On the Cognitive Parsimony of Paralogical Arguments and their Impact in Automated Persuasion: Findings and Lessons Learned for Building Automatic Counter-Arguers

Antonio Lieto[1]

Abstract

Persuasive technologies can adopt several strategies to change the attitudes and behaviors of their users. In this work I synthesize the lessons learned from three empirical case studies on automated persuasion that have been carried out in the last decade in the contexts of: persuasive news recommendations, social robotics, and e-commerce, respectively. In particular, such studies have assessed, in the technological domain, the effects of nudging techniques relying on well known persuasive argumentation schemas and on framing strategies. In discussing the main findings, I will argue that the obtained persuasive effects are due to the fact that such techniques leverage on cognitive mechanisms that refer to the "system 1" types of automatic processes hypothesized in the context of the dual process theory of reasoning. As a consequence of this state of affairs, any automated persuasive systems (used for ethical purposes: e.g. suggesting good health habits) should be able to re-use these types of system 1 strategies in order to gradually take the users in an argumentation territory where system 2 processes can take place.

---

[1] Università di Torino, Dipartimento di Informatica and ICAR-CNR

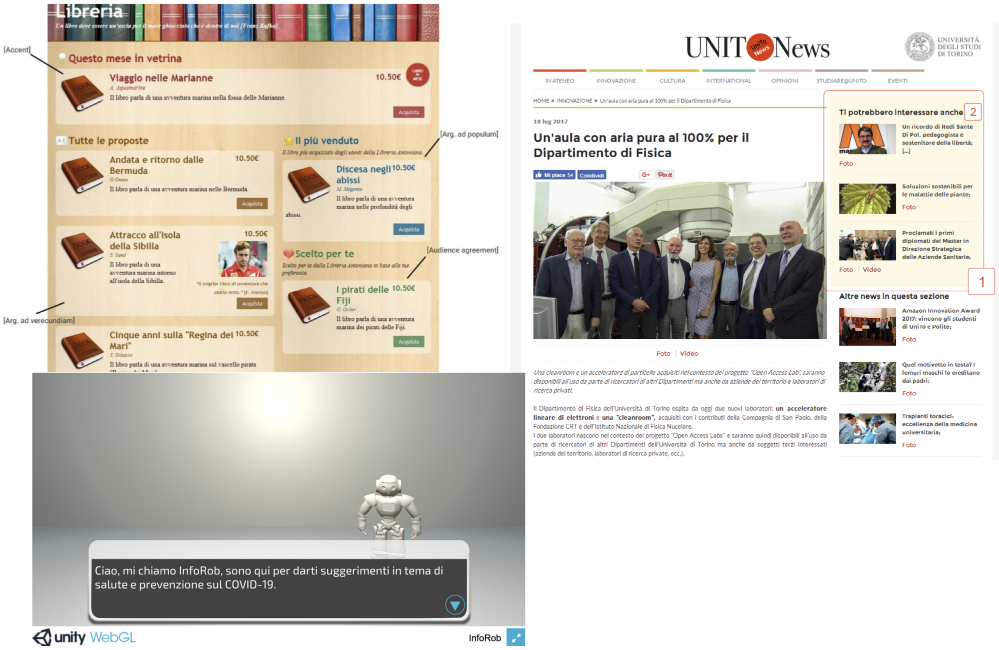
1. Introduction

Recently, the exploitation of well-known cognitive tendencies to design and evaluate the effect of nudging elements in digital environments is gaining widespread attention in the field of persuasive technologies [1][2]. The work framed by this paper can be ascribed to this class of analysis. In particular, our focus revolves around a well known class of persuasive techniques, used since the antiquity in human-human communication (e.g. in rhetoric and argumentation) known as fallacies. Fallacious schemata and arguments have been described as arguments that "even if invalid from a formal point of view, appear as plausible and therefore are psychologically persuasive" [3], [4] [5]. During the centuries different research areas such as logic, rhetoric and argumentation theory have pointed out that fallacious arguments are suitable to be used as techniques for achieving persuasive goals [6][7]. In the last decades, in the field of argumentation theory (and computational argumentation), a number of criticisms have been raised about the use of classical logic as an instrument for the analysis of fallacious arguments, and some alternative solutions have been proposed in order to justify the use of such arguments in certain contexts (e.g. in the case of the "New Dialectic" approach proposed by Douglas Walton [8]. By following this strand, an important aspect to point out regards the connection between inferential validity and rationality: a fallacious argument, is indeed not to be intended as "irrational" but as an heuristic shortcut [9]. In fact, since the psychological/cognitive aspect plays a crucial role in the dynamics of persuasion, a fallacious argument is usually an invalid argument endowed with psychological plausibility and a proper heuristic value. It is within this overall rational assumption that I present, below, the arguments and techniques that have been used in the three different case studies above mentioned.

2. Fallacies, Framing and Persuasive Technologies at Work

B.J. Fogg coined the term *captology* (*Computers As Persuasive Technologies*) in the 1990s, with the aim to describe a new research area which views computer technologies as potential persuaders and concentrates on both their design and their analysis [10].

