charitable Giving	68
A.4	Implementing Donations	69

1 Introduction

The purpose of this study is to understand whether individuals who are exposed to false negative information regarding a charity, upon correction, donate differently than they would if they were never exposed to the false information in the first place. There are three main reasons why this merits attention. *(i)* There is reason to believe that a difference may indeed occur. For instance, many psychological experiments find that correction does not completely mitigate exposure to mis- and disinformation in a cognitive sense due to a continued influence effect (CIE) from mis- and disinformation (Chan et al., 2017). *(ii)* The economic literature on corrected mis- and disinformation is limited and as a result we do not know much about it.¹ While there are plenty of psychological studies, differing traditions between psychological and behavioral economic studies mainly with regard to monetary consequences make these findings insufficient with regards to economic decision making. Nevertheless, implications that arise from lingering effects from mis- and disinformation even after correction is substantial for the development of behavioral economic theory, especially the role of intuition versus reason (Kahneman, 2003). *(iii)* Mis- and disinformation is an increasingly problematic and complex problem that threatens the fabric of society (Colomina et al., 2021; Palmertz et al., 2024), if there is a stark difference between correcting mis- and disinformation versus preventing mis- and disinformation from ever being spread in the first place, organizations and policymakers should be aware of this to assess how they want to tackle mis- and disinformation. While an answer to the question of this study does not generalize to all contexts involving mis- and disinformation, it serves as a stepping stone.

In constructing a non-deceptive method with monetary consequences to answer the question of this study, the main contributions of this study to the field of economics and behavioral science in general is: *(i)* Broadening our understanding of the usefulness of corrections as opposed to prevention, when applied to charitable giving. *(ii)* A method which can be suggested for future researchers interested in exposing participants to false information, without relying on deception.

In this study I test the correction of mis- and disinformation pertaining to charitable giving. There are other usual targets of mis- and disinformation (such as governments or businesses) which could be explored further in the future. Charities collect

¹Subsection **3.3** provides a review of existing economic literature.

donations from the public by bringing attention to people in need (Merchant et al., 2010; Small & Verrochi, 2009). Donating money to a charity is a simple way to help people, at an expense that one has control over. However, if a donor is exposed to false negative information regarding a charity – especially if it has to do with the believed impact of a donation – he or she could potentially decrease or discontinue their allocation of money to the charity (Caviola et al., 2014), potentially resulting in a welfare-loss to society. Correction of mis- and disinformation pertaining to charitable giving is important, relevant, and lends itself well to be studied with economic experiments with monetary consequences.

I design a between-subject randomized experiment where the focus is on whether participants donate differently to the American Red Cross depending on whether they are exposed to false negative information which is immediately corrected (treatment condition), or whether they are never exposed to the false negative information (control condition). I hypothesize that donations are lower among those in the treatment condition compared to the control condition. I use a two-tailed t-test to test whether there is a statistically significant ($p < 0.005$) difference in means between the two conditions. As a robustness check I run an OLS multiple regression with heteroskedastic robust standard errors, with the added covariates gender, nationality, and age.

After a period of 61 days of data collection (March 1, 2024 until April 30, 2024), I have a sample of 175 students at the Stockholm School of Economics. The difference in mean amount donated to the American Red Cross between the two conditions is 1.161 USD – and surprisingly, this is in the opposite direction of my hypothesis. However, this should be interpreted with caution as it is only suggestive ($p < 0.05$) and in the opposite direction of my hypothesis. In addition, this result is not robust. The results thus largely suggest no robust effect of the treatment compared to the control. I do not find support for a CIE or intuitive thinking rather than reason. Future experiments may benefit from hypothesizing effects in both directions and by aiming to find smaller effect-sizes.

2 Background

2.1 Misinformation, Disinformation, Fake News

In economic academic and policy contexts, both misinformation and disinformation have become increasingly problematic and is sought to be better understood through behavioral science (OECD, 2022). An understanding of the difference between the two can be useful if one wishes to understand the meaning of "fake news", weaponized false information, or any situation where an actor is intentionally trying to misinform another party. Disinformation is often distinguished from misinformation on the basis of intent. While misinformation merely conveys inaccurate information, disinformation does so in a deliberate manner for profit or to cause public harm (Bennett & Livingston, 2020; Fallis, 2015; Kapantai et al., 2021). To define these concepts in economic terms, one suggestion is to define disinformation as: misinformation that is deliberately disseminated by a person to obtain utility by making other people's behaviors change, and misinformation as: information that is not objectively correct (Harashima, 2023).

Whether it be tactics of hybrid warfare, deceitful political opinion formation in general, or attacks against certain businesses or organizations, these can be categorized as tactics of disinformation. The distinction between mis- and disinformation may not always be clear cut, as intent cannot always be proven. What is exceedingly simple to see however, is that disinformation is rampant, and through modern channels of media and by increasingly sophisticated methods it threatens not only individuals, groups, businesses and organizations, it also threatens the very fabric of democratic societies (Colomina et al., 2021; Palmertz et al., 2024). The threat is recognized all the way up to the government level across the world (Bennett & Livingston, 2020; Kapantai et al., 2021; Swedish Ministry of Finance, 2023).

In this study I sometimes use "false information" as a substitute for typing out "mis- and/or disinformation", this choice greatly simplifies readability. As "fake news" is a type of disinformation, in that it is false information that is intentionally deceitful with the added property of resembling real news (see for instance Allcott and Gentzkow (2017), Gelfert (2018), Jost et al. (2020), Lazer et al. (2018), Pennycook et al. (2018), and Tandoc Jr et al. (2018)), it falls into the broad category of

"false information". I write "false negative information" to specify negative connotations. To further improve readability, the false negative information that those in the treated condition are exposed to is referred to as the "artifact of false negative information" or "artifact" for short.² While one could potentially classify the artifact as disinformation, using the term artifact is prudent with regards to the origins and intentions of the artifact. Table 1 provides a simple overview of all relevant terms and their definitions.

Table 1: Definition of Misinformation, Disinformation, Fake News, False Information, and Artifact.

Term	Definition
Misinformation	Information that is objectively not correct
Disinformation	Misinformation that is deliberately disseminated
Fake News	Disinformation that poses as real news
False Information	A term that is used throughout this paper as a catch-all term
Artifact	The false negative information that those in the treated condition are exposed to

2.2 The Impact of False Information on Society and Charitable Giving in Particular

In Sweden, due to recent geopolitical advancements there is a growing concern for being victim to hybrid warfare, which can take the form of external false information campaigns that aim to interfere with democratic decision-making processes (Palmertz et al., 2024). The government is taking steps to better understand and mitigate these problems (Swedish Ministry of Finance, 2023), and increase resilience and psychological defence capabilities (Palmertz et al., 2024). As such, the potential negative impact of false information on Swedish society as well as more globally is crystal clear. In this study, however, I focus on the specific case of false information in the context of charitable giving.

²"Artifact" is useful terminology borrowed from literary and rhetorical scholarship where the object of a (typically) qualitative analysis is frequently referred to as an artifact.

Sweden is most likely not more susceptible than other countries to false negative information when it pertains to charities. Sweden enjoys high levels of social trust, it is shrinking in vulnerable groups e.g. people that are unemployed, young, in poor health, on sickness benefit or activity compensation, and in certain political groups such as those sympathizing with the Sweden Democrats or those without any party preference (Rothstein & Holmberg, 2022). Nevertheless, social trust is still high, and Sweden does enjoy high social capital as evidenced by the Social Capital Index (SolAbility Sustainable Intelligence, 2023), and remains in the top quintile of charitable givers as indicated by the World Giving Index (Charities Aid Foundation, 2023). Still, it should not be glossed over that polarization and distrust in NGO's (which encompasses charity organizations) is increasing and large when compared to 26 countries (sample size ranging from 1,082 to 1,500 for every country) in the 2023 Edelman Trust Barometer Global Report (Edelman, 2023a, 2023b).³

NGO's can become severely affected by falling levels of distrust, as the sector is heavily reliant on public perception (Gaskin, 1999). False information regarding charities could exacerbate negative perceptions of these institutions. Still, in Sweden lower trust in non-traditional as opposed to traditional media (Edelman, 2023a, 2023b) could make false negative information less of a concern as the barrier for false information to spread on traditional platforms is considerably higher than on social media. For example, in countries with lesser trust in NGO's and relatively high trust in non-traditional media sources, one may expect false negative information from non-traditional media sources to be more effective than in Sweden.

Increasing polarization can be the ends to which false information is the means, potentially empowered through newer forms of media. I would like to make note that common objects of criticism such as "filter bubbles" may be less of a concern than typically believed. There are suggestions that the effect of filter bubbles and algorithmic thinking is relatively modest compared to individual choices when it comes to media consumption (Bakshy et al., 2015; Flaxman et al., 2016). Confirmation bias, meaning that individuals actively seek out what aligns with their already established attitudes, is likely strong. Still, some initial exposure to false information could shape the belief which drives the confirmation bias. Likely, some interplay of confirmation bias and filter bubbles coupled with increasing false in-

³The sample of countries consists of Argentina, Australia, Brazil, Canada, China, Colombia, France, Germany, India, Indonesia, Ireland, Italy, Japan, Kenya, Malaysia, Mexico, Nigeria, Saudi Arabia, Singapore, Africa S., Korea, Spain, Sweden, Thailand, The Netherlands, UAE, UK, U.S.

formation campaigns and harsher political rhetoric (for instance as a consequence of economic downturns (Funke et al., 2016)) accounts for increasing polarization, as does many weak online ties (i.e. a large social network) for the dissemination of information (Bakshy et al., 2012). In societies where false information on social media is more commonly perceived as authentic, new media may be a more powerful driver of increasing polarization and distrust.

2.3 Advancements in Technologies for Spreading False Information

Rapid advancements in artificial intelligence increases the the perceived authenticity of false information (Kreps et al., 2022; Nightingale & Farid, 2022; Pantserev, 2020), and the potential for false information to reach a target audience (Aïmeur et al., 2023; Bontridder & Pouillet, 2021; Hajli et al., 2022; Kozyreva et al., 2020). This has led to a need to reinvent ways to detect false information. Promising efforts have been put forward, such as deep-learning for fake news detection (Kaliyar et al., 2020), combining linguistic cue (machine learning) with network-based behavioral data when searching for disinformation (Conroy et al., 2015), and techniques to detect Deepfakes (synthesized media imitating someone) (Gupta et al., 2024). It stands to question how useful false information detection is once the false information is already widely spread. Even if the false information is detected, there are challenges in correcting it and reaching those who have already been exposed to it. It may be difficult to quantify how people who have been exposed responded to the false information, and it may be challenging to correct their misconceptions.

False information is not only problematic because it has the power to misguide people’s beliefs. It is also a problem in and of itself, the mere existence of false information is damaging whether or not it is believable. For instance, it has been shown that Deepfakes lead to less trust in news on social media and a deterioration in online civil culture in democratic societies – even if the false information is not widely believable (Vaccari & Chadwick, 2020). As both governmental and all types of non-governmental entities (e.g. NGO’s, firms) are increasingly easier targets for agents who want to spread false information, it is in their interest to know how useful the detection of false information is, and whether it is more important to stop it from being able to spread as opposed to detecting and correcting it in hindsight.

While the former option may be more difficult to achieve, it may turn out to be the most effective way to establish trust and a favorable perception in the public eye.

3 Literature Review

3.1 Corrected False Information – Continued Influence

In psychology there have been great advancements in documenting how corrected false information differs from never having been falsely informed in the first place. Continued influence of false information even after receiving a correction is known as the continued influence effect (CIE) (Chan et al., 2017; Ecker et al., 2022; Johnson & Seifert, 1994; Walter & Murphy, 2018; Walter & Tukachinsky, 2020). CIE relies on activation of information. False information and correction coexist and compete for activation. Avoiding CIE may require both successful integration (the individual properly processes the correction) and retrieval (the individual properly remembers the correction) (Ecker et al., 2022). Meta-analyses of psychological lab experiments have found that correction does not satisfactorily remove the effects of misinformation (effect-size of persistence of misinformation in the face of debunking is large: Cohen’s $d = 0.75$ – 1.06 (Chan et al., 2017)).

However, the psychological literature has not relied on outcomes from economic decision making but more so on the cognitive understanding of events and memory (Chan et al., 2017; Lewandowsky et al., 2012; Walter & Tukachinsky, 2020). Still these studies generate interesting findings for economists as CIE could be prevalent in economic decision making. To the best of my knowledge, neither psychology nor any other field offers studies that examine CIE in the context of charitable giving, given monetary consequences.

A multitude of literature offers suggestions for how to properly correct false information and mitigate CIE, on the basis that correction is typically ineffective (Lewandowsky et al., 2012). There are three broadly accepted suggestions: (a) raising awareness of misleading information prior to exposure; (b) repeated corrections; (c) alternative explanations to the false information.

3.2 Drivers of Belief in False Information

In psychology, the predominant view is that false beliefs are formed similarly to real beliefs. In general individuals are inclined to be biased in their assessments of information validity, the process is usually not very deliberative (Ecker et al., 2022). Neo-classical economists may suggest that individuals can be subject to making decisions based on false information (Stigler & Becker, 1977), yet rationality drives belief in information. Under behavioral economics, bounded rationality suggests that the cognitive limitations of knowledge and computational capacity constrain people’s ability to make optimal economic decisions. Such cognitive failures could explain why some false information is believed, and also why correction may fail. Especially important, further explored in section 4, is System 1 and System 2 thinking, the everyday analogy for which would be intuition and reason (Kahneman, 2003).

Fundamental mechanisms stemming from psychology are argued to explain why individuals are susceptible to false information. Mechanisms can generally be divided into cognitive and socio-affective drivers (Ecker et al., 2022). Examples of cognitive drivers are: (1) Intuitive thinking, rather than analytical thinking. This can be likened to thinking fast, rather than slow, a very familiar concept in behavioral economics (Kahneman, 2003). (2) Cognitive failures, such as forgetting knowledge or forgetting corrections. Cognitive failure can also arise due to media failing to tag false information when it is common practice in a medium (perhaps in combination with the exposed person lacking analytical thinking) (Pennycook et al., 2020). Another potential for cognitive failures can be a lack of tangible information (Bachke et al., 2017; C. Cryder & Loewenstein, 2011; C. E. Cryder et al., 2013). (3) Illusory truth, such as prior exposure to a message, i.e. repetition can also affect the perceived legitimacy (Ecker et al., 2022; Pennycook et al., 2018).