This area is presently commonly referred to as "persuasive technologies". In this field, the connection between fallacies and technology-based persuasion has been first pointed out by [11, 12], where the authors created a persuasion matrix mapping some well-known fallacious arguments to some design features available in websites and mobile apps. This kind of theoretical grounding has also directly inspired the current empirical investigation since the connection with the huge theoretical background provided by the disciplines that, over the centuries, have dealt with all the major aspects of fallacious arguments (in logic, rhetoric, and persuasion) represents a unique and reliable source of knowledge to exploit for the study and design of "computer-driven" persuasion mechanisms. Figure 1 reports some screenshots from the different scenarios (i.e. e-commerce websites [12], persuasive news recommendations of a real online Magazine of the University of



Libreria

Un libro deve essere un'ascia per il mare ghiacciato che è dentro di noi [Franz Kafka].

Questo mese in vetrina

Viaggio nelle Marianne 10.50€

A. Aquamarine

Il libro parla di una avventura marina nella fossa delle Marianne.

Tutte le proposte

Andata e ritorno dalle Bermuda 10.50€

G.Green

Il libro parla di una avventura marina nelle Bermuda.

Attracco all'isola della Sibilla 10.50€

S. Sand

Il libro parla di una avventura marina intorno all'isola della Sibilla.

"Il miglior libro di avventure che abbia letto." [F. Alonso]

Cinque anni sulla "Regina dei Mari" 10.50€

T. Tobacco

Il libro parla di una avventura marina sul vascello pirata

Il più venduto

Il libro più acquistato dagli utenti della Libreria Antoniana.

Discesa negli abissi 10.50€

M. Magenta

Il libro parla di una avventura marina nelle profondità degli abissi.

Scelto per te

Scelto per te dalla Libreria Antoniana in base alle tue preferenze.

I pirati delle Fiji 10.50€

O. Ocher

Il libro parla di una avventura marina dei pirati delle Fiji.

UNITO News

UNIVERSITÀ DEGLI STUDI DI TORINO

HOME » INNOVAZIONE » Un'aula con aria pura al 100% per il Dipartimento di Fisica

18 lug 2017

Un'aula con aria pura al 100% per il Dipartimento di Fisica

Una cleanroom e un acceleratore di particelle acquisiti nel contesto del progetto "Open Access Lab", saranno disponibili all'uso da parte di ricercatori di altri Dipartimenti ma anche da aziende del territorio e laboratori di ricerca privati.

Il Dipartimento di Fisica dell'Università di Torino ospita da oggi due nuovi laboratori: **un acceleratore lineare di elettroni** e **una "cleanroom"**, acquisiti con i contributi della Compagnia di San Paolo, della Fondazione CRT e dell'Istituto Nazionale di Fisica Nucelare.

I due laboratori nascono nel contesto del progetto "Open Access Labs" e saranno quindi disponibili all'uso da parte di ricercatori di altri Dipartimenti dell'Università di Torino ma anche da soggetti terzi interessati (aziende del territorio, laboratori di ricerca private, ecc.).

Ti potrebbero interessare anche

Un ricordo di Redi Sante Di Pol, pedagogista e sostenitore della libertà; [...]

Soluzioni sostenibili per le malattie delle piante;

Proclamati i primi diplomati del Master in Direzione Strategica delle Aziende Sanitarie;

Altre news in questa sezione

Amazon Innovation Award 2017: vincono gli studenti di UniTo e Polito;

Quel motivetto in testa? I lemuri maschi lo ereditano dai padri;

Trapianti toracici: eccellenza della medicina universitaria;

Ciao, mi chiamo InfoRob, sono qui per darti suggerimenti in tema di salute e prevenzione sul COVID-19.

Figure 1: The three case studies assessing the persuasive efficacy of fallacious arguments and framing: an e-commerce website (top left), an online newspaper (top right) and a virtual robot trying to convince people to follow COVID-19 rules and to get vaccine shots (bottom left).

Turin [13] and social robotics [14] that have been used as testbed for assessing, in the technological realm, the persuasive efficacy of different techniques known in rethorics and in the context of the studies of human reasoning and decision making.

In such different case studies, the following arguments and strategies have been adopted and analyzed. The first one is known as "appeal to the majority" (or *Argumentum Ad Populum*). It consists of accepting a certain thesis based on the mere fact that the majority of people accept it. Its typical characterization is the following: "Most people think that X is true/false, then X is true/false" (where "X" can be any statement).

This argument can be compared to those strategies commonly used in the realm of persuasive technologies, which owe their persuasive potential to the exploitation of social dynamics. In particular, Fogg ([10] refers to well-known social psychology theories (e.g., social comparison and conformity [[15]]), which can be extended to include computer technologies. According to social comparison theory, people who are uncertain about the way they should behave in a situation proactively collect information about others and use it to build their own attitudes and behaviours. By contrast, conformity theory focuses on normative influence, stating that people who belong to a group usually experience a pressure to conform to the expectations of the other group members. In the context of the three case studies mentioned above, the argumentum ad populum has been exploited as follows: in the e-commerce websites, it is has been associated to the case in which "best seller" products are displayed in the interface of the website (in this case the persuasion strategy hidden under this interface choice is based on the following argument: "Most people buy/like X, then it is positive to buy X"). In the contexts of persuasive news recommendations and persuasive dialog with a virtual robot, on the other hand, this strategy has been employed via natural language, with expressions like "The Most Popular news is..." or "The majority of people think that ...", having the goal of triggering a certain choice as expression of a majority.