Examples of socio-affective drivers of belief in false information are: (4) Source cues giving favor to elites or in-groups, one example is giving favor to elite rhetoric regarding authenticity of certain news media (Ladd, 2010; Van Duyn & Collier, 2019). Source credibility is subjective (Castillo et al., 2011; Fragale & Heath, 2004; Li & Sakamoto, 2014), and thus sources are perceived as varyingly trustworthy. (5) Emotional content, appeals to emotion can increase persuasion, and distract people from important diagnostic cues such as source credibility. (Ecker et al., 2022). (6) Worldview, which can be tied to selective exposure and confirmation

bias. Selective exposure means that within an environment one gives particular attention to that which confirms one's attitudes, to limit cognitive dissonance (Spohr, 2017). Confirmation bias means seeking out and assigning validity to information that supports one's existing attitudes (Lazer et al., 2018).

In accordance with bounded rationality, loosely speaking, descriptions of both cognitive drivers and socio-affective fit into what is known as "fast thinking", which is explored in depth in section 4. Intuitive thinking (1) is obvious; forgetting knowledge (2) could be related to the accessibility dimension of information which may be more or less accessible and in turn forgetful; and illusory truth (3) could be connected to cognitive ease. Socio-affective drivers (4) and (6) build on reasoning similar to judgement heuristics, as does driver (5) specifically for affective heuristics (Kahneman, 2003). The point to make is not that fast thinking encapsulates all of the psychological explanations for belief in false information, it is rather that fast thinking could lead to belief in false information in some cases. Note that processing false information is not fundamentally different from processing truth. Some lies are credible even to the reasoning mind and therefore one cannot ascribe fast thinking as the singular explanation for why people believe in false information. One may however more reliably claim that if a convincing correction does not succeed in making a person act as they would prior to the false information, it may merit the explanation of fast thinking.

3.3 Corrected False Information in Economics

The economic literature on corrected false information has emerged in various topics and seems most emphasized in studies of policy preferences. It seems underdeveloped across other settings. This is not surprising, as false information plays a large role in the political sphere. In the following paragraphs, some of the economic literature on corrections is reviewed. To the best of my knowledge, this is representative of the existing corpus specifically relating to economics, not exhaustive however as the literature is scattered and thus additional articles may bear limited relevance.

In financial economics, a natural experiment was brought on by six-year-old information being mistaken for news. The old information was a potential bankruptcy filing by the United Airlines' parent company. The stock price remained lower than

predicted by a simple asset-pricing model for almost a week despite the information being corrected as old news (Carvalho et al., 2011). Other financial economic papers are concerned with the effect of different types of news on the stock market, note however that the emphasis is not on corrected false information (see for instance, the literature regarding old news re-emerging (Fedyk & Hodson, 2023) and the impact of fake news on stock returns (Arcuri et al., 2023; Clarke et al., 2020)).

Jost et al. (2020) focus on testing the persistence of anchoring. Through their online survey, a within-subject experimental setting with student participants (N=240) tests fake news with a high anchor, varying the amount of awareness of fake news (thus, some participants knew that the fake news were indeed fake) and source authenticity. Participants in different treatment conditions evaluate the value of a train station, which is not general knowledge. Good evaluations are rewarded by chocolate. The results show that both awareness and source authenticity is key in combating fake news. All participants, among which some are corrected, are affected by the anchor. Some scholars may argue that this is an unsurprising insight as anchoring is one of the most well established judgemental heuristics in behavioral economics (Tversky & Kahneman, 1974). Anchors are not statements that are to be argued for or refuted (Furnham & Boo, 2011), perhaps a correction for an anchor should be a counter-anchor. Mitigation of anchoring has been shown possible for example through a consider-the-opposite approach (Mussweiler et al., 2000).

Ali and Qazi (2023) further our understanding of corrections of false information in development economics. Their between-subject randomized experiment sampling from the field in Pakistan (N=750) tests the effect of educational interventions on countering misinformation among populations with low digital literacy. Using three conditions (baseline, video, video + feedback), they find that video-based educational messages about misinformation augmented with feedback based on individual's past engagement with fake news has an impact on the ability to identify false information and is thus better than solely providing educational messages. This could be because video + feedback lowers the cost of analytical reasoning which helps individuals form accurate beliefs about news, and the effect is especially pronounced for news about items that are essential for survival. When the cost of not knowing the truth is large, participants make better assessments.

Many researchers in economics and intersecting disciplines have studied whether Americans' policy preferences regarding immigration change as a result of correct

information, as misconceptions are common. These researchers do not expose participants to false information, they merely correct pre-existing false information. Common findings are that in spite of correction, policy preferences remain as they were before the correction (Alesina et al., 2020; Grigorieff et al., 2020; Hopkins et al., 2019). Korean findings mirror those from the U.S. (Choi et al., 2023).

There are recent economics papers related to a larger literature in political science.⁴ Researchers are interested in how persuasive the "alternative facts" of far-right politicians are, and their resilience to corrections. These studies expose participants to false information and correct them (similar to this study but the false information comes from statements by politicians) as opposed to the studies mentioned in the paragraph above that simply provide participants with correct information regarding immigration. Barrera et al. (2020) conduct a randomized online survey experiment (N=2480), participants are assigned into one of four conditions: a control group reading no text, an alt-facts group reading alternative facts, a facts group reading facts, and a fact-checking group reading alternative facts and then facts. The results show that alternative facts are highly persuasive, and that correction does improve factual knowledge (cognitively, participants understand and accept the correction) but does not affect policy conclusions or support for the political candidate. Mechanisms driving beliefs in false information and correction are not tested for, however their results are in line with their hypothesis that raising the salience of immigration issues leads to some participants favoring anti-immigration policies.

Henry et al. (2022) run a preregistered, randomized study surveying voting-age Facebook users (N=2537) and set out to test whether correction has any effect on participants dissemination of alternative facts. Three conditions are employed: participants exposed to false statements by a far-right party; participants exposed to false statements by a far-right party followed by fact-checks; participants exposed to false statements by a far-right party and offered whether to be exposed to the fact-checks. Both imposed and voluntary fact-checking reduces dissemination, with

⁴For instance, researchers have aimed to evaluate correction on Trump supporters' beliefs and voting during the 2016 U.S. presidential election. Swire et al. (2017) survey voting-age U.S. citizens (N=1776), between-subject results find that political partisanship and correction source affects whether the correction is accepted. Within-subject results show that voters do not change their voting behavior after correction. Nyhan et al. (2020) use two between-subject randomized experiments on U.S. voting-age citizens, the first of which is preregistered (N=4186; N=1546). A similar result to Swire et al. (2017) is found, Trump voters correct their beliefs but do not change partisanship (Nyhan et al., 2020).

similar effect-sizes. They also find that each additional click required to disseminate reduces sharing.

Focusing on the American 2022 mid-term elections, Guriev et al. (2023) run a pre-registered randomized experiment on U.S. Twitter-users (N=3501) evaluating potential policies as treatments for reducing the spread of false information online and increasing the spread of true information. Participants were given the possibility to share 4 tweets (two of which are false news) and were assigned to one of 5 conditions: no policy treatment (baseline); extra click treatment (required participants to click an additional time before sharing); priming fake news circulation treatment (participants were reminded to be careful with regards to fake news circulating); offer fact checking treatment (participants could choose to read corrections); ask to assess tweets treatment (participants were to evaluate tweets before sharing them). All treatments are found to decrease sharing to varying degrees. Compared to the baseline, the percentage point decrease in sharing false information was 3.6, 11.5, 13.6, and 14.1 respectively. Compared to the baseline, only priming fake news circulation increases sharing of true news (offering fact checking is the worst policy in this regard, even when compared to the baseline) which leads to the conclusion that such a policy is most effective in positively changing the composition of fake and true news in online environments.

Evident from the reviewed literature, current experiments on corrected information is not typically contingent on direct monetary consequences for participants. This is not a criticism of the existing literature, as for policy experiments people's preferences may still be monetarily motivated (e.g. a perception of immigration as a waste of tax money) and choosing ones preferred policy is free of charge in the real world. The fact that the literature reviewed is mostly survey-based is probably not indicative of any weaknesses in findings as lab-and survey experiments have been found to exhibit similar levels of internal validity (Arechar et al., 2018; Prissé & Jor-rat, 2022). Furthermore, survey-designs has allowed for large samples which lends credibility to the findings of related literature. On the other hand, my study follows the notion of $p < 0.005$ as the threshold for statistical significance, as suggested by Benjamin et al. (2017) which helps combat false positive results. This stringency is yet to be adapted broadly, and this includes the study of corrected false information.

3.4 Drivers of Charitable Giving

What drives people to engage in charitable giving? In an economic framework, people donate if the willingness to donate exceeds or is equal to the cost of donation.⁵ People’s preferences and constraints decide their individual willingness to allocate money across charities. In an influential paper impure altruism, where utility from charitable giving is argued to be derived both from altruistic and selfish reasons, is raised as a theoretically sound explanation for why people donate (Andreoni, 1990). Thus, impure altruism could drive donations from an economic perspective. There is a vast literature across disciplines that aims to isolate and explain drivers more specifically.

In an interdisciplinary review of more than 500 articles (Bekkers & Wiepking, 2011), eight mechanisms for charitable giving are uncovered: (a) awareness of need; (b) solicitation; (c) costs and benefits; (d) altruism; (e) reputation [external affirmation]; (f) psychological benefits [internal affirmation]; (g) values; (h) efficacy [perception of making a difference]. Briefly stated, I argue that these mechanisms can be represented by a utility model of impure altruism, subject to constraints and given a set of preferences when exposed to some information.⁶ Importantly, dismissing pure altruism or warm-glow altruism as mechanisms evidently neglects empirical findings. Regarding efficacy (h), survey studies have shown that when people perceive that their donations are not making a difference as intended, they are less likely to donate (Arumi et al., 2005; Bekkers & Wiepking, 2011; Borgloh et al., 2013; Diamond & Kashyap, 1997; Duncan, 2004; Mathur, 1996; Radley & Kennedy, 1992; Smith & McSweeney, 2007). Thus an altered perception of donation impact should be problematic for donors. Further, some studies suggest that quality certifications affect donorship (Adena et al., 2019; Brown et al., 2017). Low overhead costs (administration costs etc.) in charities are preferred over cost effectiveness (lives saved per

⁵In an interesting working paper, charities are analyzed as optional price discriminators (Fennell, 2022). I believe this could serve as an interesting avenue for researchers who wish to understand welfare effects from false information and correction in the context of charitable giving.

⁶Impure altruism encapsulates both the altruistic and selfish mechanisms (d, f). Awareness, solicitation and costs and benefits (a, b, c) can widely be said to be dealt through cognitive processes and exposure to information. I discuss my stance on cognition in section 4, note that regarding costs and benefits there is extensive documentation that increasing benefits to participants of experiments can crowd out the intrinsic motivation behind donating (Titmuss, 2018). Regarding reputation (e), participants are anonymous thus I do not expect external affirmation to be a mechanism for charitable giving in my study. I consider values and efficacy (g, h) as part of people’s preferences (see section 4).

dollar) as the former is easier to evaluate. This is due the evaluability bias which suggests that people weight an attribute as per the ease of its evaluation (Caviola et al., 2014).

Psychological factors that impact donations, especially monetary donations to disaster victims (such are the beneficiaries of the three charities in this study), have been reviewed thoroughly by Zagefka and James (2015). Factors such as age, sex, and level of education are correlated to donorship. Both positive and negative moods have been shown to boost donations, as opposed to neutral moods (Aknin et al., 2012; Niesta Kayser et al., 2010). Mood affects whether to donate and how much to donate, the latter is related to feelings toward the recipient (Oppenheimer & Olivola, 2011). Emotions such as guilt, nostalgia, and elevation among other emotions could also factor in. Incidental emotions have also been found to affect the allocation of donations (Ibanez et al., 2017). Feeling sad for the recipients of the donation has been shown to increase donations (Oppenheimer & Olivola, 2011; Small & Verrochi, 2009), especially when the recipients belong to the same social group as the donor (Weisbuch & Ambady, 2008). Donorship has further been shown to increase happiness in the donor (Aknin et al., 2012), more so than by purchasing for oneself (Dunn et al., 2008), and the frequency of donations is more important for happiness than the sum of the donations. Several smaller donations are better than a single large one (Oppenheimer & Olivola, 2011). Interestingly regarding emotions, when false negative information is corrected, people are found not to cling to the most emotional version of events, according to the psychological literature (absent economic decision-making). This being said, emotive misinformation is more likely to be disseminated (Lewandowsky et al., 2012). In this study I do not isolate emotive mechanisms and their extents with surgical precision, however I find it reasonable that there may be e.g. emotive negative responses to false negative information.

Different types of information have been found to affect donations to varying degrees. Tangible (specific and concrete) information in large quantity, especially when the information relates to how the donation beneficiaries will be better off from the donation, can increase donations (Bachke et al., 2017; C. Cryder & Loewenstein, 2011; C. E. Cryder et al., 2013). When attempting to isolate drivers that influence donations, one should not forget to consider whether the information given to participants allows the experimenter to measure the effect they intend to measure, or if participants are in fact responding to some ambiguity in the given information.

Trade-offs are an important consideration in charitable giving. In this study, there are monetary consequences to decisions but participants do not face trade-offs as they do not get to keep any of the money they are given to allocate. When facing trade-offs, any risk associated with giving can be used as an excuse not to give. It has been found that when participants in an anonymous setting are faced with two types of risk: own payoff and charity payoff, their risk aversion differs depending on whether they face trade-offs. When participants do not make trade-offs their aversion to both risks is comparable, however when participants make trade-offs they show higher aversion to charity risk than own risk, indicating self-serving behavior (Exley, 2016). Having participants face personal trade-offs would better reflect how charitable giving manifests in real life. Furthermore, being given a sum of money, all of which has to be donated, may be perceived as making a decision with another agent’s money. This may have some important implications, since one’s own money has been found to be perceived as more valuable than other people’s money on account of psychological distance (Polman et al., 2018). In this study, this could make a relatively small donation feel insignificant (participants are given 10 USD to allocate), which could potentially introduce more noise on account of participants feeling indifferent about their choices. However, there is evidence that in general, decision-making given other people’s money is similar to decisions with one’s own money when there is an identifiable victim but the willingness to pay differs (Ifcher & Zarghamee, 2020).

4 Theoretical Framework

Bounded rationality suggests rational economic decision-making is constrained by cognitive limitations. Such limitations can be knowledge or computational capacity. The neo-classical subjective expected utility (SEU) theory relies on a number of assumptions such as there being a fixed set of alternatives, knowing the probability distribution of outcomes (this is subjective) for each alternative, and maximizing the expected value of a utility function. By relaxing one or more of these assumptions we generate theories of bounded rationality (Simon, 1990).

Bounded rationality can explain the CIE that is found so prevalent in psychology. Bounded rationality fits the framework where corrections are dependent on correct processing and remembering of information. Cognition can be divided into

two systems, System 1 and System 2 (Stanovich & West, 2000). System 1 can be fast, automatic, effortless, associative, and emotionally charged. System 2 can be slow, serial, effortful, deliberately controlled, relatively flexible, rule-governed (Kahneman, 2003). System 2 can monitor System 1 but this does not guarantee that intuitive judgements are hindered from expression (Gilbert, 1989). The accessibility dimension of bounded rationality lends credibility to the idea that false negative information and correction coexist and compete for activation, some thoughts are more accessible than other thoughts, in cases when the false information is much easier to process than the correction this may lead to CIE. Natural assessments are those that are carried out by System 1, and a natural assessment of particular importance is the evaluation of stimuli as good or bad (Kahneman, 2003). This can be a problem for the correction of false negative information, since the false negative information may be more accessible due to it painting the American Red Cross as bad: dishonest and inefficient. CIE can arise from a lack of System 2 monitoring. Beliefs about the American Red Cross are then worse than prior to exposure to false negative information, and preferences do not revert to where they were before, and giving to the American Red Cross declines. This line of thinking motivates the alternative hypothesis in this study.

The way that the artifact and correction affect decision-making in my experiment can be formalized by assuming 3 different options of charitable giving (reflecting participants' choice in this study). Let $\mathbf{g}_{initial} = (g_{1,initial}, g_{2,initial}, g_{3,initial})$ be the vector containing different amounts of each of the options and we can call $\mathbf{g}_{initial}$ the giving bundle. Let g_1 be giving to the American Red Cross and g_2, g_3 either of the remaining two organizations. The giving set X provides all of the alternatives givers can choose and we assume $X = \mathbb{N}_+^3$ as participants donate in increments of USD. $\mathbf{g}_{initial}$ describes the choice to give as a result of preference relations before exposure to false negative information. $\mathbf{g}_{false} = (g_{1,false}, g_{2,false}, g_{3,false})$ is the giving bundle after exposure to false negative information, we assume $g_{1,initial} \neq g_{1,false}$. Even given a stringent neo-classical definition of preferences as exogenous, i.e. in general stable over time and context (Stigler & Becker, 1977), factors such as changed incentives, beliefs of expected payoff, and information are thought to be capable of altering preferences (Cohn & Maréchal, 2016; Fehr & Hoff, 2011). Note that beliefs can be both correct and incorrect (Stigler & Becker, 1977). The first bundle is based on beliefs about the American Red Cross, and when those beliefs change so does the bundle as a consequence of altered preferences (Stigler & Becker, 1977). After

correction, the giving bundle reverts to the original bundle and $g_{1,initial} = g_{1,corrected}$, if the original preference relations (by way of beliefs) are re-established. What I suggest is that $g_{1,initial} \neq g_{1,corrected}$ may be prevalent, more specifically I suggest $g_{1,initial} > g_{1,corrected}$.

I have established that people may be susceptible to CIE, which can be understood through bounded rationality. If the mean amount donated to the American Red Cross is lower in the treatment condition, $g_{1i} \downarrow$, then obviously $g_{1,initial} > g_{1,corrected}$ and CIE could be driving this decision. Let's consider the two other possibilities: while I hypothesize the mentioned direction it could be useful to consider explanations if my hypothesis is wrong. If the mean amount is unchanged, $\bar{g}_{1i}$, then $g_{1,initial} = g_{1,corrected}$ and CIE may not have been of concern in this particular study. If the mean amount is higher, $g_{1i} \uparrow$, then $g_{1,initial} < g_{1,corrected}$ and there is no CIE. This finding would be difficult to reconcile with System 2. An emotive backlash to the false negative information, a System 1 response, seems more reasonable.

I believe that a utility function of impure altruism captures charitable giving, and with this type of utility charitable giving occurs because of both pure altruism as well as selfishness, which reflects real-life motives for giving (Andreoni, 1990). Impure altruism gives room for utility as a function of both pure and warm-glow altruism, which is both realistic and convenient for this study since it does not bind the analysis to considering purely selfish or purely altruistic motives, both of which are inefficient as economic theories of giving, leading to strange implications. For instance, pure altruism would make government grants and subsidies crowd out voluntary gifts, and in large economies only a sliver of the population will donate to charity (Andreoni, 1990). I do not consider inequity averse preferences in my model (Fehr & Schmidt, 1999), however I am not concerned unless the baseline mean donated to the American Red Cross is low, such that donations with the multiplier it is comparable to 1/3 of the total donation.

I do not make any contributions to the existing influential model of impure altruism, instead I simply tweak it so that private consumption of a private good $\bar{x}_i$ is not possible, givers cannot convert the private good to the public goods by linear technology, and I specify three charities instead of one. **A.3** provides a more thorough justification and walk-through.

5 Method

5.1 Preregistration

In order to limit researcher degrees of freedom which may lead to false positives, and thus contributing to good scientific practice (Simmons et al., 2011), I created a preregistration which is available on <https://aspredicted.org/d3zw7.pdf> and it can also be seen in its entirety in the appendix under **A.1**. It was published on February 29, 2024, one day before data collection which started on March 1, 2024 and ended on April 30, 2024. Essential parts of the method presented in the coming pages, such as variables and analyses, can be found already in the preregistration, which I strictly followed.⁷

5.2 Choosing Charity Organizations

Participants are able to choose between three charity organizations to allocate the donation, namely the American Red Cross, USA for UNHCR, and Doctors Without Borders USA. These charities are among the top 10 most popular charities in Sweden (see figure 1), and they are relatively similar. Trying to add another organization could compromise the commonalities between the organizations and how familiar participants are with the organizations. Using only two charities may limit participants in their allocations in case they hold a negative view of one of the charities since prior. The chosen charities are selected on the basis of *(i)* being well known in Sweden, proven by the fact that their Swedish counterparts are consistently among the top 10 most donated to by Swedish citizens (Giva Sverige, 2020, 2021, 2022), *(ii)* being similar, having overlapping areas of operations in the humanitarian domain which are not exclusive to certain age groups (e.g. children) or diseases (e.g. cancer, organ-specific), and *(iii)* having organizational legs outside Sweden, such that organizations that are well known in Sweden can be chosen from a distant country.

Regarding the first and second requirement, it serves the experiment to choose charities that are perceived as being as similar as possible. While little is known

⁷I did not mention visualizing the data in the preregistration, note that I visualize the data for my main analysis in a box plot in figure 3. I consider this to be a minor addition that does not conflict with the intention of the registration.

about how giving behavior is affected by competing charities (Schmitz, 2021), using charities which are similar is a conceivable way of eliminating biases toward charities with highly differentiated profiles. As such, a starting point is to choose charities participants are likely to be familiar with. It is reasonable to assume that charities which are the most donated to are more well known than charities that amass lesser donations. Since charities are to be similar, they need to overlap with regards to their operations. As for the American Red Cross, USA for UNHCR, and Doctors Without Borders USA, these are all tied to disaster relief efforts and are not selective in terms of e.g. age or disease.

Regarding the third requirement, I deemed it important to distance participants from any prior news or false information relating to the charities that may influence their decision. If a participant has prior negative connotations to one of the three charities because of information (true or false), they may already have excluded one of their options. It is safer to choose the US legs of the organizations as participants are more likely to be distanced from news regarding those legs. While donations to both the Swedish and American legs fund operations in other areas of the world, the American population is much bigger than the Swedish one, by extension there is more fact-checked false information to choose from, which is of course another motivation for choosing US legs. Moreover, as the survey was conducted entirely in English there was no need to translate the artifact of false negative information from its original language, thus keeping the artifact intact. A potential weakness of choosing US legs is that they could feel distant to participants, social group is an important factor when donors make allocation decisions (as mentioned in 3.4). This could be a source of noise stemming from indifference between what organization to donate to.

Figure 1 shows the top 10 charitable organizations that Swedish people donated to in 2022. Note that the chosen charities are all in the top 6 most donated to organizations.

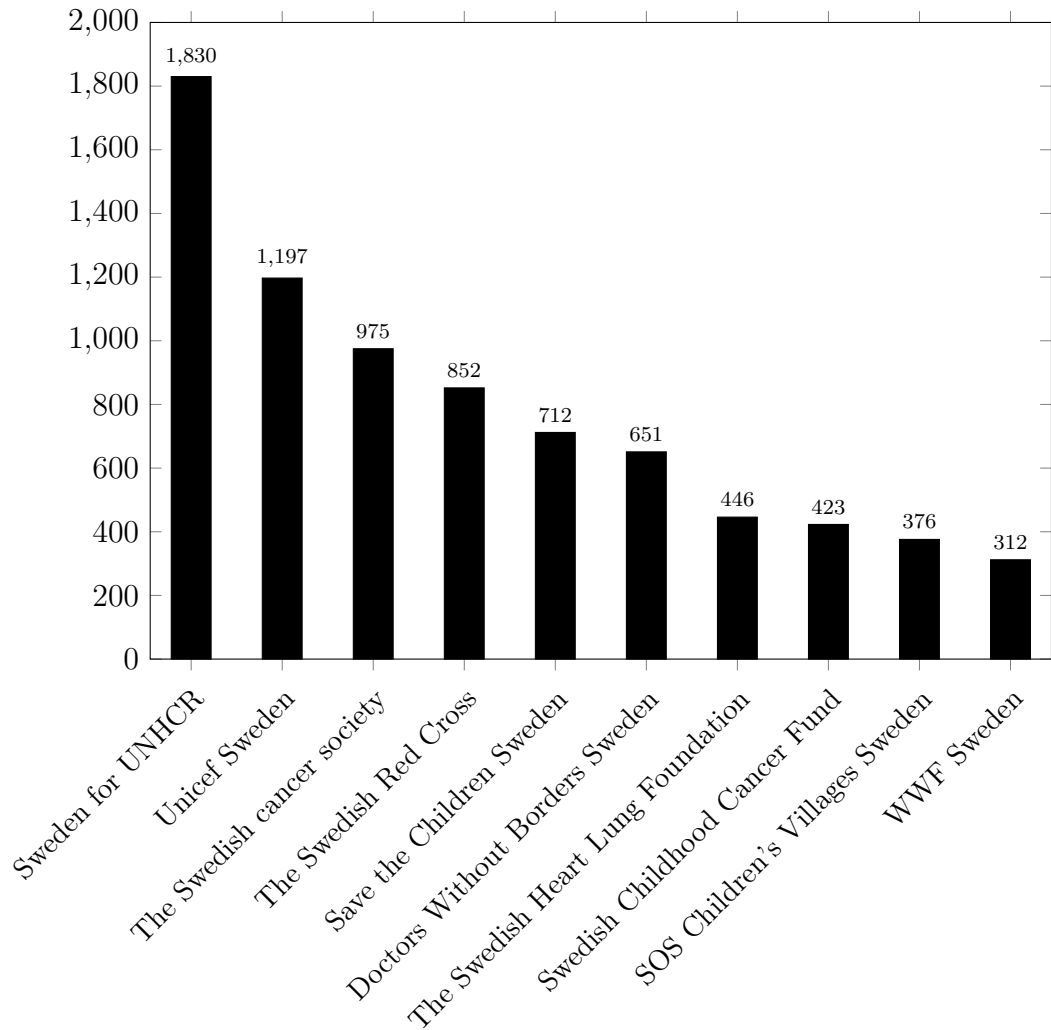


Figure 1: The 10 charities in Sweden that raised the most funds in 2022, the y-axis is denoted in millions of SEK (Giva Sverige, 2022).

5.3 Choosing, Presenting, and Correcting an Artifact of False Negative Information

I deemed it as appropriate to select an existing artifact of false negative information that has been corrected. Not only does it seem counter-intuitive to fabricate a lie if deception is inappropriate, furthermore it may be confusing for participants to be exposed to a piece of text proceeded by correction without any explanation for where the text originated – the experimenter may seem deceptive. There is also a challenge in arguing that a self-produced piece of false information would have

an effect in the real world without relying on extensive pilot-studies or inherently unverifiable qualitative argumentation. When choosing an artifact I was looking for one that has been spread in real life, as it could indicate believability. I was looking for artifacts that could potentially provoke negative emotions and/or factors that could make participants give less to the charity in question in the absence of a subsequent correction.

There are several websites dedicated to fact-checking false information. The California Berkeley Library website provides sources of fact-checking institutions like PolitiFact, FactCheck.org, OpenSecrets.org, Fact Check (Washington Post), Snopes, and Duke Reporters' Lab: Fact Checking (UC Berkeley Library, 2024). I searched these websites for a suitable artifact of false negative information, where the search concerned the three charities chosen in this study. The chosen artifact relates to the American Red Cross and it can be found on PolitiFact. PolitiFact is owned by the Poynter Institute for Media Studies, a non-profit institute that provides fact-checking, media literacy and journalism ethics training. The following is an excerpt from the PolitiFact article, title included (Putterman, 2019):

Facebook post repeats unsubstantiated claims on Red Cross CEO pay, donations

Over its 138-year history, the American Red Cross built a reputation for helping people in times of need. But the organization has come under intense scrutiny in recent years over its emergency response efforts and fiscal management. Questions of propriety have dogged the \$2.7 billion charity following Hurricane Katrina in 2005, the Haitian earthquake in 2010 and Hurricane Sandy in 2012.