Another argumentation schema tested in the case studies is the so-called "appeal to the authority" (also *Argumentum Ad Verecundiam*). It is based on the exploitation of the 'halo effect' [16] according to which a positive evaluation on a given dimension (e.g., physical beauty) produces a halo which determines an extension of such an evaluation to other, unrelated, dimensions (e.g., expertise in a certain field). As indicated in [11], [12]: "such argument refers to cases of inappropriate transfer where some theses are assumed to merely hold because the people asserting them are, wrongly, assumed to be authorities about a certain topic due to their achievements and fame obtained in other, unrelated fields". In the context of the e-commerce websites, the argumentum ad verecundiam has been associated to the presence, in one or more parts of the website, of improper testimonials for certain products (e.g. "Mario Draghi suggests this cooking book"). In the cases of online news magazine and virtual robot, on the other hand, this argument was again delivered via natural language sentences of the form "According to (Person/Organization X) [claim Y])[2]

Another investigated technique is the one of personalization. This technique can also be considered persuasive they explicitly recall he so-called *audience agreement* technique: a well-known strategy [6, 12] suggesting that persuaders should only use arguments that they know to be already accepted by their audience in order to be effective. the because they Such technique can be regarded as fallacious because it also assumes a sort of static the persistence which, although being reasonable, cannot be taken for granted. Personalized information does not only save users the effort to examine a huge amount of content, but it is also more likely to draw their attention and, in case the system suggestions are accepted, it can cause longer-lasting and deeper changes. Tailoring is somehow similar to the so-called *audience agreement* technique, a well-known strategy in rhetoric and theory of argumentation [6, 12] which suggests that persuaders should only use arguments that they know to be already accepted by their audience in order to be effective. This technique was investigated only in the context of the e-commerce website with personalized recommendations.

Another analyzed strategy is the *Accent fallacy* [5, 17], which occurs when a particular emphasis on a part of a sentence is used to manipulate the actual

---

[2] Where Person/Organization X does now have any expertise for supporting the claim Y.

meaning of a proposition. It is commonly adopted with a persuasive intent in computer technologies, especially in its visual variant where certain elements are made more visually prominent in order to emphasize (or de-emphasize) them. A common example of the (visual) accent fallacy occurs when special offers (e.g., discounts) are highlighted with big fonts and bright colors, while the possibly restrictive conditions to enjoy them are made scarcely visible. This kind of presentation is fallacious since the inference drawn by the users is than one of considering relevant the emphasized information (e.g., the *suggested* conclusion is: take the special offer!) and not relevant the de-emphasized one (in our example: the restrictive constraint conditions)[3]. The accent fallacy can be compared to the concept of misplaced salience in Human–Computer Interaction, which is known as one of the factors limiting situation awareness [18] due to the emphasis it provided to irrelevant cues, leading users to confusion activities and inappropriate behaviors. This technique, in its visual variant, has been investigated in both the e-commerce case study and in the one about persuasive recommendations of news.

Another persuasive technique exploited in our work is the so-called *framing* [19], [20]. It refers to the role of the context in shaping people's decisions. In fact, using a particular wording instead of another might determine a different configuration of a given problem that consequently, may lead to a given interpretation of a sentence's meaning. A corollary of the framing effect consists of the fact that there is an asymmetry between prospective losses or wins in the people choice's architecture [20]. This effect, known and theorized in the *prospect theory*, means that the more something is perceived as scarce, the more the prospective loss is valued as problematic (and this usually leads to a less risk-seeking behaviour [21], [22] or to an action aimed at removing this sense of potential loss). The effect of framing has been investigated in both the persuasive news recommendation and in the case of the virtual robot.

3. Main Results and Lessons Learned

Overall, 26 people (11 female and 15 male) and 63 people (23 men and 40 women) were analyzed respectively for the first (e-commerce) and the third (virtual robot) case studies, while 20,933 anonymous users were analyzed in the context of the persuasive news recommendations delivered for online magazine of the University of Turin. In all these case studies, the evaluation has been done first in a "non persuasive" setting (i.e. the absence of any persuasive strategy) and then by using the above listed strategies and arguments in order to assess the different decision choices between the two scenarios. The main figures emerging from these different case studies are reported in the table dyspayed in Figure 2

---

[3] It is worth-noticing that the mere use of color or of different fonts to highlight a particular aspect of a text or of an interface does not constitute, *per se* a fallacy. We are in the presence of a visual accent fallacy only in the case in which the element put in evidence (or voluntarily hidden) has the goal of driving the users towards a conclusion (e.g., “buy the book X instead of Y”) that is not logically justified by the premises.

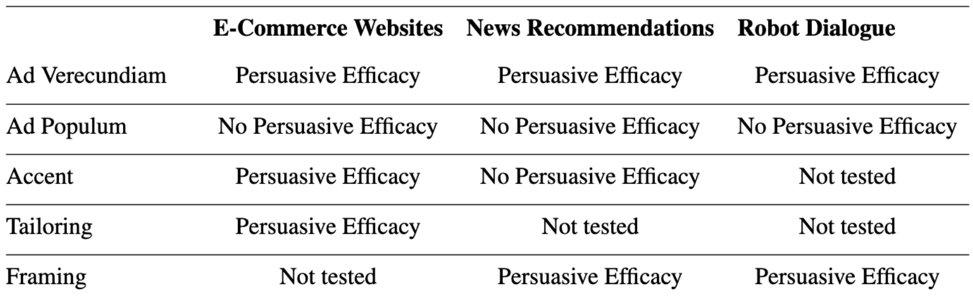
(for all the details please refer to the specific papers). In particular, it turns out that - also in the context of human-machine interaction - these techniques (and their mix) can play, with some notable exception, a measurable persuasive role.

More specifically: in the cases of e-commerce websites, users were influenced by fallacious arguments in almost half of the cases in a persuasive environment. For what concerns the effectiveness of the persuasion strategies, the use of the accent resulted to be the most successful strategy, followed by the audience agreement and by the argumentum ad verecundiam. Interestingly, however, in ([12] we also noticed a big exception: the argumentum ad populum was totally ineffective in our evaluation. Not only books presented as the "best sellers" were not chosen by any participant, but, in a follow-up focus group, a couple of them also stated that knowing what other people had bought was useless to them, and that suggestions based on the preferences of other people were far less relevant than personalized ones.

In the contexts of the news recommendations, a similar result for the *Argumentum Ad Populum* was obtained: it resulted to be not effective, and these results started to be aligned with other related works in other scenarios (see, e.g., [23–26]. In contrast, coherently with our expectations, framing appeared to be effective at promoting user clicks on recommended news, with a preference for negative framing, thus suggesting that sentences implicitly evoking a sense of loss are more effective than more neutral ones ([13].

Finally, in the context of the dialogical robot aiming at persuading users about the usefulness of following the anti-COVID rules and taking vaccine shots ([14], two main elements of interest emerged that are of interest for our purpose. First: the use of ad verecundiam and of the framing techniques seems to provide a persuasive effect if compared to the situation in which these techniques are not used. The ad populum, on the other hand, does not have any persuasive efficacy. As anticipated above, this datum is compliant with the other findings described in the two case studies mentioned above and on a surging amount of literature showing how this argumentation schema (widely used in classical rhetoric) does now show any effect when instantiated in the technological domain. Second: also for the techniques showing a measurable level of persuasive efficacy, the use of such strategies as sole elements of nudging does not seem sufficient to determine a long-lasting attitude change with respect to a particular belief on a topic. They can, however, be effectively coupled with other techniques like, for example, the use of a narrative strategy and the adoption of an ethical stance.

Overall, in the context of the case studies briefly reported above, the main element emerging from our analysis is that arguments and techniques widely used in the context of human-human persuasion also have (at some level and with the only exception of the argumentum ad populum) some impact in other human-machine persuasive contexts. In the section below, I argue that this success is due to the fact that the mechanisms activated by such strategies leverage on a fast processing of information belonging to what is called "system 1" in the context of the dual process theories of reasoning and rationality.



	E-Commerce Websites	News Recommendations	Robot Dialogue
Ad Verecundiam	Persuasive Efficacy	Persuasive Efficacy	Persuasive Efficacy
Ad Populum	No Persuasive Efficacy	No Persuasive Efficacy	No Persuasive Efficacy
Accent	Persuasive Efficacy	No Persuasive Efficacy	Not tested
Tailoring	Persuasive Efficacy	Not tested	Not tested
Framing	Not tested	Persuasive Efficacy	Persuasive Efficacy

Figure 2: Synthetic table concerning the overall analysis about the persuasive efficacy found for the analyzed strategies of argumentum ad verecundiam, argumentum ad populim, accent, tailoring and framing instantiated in the different technological contexts.

4. The Fast Track of Persuasive Arguments

The state of affairs depicted above can be grounded our on two different but interconnected theories coming from cognitive psychology, namely the ELM theory (Elaboration Likelihood Model (ELM) elaborated by [27] and the dual process theory and rationality [28–30]. While the first theory hypothesize two different information processing route (a central and a peripheral one, differing for the attention that a given receiver provides to the source of the message and, therefore, suggesting that the messages taking the peripheral route, i.e. the one where less attention is provided, are more akin to trigger fast and automatic cognitive mechanisms that are not subject to any form of deliberative control), the second one ascribes the emergence of errors in reasoning tasks to the execution of fast, associative and automatic processes (belonging to the so called System 1 processes), while slower processes (System 2 processes) are assumed responsible for the slow and cognitively demanding activity of producing answers that are correct with respect to the canons of normative rationality. The dual process approach was originally proposed to account for systematic errors in reasoning tasks: systematic reasoning errors (consider the classical examples of the selection task or the so-called conjunction fallacy) should be ascribed to fast, associative and automatic processes, while slower processes are responsible for the slow and cognitively demanding activity of producing answers that are correct with respect to the canons of normative rationality. An example is the well-known Linda problem, in which participants are given a description of Linda that stresses her independence and liberal views, and then asked whether it is more likely that she is (a) a bank teller or (b) a bank teller and active in the feminist movement. Participants tend to choose (b), since it fits the description of Linda (following the "heuristic representativeness"), even though the co-occurrence of two events cannot be more likely than one of them alone. Overall dual process theories suggests that our decision making processes are governed by two types of interacting cognitive systems, which are called respectively system(s) 1 and system(s) 2. Systems of the type 1, referred also as

S1, operate with rapid, automatic, associative processes of reasoning. They are phylogenetically older and execute processes in a parallel and fast way. Type 2 systems, referred also as S2, are, on the other hand, phylogenetically more recent and are based on conscious, controlled, sequential processes (also called type 2 processes) and on logic based rule following. As a consequence, if compared to system 1, system 2 processes are slower and cognitively more demanding.