A popular social media post draws on some of that scrutiny, while managing to spread its own inaccuracies.

The post, which has been shared on Facebook more than 242,000 times since April 2018, reads: "Marsha J. Evans president and CEO of the American Red Cross. Salary of \$651,957 plus expenses. Only 9 cents of every dollar donated to the American Red Cross goes to helping people in need."

The post was flagged as part of Facebook's efforts to combat false news

and misinformation on its News Feed. (Read more about our partnership with Facebook.)

There are a few issues here.

First, Marsha J. Evans has not been the president and CEO of the Red Cross since she stepped down in 2005. The current CEO — Gail McGovern — does not make quite that much per year. More importantly, there’s no evidence that only 9 cents per dollar goes toward helping people in need. [...]

The most useful part of the article is the claim that only 9 cents per dollar goes to helping people in need. If it were true (which it is not) donations to the American Red Cross would be less efficient than previously thought, which can potentially upset givers (Caviola et al., 2014). For the sake of participants ability to easily and quickly read through the text, and to effectively present participants with clear false information that has been satisfactorily corrected, I deemed the CEO salary to be less useful. While the stated salary is indeed false and the CEO is not even the correct person at the time, the article goes on to state that the current CEO makes a similar amount of money; bonuses, incentives, retirement, benefits, expenses and other allowances included. It is pointed out that the salary is in range for the industry standard and that Gail McGovern has donated more than \$500 000 since becoming CEO in 2008. In order to avoid an unsatisfactory correction and an added amount of text which decreases the ease with which participants can understand it, I excluded CEO salary from the artifact. While this choice may exclude a source of negative emotions that the real-life artifact triggered in people, it increases the internal validity of the experiment as the correction may not be confused with nuance. One can correct something but one can also add nuance to a claim, and these two concepts need to be kept entirely separate.

Table 2, which draws from the artifact of false negative information, the PolitiFact fact-check, and the American Red Cross official website that accounts for how donations are administered (The American Red Cross, 2024), provides the texts that participants in the control and treatment groups read. The treatment group reads every piece of text in the table in falling order, while the control group reads all but the false negative information. See subsection 5.4 for an explanation of why the text is presented to the groups in this manner.

Table 2: Text function and content.

Text Function	Text Content
Neutral text	The American Red Cross states that on average, approximately 90 cents of every dollar donated (90% of a given donation) goes directly to helping people in need. The remaining 10 cents (10%) is used to maintain day-to-day business operations. This includes their enterprise-wide computer, telecommunications, volunteer management, HR, and payroll systems for tens of thousands of employees and hundreds of thousands of volunteers, as well as fundraising, communications, and other support services. In 2018, information regarding the American Red Cross donations circulated in media.
False negative information	2018 information: Only 9 cents of every dollar donated to the American Red Cross goes to helping people in need.
Correction	Later on, the 2018 information was labeled as false information by the media platform where the statement had spread. Among others, fact-checking website PolitiFact, which operates under the Poynter Institute (a non-profit journalism organization) has also deemed the information as false. The American Red Cross has also countered the false information.

The neutral text is initiated by a statement of how much of a dollar donated to the American Red Cross goes directly to helping people in need, and how much that goes to day-to-day business operations. Said operations are elaborated as per the official statement of the American Red Cross official website. Their own description is used as extensively as possible. Some text is altered nevertheless: instead of specific numbers for the amount of employees and volunteers, rougher estimates are used to avoid potential changes in estimates between when the text is gathered and when the experiment is run. This also improves the longevity of the text content,

should it be used in a future experiment. The same reasoning motivates the use of the word "approximately" before "90 cents of every dollar". Percentage metrics for the donations (90% and 10%) are used to familiarize participants with the US currency system, which is necessary if all participants in the treatment group are to understand the meaning of the false negative information that follows. The neutral text is closed off by giving mention to the artifact as having circulated in »media«, "social" is omitted.⁸

While neutral text may not always appear before false negative information in the real world, the text serves as a neutral ground for the treatment- and control groups and assures they get a similar content of text to combat demand effects in the treatment group. Regarding the American Red Cross statement of 90 cents going directly to helping people in need: I considered whether to include some clarification that this is in range for a usual percentage for the large and well known charity organizations (according to their own information), however I discuss in **5.4** that participants may assume similar percentages for other charities. The American Red Cross touts its pride regarding the number on the official website (The American Red Cross, 2024). However, as the PolitiFact article explains, some sources claim that the estimate is too high (Putterman, 2019). As it is not within the scope of this study to evaluate the accounting of these differing estimates, I decided to use the statement from the American Red Cross while clarifying that it is their own statement. Lastly, the choice to use the wording "helping people in need" reflects the wording of the artifact, thus participants in the treatment group are clear on how the claims of the American Red Cross is different from that of the artifact.

The false negative information (the artifact) stands as it was written originally, "Only 9 cents of every dollar donated to the American Red Cross goes to helping people in need". The artifact is bold primarily because it helps treated participants clearly distinguish the false information when the attention check comes (see subsection **5.4**). Additionally, as the neutral text and the false negative information make up most of the text content, I worry that participants in the treatment group who are taking the experiment in whichever environment may not be attentive to the false negative information. As the artifact is particular to the treatment group, it is important that it is apparent to them. I deemed it unlikely that the boldness of the text would affect the credibility to participants but it could enhance the take-up

⁸Iterations of the artifact have appeared across different media (The American Red Cross, 2019).

of the text. Future experiments could make other considerations. Prefacing the artifact with "2018 information:" is done to eliminate any source of confusion to participants regarding what exactly is the false information.

The correction aims to clarify that the artifact is indeed false, where repetitions are used which is one of the aforementioned methods of correcting false information. The correction abides by reality while leaving out the fact that the artifact was a social media post. "False information" is used as the term coincides with how Facebook has worded their disclaimers for false information. A mention of the PolitiFact assessment is made, "false" is their wording for their assessment of the artifact of false negative information. A counterclaim from the American Red Cross is also included. In real life, it is not certain that those who are falsely informed receive a thorough correction. However, as the experiment relies on correction of false negative information, the correction should be thorough enough so that participants do not question the experimenters' intention or the sincerity of the correction. There is of course no mention of the artifact in the correction, as there is nothing in the correction that would give the control group an idea of the contents of the artifact.

Throughout the neutral text, the false negative information and the correction, "information" from 2018 describes the artifact. This allows participants to easily follow the text and works seamlessly when the information is labeled as "false information" in the correction.

5.4 Experimental Design

There has long been unity in the proscription of deception in economic experiments on the grounds of it jeopardising the internal validity and contaminating the participant pool (Bonetti, 1998; Krawczyk, 2019). The perception of deception as an absolute no-go zone has evolved over time. There may indeed exist situations where the benefits of carefully orchestrated deception outweigh the potential costs, and there is no readily available definition of deception that is widely accepted (Bonetti, 1998; Cooper, 2014; Krawczyk, 2019). Nonetheless, there is consensus that deception in the sense that participants do not trust the experimenters' intention should be a last resort in order to test especially important issues that could not be resolved otherwise. The correction of false information is a highly relevant and important

topic, and I argue that it can be done without relying on deception of participants. There is thus no deception of participants in this study. The following figure illustrates the between-subject design that does not rely on deceiving participants, it is designed as the simplest and most internally valid study design: a between-treatment randomized control trial.

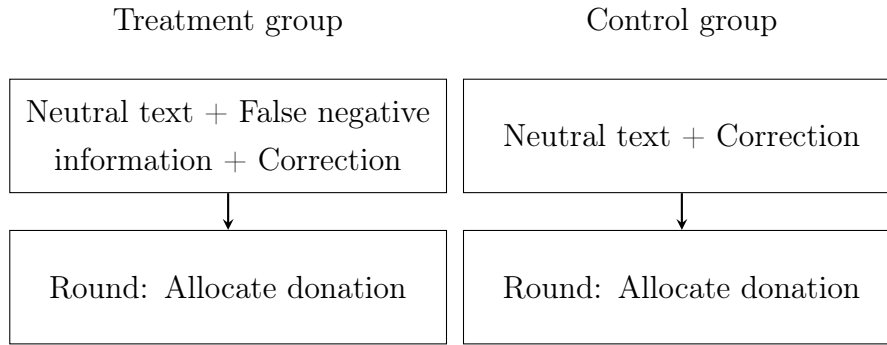


Figure 2: Between-subject experimental design.

Between-subject treatments in a RCT is the classical experiment and the golden standard when it comes to inferring causality from treatment, given randomization of participants into treatment- and control groups. Within-subject treatments, where all participants responses are collected at several rounds, may be less internally valid as anchoring from previous rounds may affect following rounds and demand effects may be at play. However they have a higher potential for statistical power as one does not need to split the participant pool into groups and they can potentially reduce measurement error (Charness et al., 2012). A deceptive design could be done with, for example, a within-subject treatment with three rounds where the participant allocates a donation, then reads the false negative information before the second round of donation allocation and receives the correction before the third round of donation allocation. Such a design may seem to erase concerns of confounding the treatment as there is a clear separation and measurement between the false negative information and the correction. However, the internal validity is damaged by a blatant use of deception. Participants who are deceived become informed of the deception, unlike in experiments where participants walk away unknowing of it. The latter type of experiments, such as in labor economics papers on discrimination, are considered less problematic as participants are not aware they are being deceived (Cooper, 2014).

The key to a successful between-treatment design is to get around the problem of

confounding the false negative information – it needs to be isolated and exclusive to the treatment group, and the correction needs to dispel the false negative information convincingly. The control group cannot read text that is different from what the treatment group reads except for the false negative information. This is challenging because it means that the control group needs to read the correction without having read the false negative information. And the correction needs to be without concrete meaning for the control group. The suggested solution is given by Table 2 under subsection **5.3**.

I need to ensure that all participants understand that the false information has been corrected. Once participants have read their assigned text, there is an attention check to make sure they indeed pay attention. Attention checks are on the same page as the texts, just below them as this makes it exceedingly simple to retrieve the needed information. The attention checks reads as follows in the treatment group:

What is true about the 2018 information regarding the American Red Cross, which is provided in the bold (fat) text above?

- A: The information has been labeled as false information
- B: The information has been confirmed by PolitiFact
- C: The information was published by the New York Times

For participants in the control group, they are met by an attention check which is almost precisely the same but it does not expose the content of the artifact:

What is true about the 2018 information regarding the American Red Cross?

- A: The information has been labeled as false information
- B: The information has been confirmed by PolitiFact
- C: The information was published by the New York Times

Participants who answer incorrectly are not excluded from the main analysis as that may introduce non-random sample attrition but I exclude them in a robustness test. Lastly, when participants make their donation decision they do so in increments of USD and the total donation must sum to 10 USD. The survey cannot be completed until a correct amount is entered. Participants are informed that 5 of them are chosen at random for the donations. There are varying minimum donation restrictions

for the charities, the largest one being 10 USD. I thus considered whether, if a charity does not receive its minimum requirement once the 5 donations have been selected, no money would be donated or if the money should be split between the other charities. However, these choices may alter how participants allocate their donations. In order to emulate a situation without a minimum donation restriction, donations that are larger than 0 USD but lower than the requirement are matched to reach the thresholds.

I introduce a multiplier of 1.5 that is given to donations made to the American Red Cross. In the control group, the multiplier may attract donations and we may see downward difference for the treatment group. Without a multiplier, I fear that if the mean donation for participants in the baseline is small, there may not be sufficient room for participants in the treatment group to give a lower mean donation.

When participants make the choice to allocate 10 USD between three charities they know about the American Red Cross' statements about their overhead costs but not of the other two charities. It is reasonable to believe that if participants behave akin to descriptions of bounded rationality, then they use the assumption that overhead costs are similar as a judgemental heuristic to evaluate payoffs (Kahneman, 2003). According to SEU as described in section 4, participants would calculate that a 1 USD donation to the American Red Cross is to be multiplied by 0.9 to reflect the overhead cost and 1.5 for the multiplier, the product of which is 1.35 USD. If the artifact were true, a 0.09 overhead would yield 0.135 USD.⁹ By way of bounded rationality, fast intuitive thinking rather than slow deliberate calculation suggests participants understand the difference is dramatic but perhaps don't go through the calculation (Kahneman, 2003).

In the main test and in the robustness test (see 5.5), participants that do not complete the survey are excluded. In the robustness test I exclude participants that answer wrong on the attention checks, I also exclude observations that fail to enter their age, nationality or gender.

⁹According to neo-classical rational choice, correct or incorrect beliefs of the payoff of a good determines the preference over that good (Stigler & Becker, 1977). One aspect of being exposed to the artifact is that giving to the American Red Cross could be perceived as less efficient. 1 USD only becomes $1\$ * 0.09 * 1.5 = 0.135\$$ USD in aid rather than $1\$ * 0.9 * 1.5 = 1.35\$$. This could mean a reduction in utility from giving to the American Red Cross as people tend to prefer charities with low overhead costs (Caviola et al., 2014). It should also be noted that utility is not solely derived from monetary payoffs (Stigler & Becker, 1977).

5.5 Research Question, Testing, and Econometric Specification

The research question reads: *Do participants who are exposed to a text with false negative information about a charity (this information is immediately corrected) donate less to this charity than participants who are not exposed to the false negative information?*

I hypothesize that the mean amount donated is smaller for participants who are exposed to a text with false negative information than participants who are not. Thus the null hypothesis and alternative hypothesis are as follows:

H_0 : $\mu_{treated} = \mu_{control}$; There is no statistically significant difference in the mean amount donated to the American Red Cross between the treatment and control groups

H_1 : $\mu_{treated} < \mu_{control}$; The mean amount donated to the American Red Cross in the treatment group is statistically significantly lower than the mean amount donated in the control group.

The main test is a standard t-test, as a robustness check a multiple OLS regression with heteroskedastic robust standard errors is used. This may seem as a simplistic choice, however, Gaussian models on potentially non-normal distributions of outcome could be a better approach than employing complex or novel non-Gaussian methods which are more prone to error and less understood in general, resulting in false positives and inadequate opposition (Knief & Forstmeier, 2021).