In my view, both ELM and the dual-process theories of reasoning share a same underlying assumption that can be summarized and intended as a sort of fast-track processing route. Namely, a feasible possibility allowing to unify both theories is that persuasive strategies can be assumed to trigger heuristic-driven and fast (e.g. type 1) processes elaborated via the the ELM peripheral route. In this scenario, all the persuasive strategies described above, can be assumed to exploit the peripheral route [31],[32],[33] and, as such, can be arguably assumed to be processed automatically, thus eluding some deliberative forms of cognitive control that are executed, on the other hand, in the central route of information elaboration and with system 2 processes.

5. Implications for building Automatic Counter-Arguers

By assuming uses of the above (and other) persuasive techniques within the boundaries that are considered ethically acceptable and that fall within the so-called *Ethical Framework for a Good AI Society* [34], two possible strategies for building systems able to automatically counter-argue against false beliefs can be used. The first one consists in the adoption of rational, system 2-based arguments, making appeal to scientific facts and knowledge and adopting scientifically sound arguments. This is the case, for example, of generative language technologies like chatGPT[4]. Figure 2 reports the sequences of a dialogue about Covid-19 vaccines (resembling the typical requests assessed in [14] and shows, in the dialogues nr. 2 and 3, how all the arguments used by the system in order to persuade users that are not willing to follow anti Covid-19 rules or to take the vaccine shots, are based on scientifically solid arguments. These type of arguments, however, despite necessary in the scientific arena have little chance to be successful for skeptic or highly polarized people. The evidences shown in the reported case studies, indeed, illustrate the persuasive efficacy of arguments and techniques that fall within commonsense reasoning schemas[5] rather than on well founded scientific arguments. The importance of commonsense reasoning in argumentation is acknowledged by [36] and, more in general, in AI, by pioneers like Marvin Minsky that suggested that the capability of dealing with commonsense knowledge and reasoning represents the grounding element to connect layers of growing thinking

---

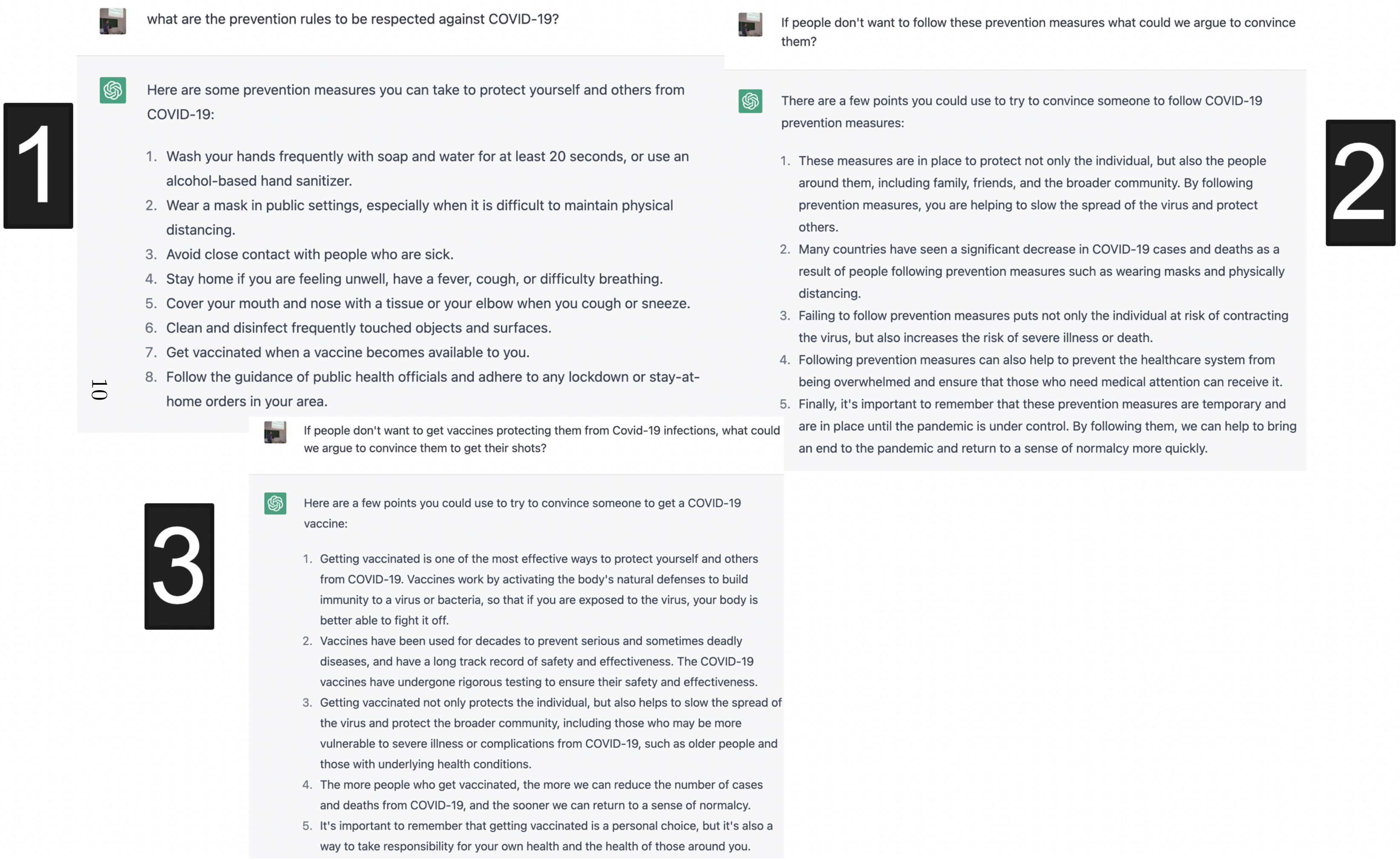
[4] `https://openai.com/blog/chatgpt/`

[5] Commonsense reasoning can be defined as the type of non-monotonic reasoning relying on the notion of typicality. This implies that, for example, not all the types of inductive or abductive inferences belong to the category of commonsense reasoning, but only those involving typical knowledge: i.e. elementary knowledge assumed to be widely accessible, whose main traits can be encoded by resorting to prototypes according to [35].

capabilities and, overall, low-level cognition (usually ascribed to fast, system 1, processes) and high-level cognition (usually ascribed to more cognitively demanding and slow system 2 processes). By taking into account this overall context, a more successful persuasive strategy could be based on the exploitation of a mix of ingredients. In particular, a persuasive strategy could be rather based on a two step procedure, where more rational arguments (i.e. system 2-driven) are used only after the initial employment of persuasive techniques (like the ones described above) leveraging system 1 processes. The use of fast-processed, persuasive techniques, used against a certain false belief/pre-conception, could play the strategic role of *taking time* while systems 2 processes start to be activated. To a certain extend, this insight could be in line with the recent findings coming from research on the depolarization of echo-chambers in social media, which show how the repeated exposition to alternative viewpoints, done with persuasive techniques like random dynamical nudging, tends to lead the people converge towards less extreme and more inclusive viewpoints [37]. In particular, the adoption of different persuasive tecniques (i.e. system 1-based in the first place, and then system 2-based) could correspond to the proposal of different viewpoints over the same argument/topic. Most importantly, however, these different viewpoints would trigger, with diverse weapons, different cognitive processes pointing towards the same persuasive goal (e.g. convincing people that it is important to get vaccinated). Adopting this kind of strategy seems also justified by recent studies in the context of diversity-seeking recommendations [38, 39] showing that, after a first emergence of cognitive resistance mechanisms (based on the automatic attempts of preservation of one's points of view), the gradual exposition to novel points of view can improve the willingness to widen and partially change one's own view.

6. Summing Up and Looking Ahead

Argumentation techniques are (and will be) widely used in the context of the next generation of persuasive and human-centered AI technologies [40]. In this paper, while reporting on the main findings obtained in the context of different case studies about the persuasive efficacy of different techniques tested in the technological setting, I have also advanced the hypothesis that the techniques proven to be more effective from a persuasive point of view are the ones triggering fast, automatic, system 1-triggering responses. The advanced hypothesis is not a *proof* of this state of affairs, but rather relies on a series of justified arguments based on the connection between the ELM and the dual process theory, and on a mounting empirical evidence coming from different application scenarios. The main underlying assumption behind the proposed hypothesis is based on a parsimony-argument: it is less laborious to process and activate system 1 processes when presented with an argument. Such processes however, also error-prone, and do not make use of more deliberative forms of rationality that is necessary in decision making. A major implication stemming from this hypothesis is in the design of two-step procedure for argumentation technologies used for ethical persuasion. For people that are, indeed, highly



what are the prevention rules to be respected against COVID-19?

Here are some prevention measures you can take to protect yourself and others from COVID-19:

1. Wash your hands frequently with soap and water for at least 20 seconds, or use an alcohol-based hand sanitizer.
2. Wear a mask in public settings, especially when it is difficult to maintain physical distancing.
3. Avoid close contact with people who are sick.
4. Stay home if you are feeling unwell, have a fever, cough, or difficulty breathing.
5. Cover your mouth and nose with a tissue or your elbow when you cough or sneeze.
6. Clean and disinfect frequently touched objects and surfaces.
7. Get vaccinated when a vaccine becomes available to you.
8. Follow the guidance of public health officials and adhere to any lockdown or stay-at-home orders in your area.

If people don't want to follow these prevention measures what could we argue to convince them?

There are a few points you could use to try to convince someone to follow COVID-19 prevention measures:

1. These measures are in place to protect not only the individual, but also the people around them, including family, friends, and the broader community. By following prevention measures, you are helping to slow the spread of the virus and protect others.
2. Many countries have seen a significant decrease in COVID-19 cases and deaths as a result of people following prevention measures such as wearing masks and physically distancing.
3. Failing to follow prevention measures puts not only the individual at risk of contracting the virus, but also increases the risk of severe illness or death.
4. Following prevention measures can also help to prevent the healthcare system from being overwhelmed and ensure that those who need medical attention can receive it.
5. Finally, it's important to remember that these prevention measures are temporary and are in place until the pandemic is under control. By following them, we can help to bring an end to the pandemic and return to a sense of normalcy more quickly.