With regards to assumptions of a t-test, potential concerns could be heterogeneity and non-normality in the distribution of the outcome variable. Non-normality can be problematic because while a large sample can credibly be assumed to lead to a normal distribution as per the central limit theorem, the distribution could be affected by donations of 0 and 10 USD if participants are driven to donate everything or nothing to the American Red Cross. Even if so, many experimental practitioners regard non-normal distributions as relatively unproblematic for t-tests given a large enough sample size (Knief & Forstmeier, 2021; Lumley et al., 2002).

I considered whether to use a Welch's t-test to account for potential heteroskedasticity at the expense of a slightly lower powered study (many researchers, especially in

the domain of psychology, advocate for the Welch's t-test to be the standard choice, even disregarding a heterogeneity-test such as a Levene's test prior to choosing a type of t-test, since they tend to be underpowered) (Delacre et al., 2017; Rasch et al., 2011; Ruxton, 2006; Wilcox et al., 2013; Zimmerman, 2004). However, given equal sample sizes between groups a standard t-test should be more robust than otherwise to unequal variances. As a robustness check a multiple OLS regression with heteroskedastic robust standard errors serves as additional testing, and should address any unequal variances between the control and treatment group. The econometric specification is as follows:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 Gender_i + \beta_3 Nationality_i + \beta_4 Age_i + \varepsilon_i$$

where the variables and subscripts denote the following:

Y = outcome variable, mean donation to the American Red Cross

β_0 = mean donation to the American Red Cross in the control group

β_1 = difference in mean donation to the American Red Cross between the control and treatment group

T = binary treatment variable: Control (0), Treated (1)

β_2 = difference in mean donation to the American Red Cross between genders

$Gender$ = mean gender as given by 2 values: male (0), female or other (1)

β_3 = difference in mean donation to the American Red Cross between nationalities

$Nationality$ = mean nationality as given by two values: Swedish (0), Other (1)

β_4 = mean change in amount donated to the American Red Cross per increase of year of age

Age = mean age as given by values from 15 to 100

ε = standard errors, heteroskedastic robust

i = subscript that specifies participant number, $i \in N$ (sample size)

In the event of a small acquired sample size, the added controls could account for small numbers of observations of e.g. males or females. Despite randomization, few observations could introduce bias to the results. Metrics on participants gender, nationality, and age also serve as descriptive statistics between the groups, however I argue that successful randomization means that any differences between these

variables between groups is random. There is no additional significance testing.

5.6 Participants and Power Calculations

The participant pool consists of students from the Stockholm School of Economics, and participants that chose to participate were randomized at $N1/N2=1$ level into the treatment and control group. All students at the Stockholm School of Economics, except students from the M.Sc. Economics students 2022 intake are in the participant pool as they may be aware of the purpose of this study and introduce social desirability bias (Nederhof, 1985).¹⁰ A survey method was used and employed through Qualtrics, see **A.2**. In related literature, survey methods are almost exclusively used and it seems as a reasonable choice for this study as well. Student e-mails are handily available and using a survey method allows for a potentially large sample. I did not need to collect participants personal data which would introduce GDPR concerns.

Rather than a lab experiment, online experiments cost less and can be performed faster, allowing larger samples sizes and thus higher powered studies (Arechar et al., 2018; Reips, 2000), and depending on the case external validity can also be improved as one may not be reliant on a limited pool of subjects (Reips, 2000). While this study is limited in its external validity due in part to participants being students at the same university and participants facing no personal monetary trade-offs, internal validity in online and lab experiments have been found as similarly valid (Arechar et al., 2018; Prissé & Jorrat, 2022), except in dictator games where people online donate less compared to in the lab (Prissé & Jorrat, 2022). In this study however, participants do not donate their own money. In addition, surveys can increase openness to the experimental procedure as some components of lab-experiments may fail to be recorded (Reips, 2000).

While surveys open up a possibility for attrition and low response rates, this study is designed to be completed quickly and without cognitively intensive labor. In lab experiments participants may simply ask the experimenter if they do not understand something, this is not possible in a survey method. Instead, I use attention checks to identify whether the text content is understood, and binding the donation amount

¹⁰For instance, participants who are aware of the purpose of this study may want to appear as rational even under anonymity, thus acting as if they are unaffected by the artifact.

to 10 USD should ensure participants do not misunderstand their allocation task. Lastly, participation is anonymous and this could lead to participants filling out multiple surveys. I deem this to be an unlikely scenario. Participants are not monetarily compensated for filling out the survey, and even if participants may feel strongly about some charity, filling out several surveys to block out or favor one charity seems unlikely.

Using the G*power software (Faul et al., 2007), a suitable aim for the sample size was made given a two-tailed independent means t-test with 80% power, Cohen's $d = 0.5$, a participant allocation ratio of $N1/N2=1$, and $p \leq 0.05$ and $p \leq 0.005$. It is two-tailed to allow for an analysis that does not exclude an effect opposite from the alternative hypothesis, which has a direction. In accordance with calls for a redefinition of statistical significance (Benjamin et al., 2017), $p \leq 0.005$ is used to define statistically significant results, this choice minimizes the likelihood of a type 1 error, i.e. a false positive. $p \leq 0.05$ is considered as suggestive evidence. Settling on an appropriate level for what to expect in terms of effect in Cohen's d was challenging due to this being an unprecedented study, typically 0.2, 0.5, and 0.8 are considered as small, medium, and large effect-sizes when comparing the difference between two means. A large anticipated effect-size requires a smaller sample size and vice versa. While the psychological literature has found large effects of persistence of misinformation in the face of debunking (Chan et al., 2017), one should be prudent when approaching economic experiments where monetary outcomes are regarded and original effect-sizes tend to be inflated even when true.

One could perhaps point to the economic literature in priming where effect-sizes have been proven to be small upon replication (Albarracín et al., 2016), and argue to exclude any testing for anything but a small anticipated effect-size. However this experiments' treatment is not using priming, which typically relies on a more subtle treatment that aims to activate mental constructs that change how participants make decisions (Bargh & Chartrand, 2000; Cohn & Maréchal, 2016; Weingarten et al., 2015). In this experiment, the treatment is very blunt and there is no assumption that participants act unaware of what considerations weigh their decisions, as the treatment text is blatant. As recruiting a very large participant pool is not an easy feat I aimed at finding a medium-size effect, if the effect-size is indeed small, which it may be, it cannot be found in this study. Such a study may be reserved for the future. However, the effect could be medium-sized given the psychological literature.

Therefore, $d = 0.5$ is considered and given all choices made, the aim was to recruit 218 participants in total to be able to draw statistically significant evidence but at least 128 are needed in order to find suggestive evidence.

The minimum detectable effect-size (MDE) for this study is reported for 80% statistical power to detect in the hypothesis test after having obtained the standard errors. The MDE that there is 80% power to detect is estimated as the standard error multiplied by 2.8 for the suggestive evidence level of $p \leq 0.05$, and the standard error multiplied by 3.65 for the statistically significant evidence level of $p \leq 0.005$. This calculation can only be made after the data has been collected as standard errors are not known beforehand.

6 Results

6.1 Descriptive Statistics

Table 3: Descriptive statistics.

Responses	
Total	175
Passed attention-check	169
Gender	
Male	103
Female	71
Other	1
Total	175
Nationality	
Swedish	98
Other	77
Total	175
Age	
Mean (years)	27.438
Min (years)	21
Max (years)	53
Total	169

Table 3 shows the descriptive statistics of the Qualtrics survey experiment. 175 responses are recorded, 97% of participants pass the attention check and 97% of participants filled in gender, nationality, and age. The 3% who did not fill in all of the fields all failed to fill in their age. For the main test the full sample of 175 is used. The robustness test uses 165 observations, it includes all who passed the attention check and filled in gender, nationality, and age.

6.2 Main Test Results

Table 4: Two-tailed t-test.

Group	Obs	Mean	Std. Error	95% CI
Control	90	4.744	0.396	[3.958, 5.531]
Treatment	85	5.906	0.371	[5.167, 6.645]
Combined	175	5.309	0.275	[4.766, 5.851]
Difference	-	-1.161	0.544	[-2.236, -.087]

t	-2.134
Pr(T < t)	0.017
Pr(T > t)	0.034
Pr(T > t)	0.983

Table 4 shows the results of the two-tailed t-test. The sample size is, in accordance with the preregistration, large enough to draw suggestive evidence given a medium effect-size (Cohen’s $d = 0.5$). Given the sample size (and reflected in Table 4) there is no evidence of a statistically significant difference in the mean amount donated between the treatment group and the control group. The alternative hypothesis $H_1 : \mu_{treated} < \mu_{control}$ does not hold, as $p = 0.983 > 0.005$. The mean difference is 1.161 USD, which means that participants in the control condition donate less than participants in the treatment condition. There is suggestive evidence in the opposite direction than hypothesized, $\mu_{treated} > \mu_{control}$; $0.05 > p = 0.017 > 0.005$. This evidence is weaker than if I had suggestive evidence in line with my hypothesis, since the suggestive evidence is for a contrary effect. The MDE to detect at 80% power is $2.8 * 0.544 = 1.523$ USD for the suggestive evidence level and $3.65 * 0.544 = 1.986$ USD for the statistically significant evidence level.

Figure 3 visualizes the data of the main test in a box plot, it shows that the median donations in the control and treatment group were 4 USD and 5 USD respectively. Groups differ with regards to the 25th and 50th percentiles, not dramatically.

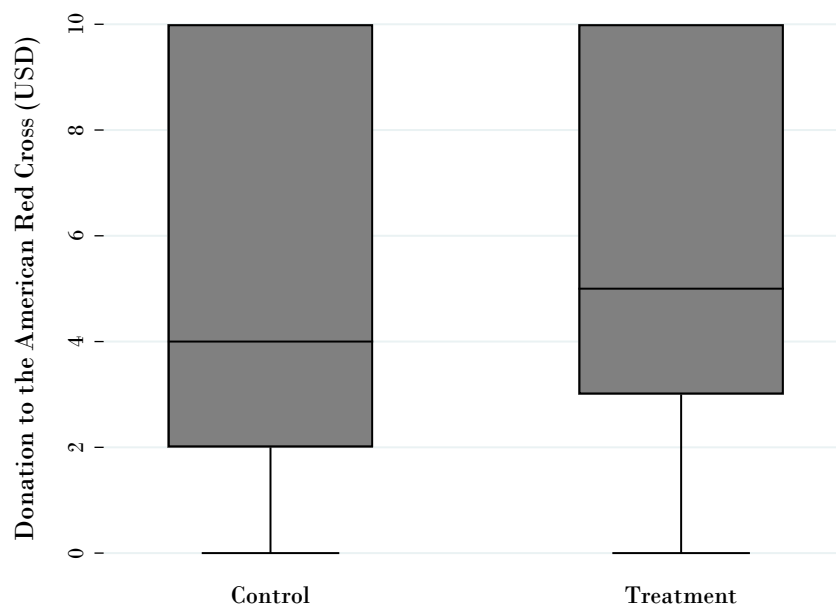


Figure 3: A box plot that represents the median donations (50th percentile) in USD to the American Red Cross with horizontal lines within the boxes, the bottom of the boxes are the first quartile (25th percentile) and the top of the boxes are the third quartile (75th percentile) and the whiskers extend to the highest and lowest values.

6.3 Robustness Test Results

Table 5: Multiple OLS regression.

	Coefficient	Std. Error	t	p
Treated Dummy	0.949	0.578	1.64	0.102
Gender	-0.568	0.574	-0.99	0.325
Nationality	-0.418	0.599	-0.70	0.486
Age	-0.052	0.058	-0.89	0.374
Constant	6.710	1.624	4.13	0.000

Obs.	165
------	-----

Table 5 shows the results of the multiple OLS regression with heteroskedastic robust standard errors and controls. The regression excludes participants that did not pass the attention check and participants that did not fill in gender, nationality, or age. The result shows a 0.949 USD difference in mean donation between the control and treatment group, the treatment group gave more to the American Red Cross than the control group. The results show that the difference is statistically insignificant. The MDE to detect at 80% power is: $2.8 * 0.578 = 1.618$ USD for the suggestive evidence level and $3.65 * 0.578 = 2.110$ USD for the statistically significant evidence level.

7 Discussion

Given 175 participants, the sample size is sufficiently large to draw suggestive evidence but not statistically significant evidence for medium effect-sizes. Given the results I do not find any support for my hypothesis, if anything I find suggestive suggestive evidence that goes in the opposite direction of the hypothesis, which could suggest some System 1 emotive backlash to the artifact. However, this suggestive evidence goes in the opposite direction of my hypothesis, thus this result should be interpreted with caution. Furthermore it is not robust to additional testing. Taken together, in this experiment there is no statistically significant nor suggestive evidence for my hypothesis $H_1: \mu_{treated} < \mu_{control}$. A replication study with higher power may be needed in order to establish an effect-size of similar magnitude to that which is found in this study. Related literature has relied on larger samples, 175 participants is a relatively small sample and quite a bit smaller than my goal of 218 participants, a goal that was already small.

This experiment could bring interest to corrections in economics. There are minuscule differences in the text content that participants were exposed to, there is potential for finding true positives in different settings with different artifacts and/or text content. Regarding the accessibility dimension of the artifact and the correction (discussed in section 4), both may have been highly accessible and thus the correction may have been easily activated. Furthermore the "bad" stimuli of the artifact may not have lingered for long enough due to the artifact text being short and the correction being immediate. The emotional nature and intensity evoked by the artifact is not known, perhaps there are more emotionally provoking artifacts.

Finding suggestive evidence in the wrong direction, albeit not robust, could still be interesting in future studies if one expects to find smaller effect-sizes. After all, this experiment is intended to find medium-sized effects, which may have been overly ambitious. Furthermore, one may speculate that a System 1 emotive backlash to the artifact could vary depending on the organization that the artifact addresses. Perhaps individuals react more negatively to smearing some organizations than others.