If people don't want to get vaccines protecting them from Covid-19 infections, what could we argue to convince them to get their shots?

Here are a few points you could use to try to convince someone to get a COVID-19 vaccine:

1. Getting vaccinated is one of the most effective ways to protect yourself and others from COVID-19. Vaccines work by activating the body's natural defenses to build immunity to a virus or bacteria, so that if you are exposed to the virus, your body is better able to fight it off.
2. Vaccines have been used for decades to prevent serious and sometimes deadly diseases, and have a long track record of safety and effectiveness. The COVID-19 vaccines have undergone rigorous testing to ensure their safety and effectiveness.
3. Getting vaccinated not only protects the individual, but also helps to slow the spread of the virus and protect the broader community, including those who may be more vulnerable to severe illness or complications from COVID-19, such as older people and those with underlying health conditions.
4. The more people who get vaccinated, the more we can reduce the number of cases and deaths from COVID-19, and the sooner we can return to a sense of normalcy.
5. It's important to remember that getting vaccinated is a personal choice, but it's also a way to take responsibility for your own health and the health of those around you.

Figure 3: chatGPT sequence of responses in a dialogue about Covid-19 rules and the arguments to use in the hypothetical case of people showing an adverse beliefs/mindset to i) using prevention measures to mitigate the COVID-19 spreading and ii) to get vaccine shots.

polarized or skeptic about well established scientific facts and theories there is, indeed, no chance of having any success in persuading them of the contrary by using rational arguments. However, there is room for having some persuasive effect when the confrontation with them (e.g. in a dialogue operated by an artificial system) uses their same weapons (on a first stance) and then gradually converge towards more rational explanations and argumentations. It is also worth-noticing, if we consider the same problem from a different angle, that the very same two-step strategy could be used to built systems able to rebut to the eventual misuse of such techniques (e.g. when employed for unethical purposes). In this specific case, in fact, the goal of the automatic counter-arguer should be first the one of detecting the eventual use of persuasive technique within an argument (like in the case of fallacies in [41], then evaluate if the use of such technique can be somehow justified by higher ethical principles (e.g. convincing users to stop smoking) or if it represents a case of unethical use. In this second case, the system could adopt the above sketched two-step procedure. In the next years, considering also the latest advancement of language technologies, the focus on strategical - argumentation-based - elements governing the dialogues of human-centered AI systems will become more and more central and will represent an important occasion for testing the proposed hypothesis.

-

References

[1] G. Theocharous, J. Healey, S. Mahadevan, M. Saad, Personalizing with human cognitive biases, in: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, UMAP'19 Adjunct, ACM, New York, NY, USA, 2019, pp. 13–17. `doi:10.1145/3314183.3323453`. URL `http://doi.acm.org/10.1145/3314183.3323453`

[2] C. Schneider, M. Weinmann, J. vom Brocke, Digital nudging: guiding online user choices through interface design, Communications of the ACM 61 (7) (2018) 67–73.

[3] M. R. Cohen, E. Nagel, J. Corcoran, An introduction to logic, Hackett Publishing, 1993.

[4] C. L. Hamblin, Fallacies, Vale Press, 1970.

[5] D. Ross, Aristotle Topica Et Sophistici Elenchi, Oxford classical texts, 1963.

[6] C. Perelman, L. Olbrechts-Tyteca, Traité de l'argumentation. La nouvelle rhétorique, Presses universitaires de France, 1958.

[7] U. Eco, The ars magna by ramon llull, Contributions to science (2017) 47–50.

[8] D. N. Walton, The new dialectic: Conversational contexts of argument, University of Toronto Press, 1998.

[9] A. Lieto, Cognitive design for artificial minds, Routledge, 2021.

[10] B. J. Fogg, Persuasive Technology: Using Computers to Change what we Think and Do, Morgan Kaufmann Series in Interactive Technologies, Morgan Kaufmann, Amsterdam, 2003.
URL `http://www.sciencedirect.com/science/book/9781558606432`

[11] A. Lieto, F. Vernero, Unveiling the link between logical fallacies and web persuasion, in: Proceedings of the 5th Annual ACM Web Science Conference, ACM, 2013, pp. 473–478.

[12] A. Lieto, F. Vernero, Influencing the others' minds: An experimental evaluation of the use and efficacy of fallacious-reducible arguments in web and mobile technologies, PsychNology Journal 12 (3) (2014) 87–105.
URL `http://www.psychnology.org/File/PNJ12(3)/PSYCHNOLOGY_JOURNAL_12_3_LIETO.pdf`

[13] C. Gena, P. Grillo, A. Lieto, C. Mattutino, F. Vernero, When personalization is not an option: An in-the-wild study on persuasive news recommendation, Information 10 (10) (2019) 300.

[14] A. Augello, G. Città, M. Gentile, A. Lieto, A storytelling robot managing persuasive and ethical stances via act-r: an exploratory study, International Journal of Social Robotics (2021) 1–17.

[15] J. Turner, Social Influence, A Volume in the Brooks/Cole Mapping Social Psychology Series, Brooks/cole Publishing Company, 1991.
URL `https://books.google.it/books?id=bFq4AAAAIAAJ`

[16] B. J. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, M. Treinen, What makes web sites credible?: a report on a large quantitative study, in: J. A. Jacko, A. Sears (Eds.), Proceedings of the CHI 2001 Conference on Human Factors in Computing Systems, Seattle, WA, USA, March 31 - April 5, 2001., ACM, 2001, pp. 61–68. `doi:10.1145/365024.365037`.
URL `http://doi.acm.org/10.1145/365024.365037`

[17] I. M. Copi, C. Cohen, K. McMahon, Introduction to logic, Routledge, 2016.

[18] M. R. Endsley, B. Bolte, D. G. Jones, Designing for Situation Awareness: An Approach to User-Centered Design, CRC Press, 2003.

[19] G. Lakoff, The all new don't think of an elephant!: Know your values and frame the debate, Chelsea Green Publishing, 2014.

[20] A. Tversky, D. Kahneman, The framing of decisions and the psychology of choice, Science 211 (4481) (1981) 453–458.

[21] K. Shimizu, Prospect theory, behavioral theory, and the threat-rigidity thesis: Combinative effects on organizational decisions to divest formerly acquired units, Academy of Management Journal 50 (6) (2007) 1495–1514.

[22] T. B. Palmer, R. M. Wiseman, Decoupling risk taking from income stream uncertainty: A holistic model of risk, Strategic Management Journal 20 (11) (1999) 1037–1062.

[23] S. Gkika, G. Lekakos, The persuasive role of explanations in recommender systems, in: A. Öörni, S. Kelders, L. van Gemert-Pijnen, H. Oinas-Kukkonen (Eds.), Second International Workshop on Behavior Change Support Systems, no. 994 in CEUR Workshop Proceedings, Aachen, 2014, pp. 59–68.
URL `ceur-ws.org/Vol-1153/Paper_6.pdf`

[24] H. Zhu, B. A. Huberman, To switch or not to switch: Understanding social influence in online choices, American Behavioral Scientist 58 (10) (2014) 1329–1344. `arXiv:https://doi.org/10.1177/0002764214527089`, `doi:10.1177/0002764214527089`.
URL `https://doi.org/10.1177/0002764214527089`

[25] J. Su, A. Sharma, S. Goel, The effect of recommendations on network structure, in: Proceedings of the 25th International Conference on World Wide Web, WWW '16, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2016, pp. 1157–1167. `doi:10.1145/2872427.2883040`.
URL `https://doi.org/10.1145/2872427.2883040`

[26] M. Cataldi, M.-A. Aufaure, The 10 million follower fallacy: audience size does not prove domain-influence on twitter, Knowledge and Information Systems 44 (3) (2015) 559–580.

[27] R. E. Petty, J. T. Cacioppo, The elaboration likelihood model of persuasion, in: Communication and persuasion, Springer, 1986, pp. 1–24.

[28] K. E. Stanovich, R. F. West, Individual differences in reasoning: Implications for the rationality debate?, Behavioral and brain sciences 23 (5) (2000) 645–665.

[29] J. S. B. Evans, K. E. Frankish, In two minds: Dual processes and beyond., Oxford University Press, 2009.

[30] D. Kahneman, Thinking, fast and slow, Macmillan, 2011.

[31] R. E. Petty, J. Barden, S. C. Wheeler, The elaboration likelihood model of persuasion: Developing health promotions for sustained behavioral change.

[32] R. E. Petty, P. Briñol, The elaboration likelihood model, Handbook of theories of social psychology 1 (2011) 224–245.

[33] D. J. O'Keefe, The elaboration likelihood model, The Sage handbook of persuasion: Developments in theory and practice (2013) 137–149.

[34] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, et al., Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations, Minds and Machines 28 (4) (2018) 1–19.

[35] E. Rosch, Prototype classification and logical classification: The two systems, New trends in conceptual representation: Challenges to Piaget's theory (1983) 73–86.

[36] D. Walton, A pragmatic theory of fallacy.

[37] C. B. Currin, S. V. Vera, A. Khaledi-Nasab, Depolarization of echo chambers by random dynamical nudge, Scientific Reports 12 (1) (2022) 1–13.

[38] A. Lieto, G. L. Pozzato, M. Striani, S. Zoia, R. Damiano, Degari 2.0: A diversity-seeking, explainable, and affective art recommender for social inclusion, Cognitive Systems Research 77 (2023) 1–17.

[39] A. Lieto, M. Striani, C. Gena, E. Dolza, A. Marras, G. L. Pozzato, R. Damiano, A sensemaking system for grouping and suggesting stories from multiple affective viewpoints in museums.

[40] E. Dietz, A. Kakas, L. Michael, Argumentation: A calculus for human-centric ai, Frontiers in Artificial Intelligence 5.

[41] Z. Jin, A. Lalwani, T. Vaidhya, X. Shen, Y. Ding, Z. Lyu, M. Sachan, R. Mihalcea, B. Schölkopf, Logical fallacy detection, arXiv preprint arXiv:2202.13758.