While connected literature has reported significant effects, these bear limited relevance for this study. Participants in this study are faced with monetary allocation decisions. Furthermore, the related literature does not widely uphold a $p < 0.005$ significance threshold and aside from the literature relating to corrections of political statements and political beliefs, there seems to be a lot to learn about corrected false information and its implications. Economics as a whole can draw insights from studies on corrected false information but policy implications from this study are limited, in part due to a sample size that is small relative to related literature. Replications and new experiments are in order for this study as well as related literature, preferably employing $p < 0.005$ thresholds. Based on this study I cannot suggest prevention rather than correction because of a CIE. I can also not claim that System 1 thinking is prevalent in the context of corrected false information and I cannot generalize those statements to other contexts. This does not mean that interest in these topics are unwarranted, the abundant literature in psychology that finds a CIE remains and so do the experiments on correction in the political sphere. Further exploration is needed and important given the current state of the world where false information is a serious threat.

Regarding the baseline mean donation to the American Red Cross, which is approximately 5 USD, it could be attributed to demand effects from the text or due to the multiplier. If the latter, participants may give favor to the efficiency of their donations. Still, participants seem to put weight on the other organizations as well. The baseline mean donation together with the remaining allocations yields $4.744 * 1.5 + 5.256 = 12.372$ USD in donations. It makes sense if participants do not consider the charities to be perfect substitutes. There is no way of knowing the extent to which different mechanisms play a role in increasing or decreasing mean donations in the baseline but the use of the multiplier is nevertheless a good way of ensuring that donations are not small or non-existent in the control group although

it could be driving skew in the distribution of the outcome variable (see figure 3), nevertheless I was aware of this possibility, as discussed in 5.5 and I am not severely concerned about non-normality. There is no clear indication of inequity aversion since the American Red Cross receives more than either of the two remaining organizations. Impure altruism is most likely a good theoretical assumption since it does not rule out selfless or selfish reasons to donate, and the extent to which participants lean either way cannot be discerned in this study.

It would have been beneficial to know the percentage of participants who believed in the artifact. Having been widely spread online is not a foolproof proxy for believability, and it is not known how many of all those exposed to the artifact online went on to interact with it. It would also have been interesting to identify the extent to which different possible drivers in the artifact undermine donations to the American Red Cross. While I do not know the extent to which participants accepted the correction, it is reasonable to assume most did. An additional question in the survey (after the allocation so no suspicion is raised ahead of the allocation) could have asked whether they believe in the correction. Instead, I just assumed that the correction is to be believed. The purpose is not to not be deceptive, the correction is repeated in the attention check, literature suggests this strengthens the correction, however the same may be the case for the artifact which is not explicitly stated in the attention check but is the object of attention. Nevertheless the correction ought to be sufficient given the text content that is presented. Utilizing a within-subject experiment could potentially identify whether the artifact is believable as one can track participants allocations along every new piece of information. Still, such a design would rely on deception and thus be problematic.

I do not analyze anchoring effects of the artifact because there is little for any anchor to pertain to. Participants are making the decision to allocate 10 USD across three charities and the intensive margin is considered, this study is not akin to anchoring a high number and finding participants donating more. Moreover, the anchoring potential is difficult to interpret should the numerical content be considered. There are several uses of numbers in both the treated and the control condition and the American Red Cross is mentioned numerous times. There may be some demand effect toward the American Red Cross but no discernible anchoring with regards to the American Red Cross.

Lastly, I want to discuss the exclusion of the artifact in the control condition. Artifi-

cially, the control group receives incomplete information (they are told of some false information but they are not made aware of the content of the false information). This may be argued to create uncertainty in the control condition due to ambiguous information, thus negatively affecting the willingness to give to the American Red Cross. If that is the case, the design of the experiment is not fit for testing the specified hypothesis and both conditions may donate less than a "true" unrecorded baseline, one due to insufficient information and one due to smearing of the charity. Nevertheless, I deem it as an unlikely factor given that no matter the content of the false information, it is refuted quite extensively, this ought to give control participants all the information they need to make a comfortable decision.

8 Conclusion

In this experiment, my results suggest that individuals who are exposed to false negative information, upon correction, do not donate significantly less than they would have if they had never been exposed to the false information in the first place. Furthermore, System 1 thinking cannot be said to dominate participants' allocation decision and there is no basis for a CIE. However, I wished to acquire a larger sample as I acquired 175 participants while I aimed for 218, and I aimed at finding medium-sized effects, a larger sample and an expectation to find smaller effects could be in order. The external validity of this study is limited, for example I have a relatively small sample of students in Sweden at a particular school, a particular artifact and overall text content, and participants do not face trade-offs.

While there is suggestive evidence that treated participants donated more to the American Red Cross, this effect is in the wrong direction than hypothesized and should be interpreted cautiously, and it is not robust to additional testing. There is thus barely support for some affective heuristic backlash to the false negative information. Maybe it would have been reasonable to consider $H_1: \mu_{treated} > \mu_{control}$, however I did not account for it.

Regardless of the direction of an effect, any deviance from $g_{1,initial} = g_{1,corrected}$ would likely be driven by System 1 heuristics since after correction there is no rational reason, no System 2 justification, to deviate. A specific heuristic cannot be discerned

by this study, the experiment does not allow for it. I can merely speculate. Interestingly, my hypothesis assumed that a correction would be insufficient in mitigating false information – this would lead to the policymaking conclusion that prevention is better than correction. If corrections in some cases increase favor to the smeared organization compared to if there was never any exposure to false negative information, then perhaps the negative effects of the mere existence of false information, regardless of believability, would still be enough to merit prevention.

9 Limitations and Future Research

Given limited time, a limited budget, and a limited access to participants, several suggestions can be made for future studies.

Using a representative population sample would make my study all the more important with regards to external validity. Even if I had found a statistically significant result aligned with my hypothesis, there is limited generalizeability, for instance given the student sample of participants.

Participants in this study are anonymous, and could not face trade-offs. The money they were given had to be allocated between the three charities. Allowing for trade-offs between participants and one charity, one may capture participants as less willing to allocate money toward the smeared charity despite of correction. Trade-offs are also more externally valid.

A lack of relevant previous studies made it challenging to assess whether finding a medium effect-size is a realistic expectation. Given a larger sample potential, future research may benefit from expecting smaller effects.

This study was not designed to discern particular heuristics that drive $g_{1,initial} \neq g_{1,corrected}$ and I do not know how believable the artifact is and whether participants believe the correction. I do not know with accuracy what responses the artifact and corrections evoke. Future studies may attempt to discern such questions.

Future studies could try to prolong the exposure to false negative information before correction, which is not logistically easy under no deception. An artifact that is

longer could perhaps allow for some time between the reaction to the artifact and the correction.

Finally, in future endeavors the non-deceptive method developed in this study, which allows for exposure to any artifact in a between-subject setting, can hopefully be used in experiments on charitable giving or on other topics when studying corrected false information.

References

- Adena, M., Alizade, J., Bohner, F., Harke, J., & Mesters, F. (2019). Quality certification for nonprofits, charitable giving, and donor's trust: Experimental evidence. *Journal of Economic Behavior & Organization*, 159, 75–100. <https://doi.org/https://doi.org/10.1016/j.jebo.2019.01.007>
- Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, 13(1). <https://doi.org/10.1007/s13278-023-01028-5>
- Aknin, L. B., Dunn, E. W., & Norton, M. I. (2012). Happiness runs in a circular motion: Evidence for a positive feedback loop between prosocial spending and happiness. *Journal of happiness studies*, 13, 347–355. <https://doi.org/10.1007/s10902-011-9267-5>
- Albarracin, D., Chen, Q., Hepler, J., Yi, J., McAdams, M., & Weingarten, E. (2016). On priming action: Conclusions from a meta-analysis of the behavioral effects of incidentally-presented words. *Current Opinion in Psychology*, 12, 53–57. <https://doi.org/https://doi.org/10.1016/j.copsyc.2016.04.015>
- Alesina, A., Miano, A., & Stantcheva, S. (2020). The polarization of reality. *AEA Papers and Proceedings*, 110, 324–328. <https://doi.org/10.1257/pandp.20201072>
- Ali, A., & Qazi, I. A. (2023). Countering misinformation on social media through educational interventions: Evidence from a randomized experiment in Pakistan. *Journal of Development Economics*, 163, 103108. <https://doi.org/10.1016/j.jdeveco.2023.103108>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100(401), 464–477. Retrieved April 27, 2024, from <http://www.jstor.org/stable/2234133>
- Arcuri, M. C., Gandolfi, G., & Russo, I. (2023). Does fake news impact stock returns? Evidence from US and EU stock markets. *Journal of Economics and Business*, 125-126, 106–130. <https://doi.org/https://doi.org/10.1016/j.jeconbus.2023.106130>

- Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental economics*, 21, 99–131. <https://doi.org/10.1007/s10683-017-9527-2>
- Arumi, A. M., Wooden, R., Johnson, J., & Farkas, S. (2005). The Charitable Impulse: Those who give to charities and those who run them talk about what’s needed to keep the public trust.
- Bachke, M. E., Alfnes, F., & Wik, M. (2017). Information and donations to development aid projects. *Journal of Behavioral and Experimental Economics*, 66, 23–28. <https://doi.org/https://doi.org/10.1016/j.socec.2016.04.015>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>
- Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. *WWW’12 - Proceedings of the 21st Annual Conference on World Wide Web*, 519–528. <https://doi.org/10.1145/2187836.2187907>
- Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle. *Handbook of research methods in social and personality psychology*, 2, 253–285. <https://doi.org/10.1017/CB09780511996481.017>
- Barrera, O., Guriev, S., Henry, E., & Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182, 104123. <https://doi.org/https://doi.org/10.1016/j.jpubeco.2019.104123>
- Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and voluntary sector quarterly*, 40(5), 924–973. <https://doi.org/10.1177/0899764010380927>
- Benjamin, D., Berger, J., Johannesson, M., Nosek, B., Wagenmakers, E.-J., Berk, R., Bollen, K., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C., Clyde, M., Cook, T., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., & Johnson, V. (2017). Redefine Statistical Significance. *Nature Human Behaviour*, 2. <https://doi.org/10.1038/s41562-017-0189-z>
- Bennett, W., & Livingston, S. (2020). *The disinformation age*. Cambridge University Press. <https://doi.org/https://doi.org/10.1017/9781108914628>

- Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology*, 19(3), 377–395. [https://doi.org/https://doi.org/10.1016/S0167-4870\(98\)00012-9](https://doi.org/https://doi.org/10.1016/S0167-4870(98)00012-9)
- Bontridder, N., & Pouillet, Y. (2021). The role of artificial intelligence in disinformation. *Data and Policy*, 3(3). <https://doi.org/10.1017/dap.2021.20>
- Borgloh, S., Dannenberg, A., & Aretz, B. (2013). Small is beautiful—experimental evidence of donors’ preferences for charities. *Economics Letters*, 120(2), 242–244. <https://doi.org/10.1016/j.econlet.2013.04.011>
- Brown, A. L., Meer, J., & Williams, J. F. (2017). Social distance and quality ratings in charity choice. *Journal of Behavioral and Experimental Economics*, 66, 9–15. <https://doi.org/https://doi.org/10.1016/j.socec.2016.04.006>
- Carvalho, C., Klagge, N., & Moench, E. (2011). The persistent effects of a false news shock. *Journal of Empirical Finance*, 18(4), 597–615. <https://doi.org/10.1016/j.jempfin.2011.03.003>
- Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. *Proceedings of the 20th international conference on World wide web*, 675–684. <https://doi.org/10.1145/1963405.1963500>
- Caviola, L., Faulmüller, N., Everett, J. A., Savulescu, J., & Kahane, G. (2014). The evaluability bias in charitable giving: Saving administration costs or saving lives? [PMID: 25279024]. *Judgment and decision making*, 9(4), 303–315.
- Chan, M.-p. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Charities Aid Foundation. (2023). CAF World Giving Index 2023 [Accessed: April 20, 2024]. https://www.cafonline.org/docs/default-source/about-us-research/wgi_report_2023_final.pdf?sfvrsn=402a5447_2
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1–8. <https://doi.org/https://doi.org/10.1016/j.jebo.2011.08.009>
- Choi, S., Choi, C.-Y., & Kim, S. (2023). Tackling misperceptions about immigrants with fact-checking interventions: A randomized survey experiment. *Labour Economics*, 84, 102428. <https://doi.org/10.1016/j.labeco.2023.102428>

- Clarke, J., Chen, H., Du, D., & Hu, Y. J. (2020). Fake news, investor attention, and market reaction. *Information Systems Research*, 32(1), 35–52. <https://doi.org/10.1287/isre.2019.0910>
- Cohn, A., & Maréchal, M. A. (2016). Priming in economics. *Current Opinion in Psychology*, 12, 17–21. <https://doi.org/https://doi.org/10.1016/j.copsyc.2016.04.019>
- Colomina, C., Margalef, H. S., Youngs, R., & Jones, K. (2021). The impact of disinformation on democratic processes and human rights in the world. *Brussels: European Parliament*, 1–19. <https://doi.org/doi:10.2861/59161>
- Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4. <https://doi.org/10.1002/pra2.2015.145052010082>
- Cooper, D. J. (2014). A Note on Deception in Economic Experiments. *Journal of Wine Economics*, 9(2), 111–114. <https://doi.org/10.1017/jwe.2014.18>
- Cryder, C., & Loewenstein, G. (2011). The critical link between tangibility and generosity. In D. M. Oppenheimer & C. Y. Olivola (Eds.), *The science of giving: Experimental approaches to the study of charity* (pp. 237–252). <https://doi.org/10.4324/9780203865972>
- Cryder, C. E., Loewenstein, G., & Scheines, R. (2013). The donor is in the details. *Organizational Behavior and Human Decision Processes*, 120(1), 15–23. <https://doi.org/10.1016/j.obhdp.2012.08.002>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch’s t-test instead of student’s t-test. *International Review of Social Psychology*, 30(1), 92–101. <https://doi.org/10.5334/irsp.82>
- Diamond, W. D., & Kashyap, R. K. (1997). Extending Models of Prosocial Behavior to Explain University Alumni Contributions. *Journal of Applied Social Psychology*, 27(10), 915–928. <https://doi.org/10.1111/j.1559-1816.1997.tb00278.x>
- Duncan, B. (2004). A theory of impact philanthropy. *Journal of public Economics*, 88(9-10), 2159–2180. [https://doi.org/10.1016/S0047-2727\(03\)00037-9](https://doi.org/10.1016/S0047-2727(03)00037-9)
- Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending Money on Others Promotes Happiness. *Science*, 319(5870), 1687–1688. <https://doi.org/10.1126/science.1150952>
- Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological

- drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Edelman. (2023a). 2023 Edelman Trust Barometer Global Report. *Edelman Trust Barometer Report*. <https://www.edelman.com/sites/g/files/aatuss191/files/2023-03/2023%20Edelman%20Trust%20Barometer%20Global%20Report%20FINAL.pdf>
- Edelman. (2023b). 2023 Edelman Trust Barometer Sweden Report. *Edelman Trust Barometer Report*. https://www.edelman.com/sites/g/files/aatuss191/files/2023-05/2023%20Edelman%20Trust%20Barometer_Sweden%20Report.pdf
- Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *Review of Economic Studies*, 83(2), 587–628. <https://doi.org/10.1093/restud/rdv051>
- Fallis, D. (2015). What Is Disinformation? *Library Trends*, 63(3), 401–426. <https://doi.org/10.1353/lib.2015.0014>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/bf03193146>
- Fedyk, A., & Hodson, J. (2023). When can the market identify old news? *Journal of Financial Economics*, 149(1), 92–113. <https://doi.org/https://doi.org/10.1016/j.jfineco.2023.04.008>
- Fehr, E., & Hoff, K. (2011). Introduction: Tastes, castes and culture: The influence of society on preferences. *The Economic Journal*, 121(556), F396–F412. <http://www.jstor.org/stable/41301349>
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3), 817–868. <https://doi.org/10.1162/003355399556151>
- Fennell, L. A. (2022). Optional Price Discrimination. *Texas AM Law Review*, 10. <http://dx.doi.org/10.2139/ssrn.4158024>
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1), 298–320. <https://doi.org/10.1093/poq/nfw006>

- Fragale, A. R., & Heath, C. (2004). Evolving informational credentials: The (mis) attribution of believable facts to credible sources. *Personality and social psychology bulletin*, 30(2), 225–236. <https://doi.org/10.1177/0146167203259933>
- Funke, M., Schularick, M., & Trebesch, C. (2016). Going to extremes: Politics after financial crises, 1870–2014. *European Economic Review*, 88, 227–260. <https://doi.org/10.1016/j.euroecorev.2016.03.006>
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42. <https://doi.org/https://doi.org/10.1016/j.socrec.2010.10.008>
- Gaskin, K. (1999). Blurred vision: Public trust in charities. *International Journal of Nonprofit and Voluntary Sector Marketing*, 4(2), 163–178. <https://doi.org/https://doi.org/10.1002/nvsm.66>
- Gelfert, A. (2018). Fake news: A definition. *Informal logic*, 38(1), 84–117. <https://doi.org/https://doi.org/10.22329/il.v38i1.5068>
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process.
- Giva Sverige. (2020). Givandet i siffror 2020 [Accessed: February 19, 2024]. https://wwwfriise.cdn.triggerfish.cloud/uploads/2021/05/givandet-i-siffror-2020_giva-sverige_210528.pdf
- Giva Sverige. (2021). Givandet i siffror 2021 [Accessed: February 19, 2024]. <https://wwwfriise.cdn.triggerfish.cloud/uploads/2022/05/givandet-i-siffror-2021.pdf>
- Giva Sverige. (2022). Givandet i siffror 2022 [Accessed: February 19, 2024]. <https://wwwfriise.cdn.triggerfish.cloud/uploads/2023/05/givandet-i-siffror-2022.pdf>
- Grigorieff, A., Roth, C., & Ubfal, D. (2020). Does information change attitudes toward immigrants? *Demography*, 57(3), 1117–1143. <https://doi.org/10.1007/s13524-020-00882-8>
- Gupta, G., Raja, K., Gupta, M., Jan, T., Whiteside, S. T., & Prasad, M. (2024). A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods. *Electronics (Switzerland)*, 13(1). <https://doi.org/10.3390/electronics13010095>
- Guriev, S., Henry, E., Marquis, T., & Zhuravskaya, E. (2023). Curtailing False News, Amplifying Truth. <http://dx.doi.org/10.2139/ssrn.4616553>
- Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence.

- British Journal of Management*, 33(3), 1238–1253. <https://doi.org/10.1111/1467-8551.12554>
- Harashima, T. (2023). Economic Rents Extracted in the Process of Consumption. <https://ideas.repec.org/p/pramprapa/119012.html>
- Henry, E., Zhuravskaya, E., & Guriev, S. (2022). Checking and Sharing Alt-Facts. *American Economic Journal: Economic Policy*, 14(3), 55–86. <https://doi.org/10.1257/pol.20210037>
- Hopkins, D. J., Sides, J., & Citrin, J. (2019). The muted consequences of correct information about immigration. *The Journal of Politics*, 81(1), 315–320. <https://doi.org/10.1086/699914>
- Ibanez, L., Moureau, N., & Roussel, S. (2017). How do incidental emotions impact pro-environmental behavior? Evidence from the dictator game. *Journal of behavioral and experimental economics*, 66, 150–155. <https://doi.org/10.1016/j.socec.2016.04.003>
- Ifcher, J., & Zarghamee, H. (2020). Behavioral economic phenomena in decision-making for others. *Journal of Economic Psychology*, 77. <https://doi.org/10.1016/j.joep.2019.06.003>
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of experimental psychology: Learning, memory, and cognition*, 20(6), 1420–1436. <https://doi.org/10.1037/0278-7393.20.6.1420>
- Jost, P. J., Pünder, J., & Schulze-Lohoff, I. (2020). Fake news - Does perception matter more than the truth? *Journal of Behavioral and Experimental Economics*, 85. <https://doi.org/https://doi.org/10.1016/j.socec.2020.101513>
- Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, 93(5), 1449–1475. Retrieved April 27, 2024, from <http://www.jstor.org/stable/3132137>
- Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61, 32–44. <https://doi.org/10.1016/j.cogsys.2019.12.005>
- Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, 23(5), 1301–1326. <https://doi.org/https://doi.org/10.1177/1461444820959296>

- Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, 53(6), 2576–2590. <https://doi.org/10.3758/s13428-021-01587-5>
- Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools. *Psychological Science in the Public Interest*, 21(3), 103–156. <https://doi.org/10.1177/1529100620946707>
- Krawczyk, M. (2019). What should be regarded as deception in experimental economics? evidence from a survey of researchers and subjects. *Journal of Behavioral and Experimental Economics*, 79, 110–118. <https://doi.org/https://doi.org/10.1016/j.socec.2019.01.008>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the News That’s Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science*, 9(1), 104–117. <https://doi.org/10.1017/XPS.2020.37>
- Ladd, J. M. (2010). The neglected power of elite opinion leadership to produce antipathy toward the news media: Evidence from a survey experiment. *Political Behavior*, 32, 29–50. <https://doi.org/10.1007/s11109-009-9097-x>
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Li, H., & Sakamoto, Y. (2014). Social impacts in social media: An examination of perceived truthfulness and sharing of information. *Computers in Human Behavior*, 41, 278–287. <https://doi.org/10.1016/j.chb.2014.08.009>
- Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The Importance of the Normality Assumption in Large Public Health Data Sets. *Annual Review of Public Health*, 23(1), 151–169. <https://doi.org/10.1146/annurev.publhealth.23.100901.140546>
- Mathur, A. (1996). Older adults’ motivations for gift giving to charitable organizations: An exchange theory perspective. *Psychology & Marketing*, 13(1), 107–

123. [https://doi.org/10.1002/\(SICI\)1520-6793\(199601\)13:1%3C107::AID-MAR6%3E3.0.CO;2-K](https://doi.org/10.1002/(SICI)1520-6793(199601)13:1%3C107::AID-MAR6%3E3.0.CO;2-K)
- Merchant, A., Ford, J. B., & Sargeant, A. (2010). Charitable organizations' storytelling influence on donors' emotions and intentions. *Journal of Business Research*, 63(7), 754–762. <https://doi.org/10.1016/j.jbusres.2009.05.013>
- Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility. *Personality and Social Psychology Bulletin*, 26(9), 1142–1150. <https://doi.org/10.1177/01461672002611010>
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European journal of social psychology*, 15(3), 263–280. <https://doi.org/10.1002/ejsp.2420150303>
- Niesta Kayser, D., Greitemeyer, T., Fischer, P., & Frey, D. (2010). Why mood affects help giving, but not moral courage: Comparing two types of prosocial behaviour. *European Journal of Social Psychology*, 40(7), 1136–1157. <https://doi.org/https://doi.org/10.1002/ejsp.717>
- Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences of the United States of America*, 119(8). <https://doi.org/10.1073/pnas.2120481119>
- Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2020). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political behavior*, 42, 939–960. <https://doi.org/10.1007/s11109-019-09528-x>
- OECD. (2022). Misinformation and disinformation. (21). <https://doi.org/https://doi.org/https://doi.org/10.1787/b7709d4f-en>
- Oppenheimer, D. M., & Olivola, C. Y. (2011). *The science of giving: Experimental approaches to the study of charity*. Psychology Press.
- Palmertz, B., Weissmann, M., Nilsson, N., & Engvall, J. (2024). Building Resilience and Psychological Defence: An analytical framework for countering hybrid threats and foreign influence and interference. *Psychological Defence Research Institute*.
- Pantserov, K. A. (2020). The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability. *Advanced*

- Sciences and Technologies for Security Applications*, 37–55. https://doi.org/10.1007/978-3-030-35746-7_3
- Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management science*, 66(11), 4944–4957. <https://doi.org/10.1287/mnsc.2019.3478>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general*, 147(12), 1865–1880. <https://doi.org/10.1037/xge0000465>
- Polman, E., Effron, D. A., & Thomas, M. R. (2018). Other People’s Money: Money’s Perceived Purchasing Power Is Smaller for Others Than for the Self. *Journal of Consumer Research*, 45(1), 109–125. <https://doi.org/10.1093/jcr/ucx119>
- Prissé, B., & Jorrat, D. (2022). Lab vs online experiments: No differences. *Journal of Behavioral and Experimental Economics*, 100, 101910. <https://doi.org/10.1016/j.socec.2022.101910>
- Putterman, S. (2019). Facebook post repeats unsubstantiated claims on Red Cross CEO pay, donations [Accessed: February 19, 2024]. <https://www.politifact.com/factchecks/2019/feb/08/facebook-posts/refurbished-meme-unsubstantiated-red-cross-ceo-pay/>
- Radley, A., & Kennedy, M. (1992). Reflections upon charitable giving: A comparison of individuals from business, ‘manual’ and professional backgrounds. *Journal of Community & Applied Social Psychology*, 2(2), 113–129. <https://doi.org/10.1002/casp.2450020208>
- Rasch, D., Kubinger, K. D., & Moder, K. (2011). The two-sample t test: Pre-testing its assumptions does not pay off. *Statistical Papers*, 52(1), 219–231. <https://doi.org/10.1007/s00362-009-0224-x>
- Reips, U.-D. (2000). The Web Experiment method: Advantages, Disadvantages, and Solutions. In *Psychological experiments on the internet* (pp. 89–117). Elsevier. <https://doi.org/10.5167/uzh-19760>
- Rothstein, B., & Holmberg, S. (2022). Social tillit i höglitarlandet sverige. *SOM-Institutets Temaserie 2022*, 1.
- Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student’s t-test and the Mann-Whitney U test. *Behavioral Ecology*, 17(4), 688–690. <https://doi.org/10.1093/beheco/ark016>

- Schmitz, J. (2021). Is charitable giving a zero-sum game? The effect of competition between charities on giving behavior. *Management Science*, 67(10), 6333–6349. <https://doi.org/10.1287/mnsc.2020.3809>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Simon, H. A. (1990). Bounded Rationality. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Utility and probability* (pp. 15–18). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-20568-4_5
- Small, D. A., & Verrochi, N. M. (2009). The face of need: Facial emotion expression on charity advertisements. *Journal of Marketing Research*, 46(6), 777–787. <https://doi.org/10.1509/jmkr.46.6.777>
- Smith, J. R., & McSweeney, A. (2007). Charitable giving: The effectiveness of a revised theory of planned behaviour model in predicting donating intentions and behaviour. *Journal of Community & Applied Social Psychology*, 17(5), 363–386. [10.1002/casp.906](https://doi.org/10.1002/casp.906)
- SolAbility Sustainable Intelligence. (2023). The Global Sustainable Competitiveness Report 2023.
- Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business information review*, 34(3), 150–160. <https://doi.org/10.1177/0266382117722446>
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5), 645–726. <https://doi.org/10.1017/S0140525X00003435>
- Stigler, G. J., & Becker, G. S. (1977). De Gustibus Non Est Disputandum. *The American Economic Review*, 67(2), 76–90. Retrieved April 8, 2024, from <http://www.jstor.org/stable/1807222>
- Swedish Ministry of Finance. (2023). Budgetpropositionen för 2024 (proposition 2023/24:1). <https://data.riksdagen.se/fil/FEBA71E5-88EF-45A8-930E-495B59540B12>
- Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society open science*, 4(3), 160802. <https://doi.org/10.1098/rsos.160802>

- Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News”: A typology of scholarly definitions. *Digital journalism*, 6(2), 137–153. <https://doi.org/10.1080/21670811.2017.1360143>
- The American Red Cross. (2019). Red Cross Statement on Inaccurate Viral Email on Charity CEO Pay [Accessed: May 10, 2024]. <https://www.redcross.org/about-us/news-and-events/press-release/red-cross-statement-on-inaccurate-viral-email-on-charity-ceo-pay.html>
- The American Red Cross. (2024). How the Red Cross Spends Your Donations [Accessed: February 19, 2024]. <https://www.redcross.org/donations/how-the-red-cross-spends-your-donations.html>
- Titmuss, R. (2018). *The gift relationship: From human blood to social policy*. Policy Press. <https://doi.org/10.1332/policypress/9781447349570.001.0001>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- UC Berkeley Library. (2024). Real news/fake news: Fact checkers [Accessed: February 19, 2024]. <https://guides.lib.berkeley.edu/c.php?g=620677&p=4333407>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media and Society*, 6(1). <https://doi.org/10.1177/2056305120903408>
- Van Duyn, E., & Collier, J. (2019). Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication and Society*, 22(1), 29–48. <https://doi.org/10.1080/15205436.2018.1511807>
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication monographs*, 85(3), 423–441. <https://doi.org/10.1080/03637751.2018.1467564>
- Walter, N., & Tukachinsky, R. (2020). A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It? *Communication Research*, 47(2), 155–177. <https://doi.org/10.1177/0093650219854600>
- Weingarten, E., Chen, Q., Mcadams, M., Yi, J., Hepler, J., & Albarracin, D. (2015). From Primed Concepts to Action: A Meta-Analysis of the Behavioral Effects

- of Incidentally Presented Words. *Psychological Bulletin*, 142. <https://doi.org/10.1037/bul0000030>
- Weisbuch, M., & Ambady, N. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of personality and social psychology*, 95(5), 1063–1079. <https://doi.org/10.1037/a0011993>
- Wilcox, R., Granger, D., & Clark, F. (2013). Modern Robust Statistical Methods: Basics with Illustrations Using Psychobiological Data. *Universal Journal of Psychology*, 1, 21–31. <https://doi.org/10.13189/ujp.2013.010201>
- Zagefka, H., & James, T. (2015). The psychology of charitable donations to disaster victims and beyond. *Social Issues and Policy Review*, 9(1), 155–192. <https://doi.org/https://doi.org/10.1111/sipr.12013>
- Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1), 173–181. <https://doi.org/10.1348/000711004849222>

A Appendix

A.1 Preregistration



A non-deceptive study on false information and charity (#164337)

Created: 02/29/2024 11:03 PM (PT)

Public: 05/29/2024 11:43 AM (PT)

Author(s)

Evan Shokor Qasim (Stockholm School of Economics) - 42415@student.hhs.se

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

The research question reads:

Will participants who are exposed to a text with false negative information about a charity (this information is immediately corrected) donate less to this charity than participants who are not exposed to the false negative information?

I hypothesize that the mean amount donated will be smaller for participants who are exposed to a text with false negative information than participants who are not. Thus the null hypothesis and alternative hypothesis are as follows:

H0: $\mu_{\text{treated}} = \mu_{\text{control}}$; There is no statistically significant difference in the mean amount donated to the American Red Cross between the treatment and control groups.

H1: $\mu_{\text{treated}} < \mu_{\text{control}}$; The mean amount donated to the American Red Cross in the treatment group is statistically significantly lower than the mean amount donated in the control group.

3) Describe the key dependent variable(s) specifying how they will be measured.

Mean amount donated (in dollars) to the American Red Cross in the control group and in the treatment group. This will be measured through an incentivized survey where participants choose how to allocate a given set of money (not their own money) between three different charities: Doctors Without Borders USA, the American Red Cross, and USA for UNHCR. Donations to the American Red Cross are given a multiplier of 1.5, the other charities are not given this multiplier.

4) How many and which conditions will participants be assigned to?

There are two conditions and participants will be randomized to one of them: the treatment condition for the treatment group (participants in this condition read a neutral text, a text with false information, and a text disproving the false information) and the control condition for the control group (participants in this condition read a neutral text and a text disproving the false information)

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

The primary test is a test of the difference between the two mean amounts donated to the American Red Cross, using a two-tailed independent t-test.

Following Benjamin et al. (2018), I define suggestive evidence as $p \leq 0.05$ and statistically significant evidence as $p \leq 0.005$. I will report the minimum detectable effect size (MDE) that I had 80% power to detect, estimated as the standard error multiplied by 2.8 for the suggestive evidence level of $p \leq 0.05$, and the standard error multiplied by 3.65 for the statistically significant evidence level of $p \leq 0.005$

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Unfinished surveys are excluded; the consent form must be agreed to before proceeding; gender, nationality, and age are not compulsory to fill in, missing observations of these values will not be excluded in the main analysis but they will be dropped in the robustness test regression. Age is restricted between 15 and 80. Surveys where participants answer incorrectly on the attention check are not excluded from the main test as that could potentially lead to non-random attrition between groups, but in a robustness test I will exclude those participants.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

The G*Power software is used to calculate the required sample size given the specified parameters: Two tails, Cohen's $d = 0.5$ (medium sized effect anticipated), $p \leq 0.005$, 80% power, allocation ratio $N1/N2=1$. Thus my aim is to recruit 218 (at least, in case of sample attrition) participants for statistically significant evidence. For suggestive evidence, $p \leq 0.05$, 128 participants at least are required. I will stop collecting observations at exactly 12:00 AM UTC +02:00 April 30 in order to be finished before the deadline of this thesis. If I get a smaller sample size than this I will anyway report the results from my planned analysis and add the caveat that the sample is smaller than intended.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

In a robustness check, a multivariate OLS regression with heteroskedastic robust standard errors will be run. As such, the econometric specification is as follows:

$$Y_i = \beta_0 + \beta_1 * T_i + \beta_2 * Gender_i + \beta_3 * Nationality_i + \beta_4 * Age_i + \epsilon_i$$



Y = outcome variable, mean donation to the American Red Cross
 β_0 = mean donation to the American Red Cross in the control group
 β_1 = difference in mean donation to the American Red Cross between the control and treatment group
 T = binary treatment variable: Control (0), Treated (1)
 β_2 = difference in mean donation to the American Red Cross between genders
Gender = mean gender as given by 2 values: man (0), woman and other (1)
 β_3 = difference in mean donation to the American Red Cross between nationalities
Nationality = mean nationality as given by two values: Swedish (0), Other (1)
 β_4 = mean change in amount donated to the American Red Cross per increase of year of age
Age = mean age as given by values from 15 to 80
 ϵ = standard errors, heteroskedastic robust
 i = subscript that specifies participant number, $i \in N$ (sample size)

I define suggestive evidence as $p \leq 0.05$ and statistically significant evidence as $p \leq 0.005$.

I will also have a attention checks for both groups,:

Treatment:

What is true about the 2018 information regarding the American Red Cross, which is provided in the bold (fat) text above?

A: The information has been labeled as false information

B: The information has been confirmed by PolitiFact

C: The information was published by the New York Times

Control:

What is true about the 2018 information regarding the American Red Cross?

A: The information has been labeled as false information

B: The information has been confirmed by PolitiFact

C: The information was published by the New York Times

A.2 Survey

2024-03-01 11:05

Qualtrics Survey Software

Consent form

Hi! This survey is part of a MSc thesis which examines how different types of information affect donations. You will: (1) read a short text; (2) be given 10 US dollars to allocate **[you do not donate your own money]** between three charities; (3) fill in your gender, nationality and age. There is no risk associated with participating, your identity and your responses are completely anonymous. The survey takes approximately 5 minutes to complete.

Regarding donations: All participants are given 10 dollars that they are asked to allocate between 3 charities in any way they like. 5 participants will be randomly chosen, and the donations that these participants made will be implemented. Once implemented, if a charity receives more than 0 dollars in total but less than the minimum donation restriction on the charity website, I will add the additional required amount to make sure the donation reaches the charity.

Participation is voluntary, you can withdraw from the survey at any given moment. The researcher of this study is Evan Shokor Qasim, student at the Stockholm School of Economics. For questions, contact 42415@student.hhs.se

If you agree to these terms, click "I agree" and go to the next page to start the first step of the survey.

- ☐ I agree
- ☐ I do not agree

T Text&Takeup

The American Red Cross states that on average, approximately 90 cents of every dollar donated (90% of a given donation) goes directly to helping people in need. The remaining 10 cents (10%) is used to maintain day-to-day business operations. This includes their enterprise-wide computer, telecommunications, volunteer management, HR, and payroll systems for tens of thousands of employees and hundreds of thousands of volunteers, as well as fundraising, communications, and other support services. In 2018, information regarding the American Red Cross donations circulated in media.

2018 information: **Only 9 cents of every dollar donated to the American Red Cross goes to helping people in need.**

Later on, the 2018 information was labeled as false information by the media platform where the statement

had spread. Among others, fact-checking website PolitiFact, which operates under the Poynter Institute (a non-profit journalism organization) has also deemed the information as false. The American Red Cross has also countered the false information.

What is true about the 2018 information regarding the American Red Cross, which is provided in the bold (fat) text above?

- ☐ The information has been labeled as false information
- ☐ The information has been confirmed by PolitiFact
- ☐ The information was published by the New York Times

C Text&Takeup

The American Red Cross states that on average, approximately 90 cents of every dollar donated (90% of a given donation) goes directly to helping people in need. The remaining 10 cents (10%) is used to maintain day-to-day business operations. This includes their enterprise-wide computer, telecommunications, volunteer management, HR, and payroll systems for tens of thousands of employees and hundreds of thousands of volunteers, as well as fundraising, communications, and

other support services. In 2018, information regarding the American Red Cross donations circulated in media.

Later on, the 2018 information was labeled as false information by the media platform where the statement had spread. Among others, fact-checking website PolitiFact, which operates under the Poynter Institute (a non-profit journalism organization) has also deemed the information as false. The American Red Cross has also countered the false information.

What is true about the 2018 information regarding the American Red Cross?

- ☐ The information has been labeled as false information
- ☐ The information has been confirmed by PolitiFact
- ☐ The information was published by the New York Times

Donation

Donate 10 US dollars (which is roughly equal to 100 SEK). Your decision must add up to 10, no more and no less.

Money donated to the American Red Cross will be

multiplied by 1.5, meaning that 50% more than the chosen allocation will be donated.



Age, gender, nationality, field

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

What is your nationality?

- ☐ Swedish
- ☐ Other

Choose your age

Powered by Qualtrics

A.2.1 Survey Distribution

(March 1, 2024)

Survey: allocate money (not your own) to charity

Dear students,

My name is Evan Shokor Qasim and I am writing my master's thesis on how different types of information affect donations to charity. I would be deeply grateful if you took 5 minutes of your time to participate in my survey. It is completely anonymous and very easy to complete. You will be given 10 dollars to allocate between three charities (**you will not donate your own money**), five participants will be randomly chosen and their donations will be implemented.

Data protection: No identifying information will be collected in this survey, participants remain completely anonymous.

The survey can be found here:

https://sse.qualtrics.com/jfe/form/SV_8cRv9cT2jeWpwFg

For questions you are very welcome to contact me: 42415@student.hhs.se

All the best,

Evan

(March 18, 2024)

Reminder: Survey: allocate money (not your own) to charity

Dear students,

My name is Evan Shokor Qasim and I am writing my master's thesis on how different types of information affect donations to charity. I would be deeply grateful if you took 5 minutes of your time to participate in my survey if you have not already done so. It is completely anonymous and very easy to complete. You will be given 10 dollars to allocate between three charities (**you will not donate your own money**), five participants will be randomly chosen and their donations will be implemented.

Data protection: No identifying information will be collected in this survey, participants remain completely anonymous.

The survey can be found here:

https://sse.qualtrics.com/jfe/form/SV_8cRv9cT2jeWpwFg

For questions you are very welcome to contact me: 42415@student.hhs.se

All the best,

Evan

(April 22, 2024)

Final Reminder: Survey: allocate money (not your own) to charity

Dear students,

My name is Evan Shokor Qasim and I am writing my master's thesis on how different types of information affect donations to charity. I would be deeply grateful if you took 5 minutes of your time to participate in my survey if you have not already done so. It is completely anonymous and very easy to complete. You will be given 10 dollars to allocate between three charities (**you will not donate your own money**), five participants will be randomly chosen and their donations will be implemented.

Data protection: No identifying information will be collected in this survey, participants remain completely anonymous.

The survey can be found here:

https://sse.qualtrics.com/jfe/form/SV_8cRv9cT2jeWpwFg

For questions you are very welcome to contact me: 42415@student.hhs.se

All the best,

Evan

A.3 Impure Altruism in Charitable Giving

The model of impure altruism presented by Andreoni (1990) assumes that individuals derive utility from selfish and altruistic reasons. He assumes people are endowed with wealth w_i to allocate between a private good x_i and their gift to a public good g_i , one of each for simplicity. $G = \sum_{i=1}^n g_i$ is the total amount of the public good, n denotes number of individuals. Gifts of everyone except individual i is given by $G_{-i} = \sum_{j \neq i} g_j$. He allows the private good to be converted to a public good through linear technology such that both may be denoted in USD. Through a utility function where g_i enters two times: (1) as public good G to capture pure altruism, (2) as private good g_i to capture pure egoism, properties being independent of each other, and subject to $x_i + g_i = w_i$ and under assumptions of U_i as strictly quasi-concave due to the preferences of the utility function being convex, he derives a function of giving as such:

$$g_i = f_i(w_i + G_{-i}, G_{-i}) - G_{-i}$$

The first argument in f_i is brought from the public goods dimension in the utility function and we call the derivative of f_i w.r.t this argument f_{ia} the marginal propensity to donate for altruistic reasons. The second argument in f_i is brought from the private goods dimension in the utility function and we call the derivative of f_i w.r.t. this argument f_{ie} the marginal propensity to donate for egoistic reasons.

I follow the model almost exactly and the implications are the same, the difference is that in my model I assume $\bar{x}_i$ as private consumption is not possible for my participants, I use a vector of public goods $\mathbf{g}_i = (g_{1i}, g_{2i}, g_{3i})$ containing different quantities of each of the options, wealth w_i is equal to 10 USD and may only be allocated to the public goods, and I do not allow for conversion through linear technology of the private good to a public good.

A.4 Implementing Donations

10 days after data collection, on May 10, 2024 I used <https://www.random.org/integers/> to generate 5 integers between the values of 1 to 175 (the number of participants) which resulted in 92, 44, 88, 156, and 174. Allocating 10 USD between Doctors Without Borders USA, the American Red Cross, and USA for UNHCR, the participants whose numbers were drawn chose (5,1,4), (2,8,0), (0,10,0), (4,2,4), and (4,4,2). Given the 1.5 multiplier to the American Red Cross, 15 USD is donated to Doctors Without Borders USA, 37.5 USD is donated to the American Red Cross, and 10 USD is donated to USA for UNHCR.